Stochastic Endogenous Time Preference *

Youichiro Higashi, Kazuya Hyogo, and Norio Takeoka†

May 12, 2010

Abstract

In the literature of dynamic models under uncertainty, it is commonly assumed that uncertainty is resolved gradually over time and its resolution is independent of the actions taken by a decision maker up to that time period. However, contingencies in the decision maker’s mind may change and even increase over time according to habits that have evolved by experiences, actions, or consumptions in the past. We provide an axiomatic model where the decision maker faces uncertainty about her own time preference or discount factors in future and her belief over these uncertainty may evolve according to histories of the past consumption. Moreover, we provide a behavioral comparison about attitude toward flexibility across different histories, and characterize a property of the representation where the decision maker’s belief over future time preference at some history is more patient than her belief at another history in some stochastic sense.

Keywords: endogenous time preference, subjective states, preference for flexibility, comparative impatience, increasing convex [concave] stochastic order.

JEL classification: D11, D81, D91,

*An earlier version of this paper is circulated under the title “Random Discounting with Habits.” We would like to thank Chiaki Hara, Katsutoshi Wakai, and the audiences of the 2009 JEA Spring Meeting, FESAMES 2009, SWET 2009, ICNAAM 2009, Osaka GCOE Workshop, and the seminar participants at Osaka Keizai, Osaka Prefecture, and Toyama Universities, and Otaru University of Commerce for useful comments. Hyogo gratefully acknowledges financial support from a Grant-in-Aid for Young Scientists (B). All remaining errors are the author’s responsibility.

†Higashi is at the Faculty of Economics, Okayama University, Okayama, Japan, higashi@e.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 Faculty of Economics, Yokohama National University, 79-3 Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan, takeoka@ynu.ac.jp.
1 Introduction

1.1 Objective

In a dynamic decision making, a state space with a filtration on it is a standard tool for modeling uncertainty. It is commonly assumed that uncertainty is resolved gradually over time and its resolution is independent of the actions taken by a decision maker (DM) up to that time period. However, contingencies in the DM’s mind may change and even increase over time according to habits that have evolved by experiences, actions, or consumptions in the past. For example, a DM who has the habits of thrift may surely anticipate that facing a trade-off between a later but larger reward and an immediate but smaller reward, she will be patient enough to choose the former. But, according to having a higher standard of life, she may become less certain about her own time preference in future and anticipate that the immediate but smaller reward may become more attractive in some contingencies.

Our purpose is to provide an axiomatic model where the DM faces uncertainty about her own time preference or discount factors in future and these uncertainty may evolve according to histories of the past consumption. In decision theoretic models, expectations of taste shocks, whether they are about discount factors or other aspects of tastes, have been associated with preference for flexibility, which can be modeled by preference over opportunity sets (for example, Kreps [16, 17] and Dekel, Lipman, and Rustichini [5]). We incorporate histories of past consumption into this framework and model a DM as a collection of preferences depending on various histories. Then, a change of the DM’s belief over future time preference according to histories should be revealed as a change in attitudes toward flexibility regarding consumption plans.

In the framework without uncertainty, endogenous determination of time preference has been investigated in economics. Since Fisher [9], it is admitted that time preference may be affected by habits and change over time. Becker and Mulligan [2] consider a model in which a time preference depends on the quantity of resources the DM invests into making future pleasures less remote, for example, investment for health or education. However, past consumption may affect current time preference in a stochastic manner as mentioned in the first paragraph, and investment decisions may involve uncertainty in their own nature. Thus, our model provides a general framework for endogenous time preference accommodating uncertainty.

In case of endogenous time preference without uncertainty, if one’s discount factor at some history is greater than that at another history, she is said to be more patient at the former history than the latter. We would like to have a comparative notion of patience also in case of stochastic endogenous time preference. In our decision theoretic framework, it is possible to provide a behavioral comparison of patience across various histories. We show that this comparison characterizes the case where the DM’s belief over future time preference at some history is dominated by her belief at another history in the sense of the increasing convex and concave order, which is a generalization of the first-order stochastic dominance. Hence, if beliefs are deterministic and degenerate, this condition is simply
saying that the discount factor at some history is greater than that at another.

For the purpose, we adopt the same domain of choice used by Gul and Pesendorfer [11]. Let $C$ be the consumption set, which is a compact metric space. They show that there exists a compact metric space $Z$ such that $Z$ is homeomorphic to $\mathcal{K}(\Delta(C \times Z))$, where $\Delta(X)$ is the set of all Borel probability measures over $X$ and $\mathcal{K}(X)$ is the set of all non-empty compact subsets of $X$. An element of $Z$, called a menu, is an opportunity set of lotteries over pairs of current consumption and future menus.

We have in mind the following timing of decisions: Imagine a DM who has experienced a history of consumption $h = (\cdots, c_2, c_1)$. In period 0, the DM chooses a menu $x$ according to preference $\succeq_h$. At the moment of this choice, the DM is uncertain about future discount factors and has a belief about them. This belief is reflected somehow on $\succeq_h$ and will be elicited as a component of a representation (see below). In period $1^-$ (which is not a part of the primitives), current discount factor $\alpha$ becomes known to the DM, and she chooses a lottery $l$ out of the menu $x$ in period 1. In period $1^+$, the DM receives a pair $(c, x')$ according to realization of the lottery $l$, and consumption $c$ takes place. At this point in time, the DM evaluates $x'$ according to preference $\succeq_{hc}$, where $hc = (\cdots, c_2, c_1, c)$ is an updated history. The DM expects another discount factor $\alpha'$ to be realized in the following period. Subsequently, she chooses another lottery $l'$ out of the menu $x'$, and so on.

We consider a collection of preferences $\{\succeq_h\}_{h \in H}$ over $Z$, where $h = (\cdots, c_2, c_1)$ is a history of past consumption and $H$ is the set of all histories. We axiomatize $\{\succeq_h\}_{h \in H}$ having the following representation: there exist a non-constant continuous mixture linear function $u : \Delta(C) \to \mathbb{R}$ and a history-dependent belief $\mu(\cdot|h)$ about discount factors $\alpha \in [0, 1]$ such that

$$U(x|h) = \int_{[0, 1]} \max_{l \in x} \int_{C \times Z} \left( (1 - \alpha)u(c) + \alpha U(z|hc) \right) dl(c, z) d\mu(\alpha|h),$$

where $hc = (\cdots, c_2, c_1, c)$ if $h = (\cdots, c_2, c_1)$. Moreover, we show that $u$ is unique up to a positive affine transformation and $\mu(\cdot|h)$ is uniquely identified for all $h \in H$.

In this model, the DM surely anticipates that her instantaneous utility function $u$ will not change over time, while she is uncertain about her future discount factor $\alpha$ and has a belief $\mu(\alpha|h)$ on it. After observing $\alpha$, the DM will choose a lottery $l \in x$ so as to maximize the expected utility of the weighted average between an instantaneous utility $u(c)$ and a continuation value $U(z|hc)$. Notice that an immediate consumption, $c$, has two effects: One is a direct effect on an instantaneous utility $u(c)$, and the other is an indirect effect on a
continuation utility $U(\cdot | hc)$ (or future belief $\mu(\cdot | hc)$) through history dependence. A menu $x$ is evaluated according to its expected maximum value with respect to her belief $\mu(\cdot | h)$.

1.2 Changing in Preference for Flexibility

We illustrate how histories of past consumption affect beliefs about time preference in future and how these beliefs are in turn reflected on behavior of the DM. A key behavior is an attitude toward flexibility and its change according to habits.

Let $C$ stand for a set of monetary payoffs. For simplicity, we use a usual notation $(c_1, c_2, c_3, \cdots)$ instead of $(c_1, \{(c_2, \{(c_3, \cdots)\})\})$. Let $0 = (0, 0, \cdots)$.

Imagine a DM who has experienced a low level of consumption in her life and has preference $\geq_L$ over menus. Now consider two options, $(0, 110, 0)$ and $(100, 0, 0)$. Suppose that the DM is patient and prefers a later but lager reward to an immediate but smaller reward, that is, $(0, 110, 0) \geq_L (100, 0, 0)$. Moreover, since the DM has experienced a low level of consumption in her life, she may have no contingency in her mind where she will become impatient and choose $(100, 0, 0)$ over $(0, 110, 0)$. Hence it is not beneficial to keep $(100, 0, 0)$ with $(0, 110, 0)$ as an option. That is,

$$
\{(100, 0, 0), (0, 110, 0)\} \sim_L \{(0, 110, 0)\} \geq_L \{(100, 0, 0)\}.
$$

However, once the DM experiences some higher level of consumption for certain time periods, she may become less certain about her time preference and expect that $(100, 0, 0)$ may be chosen over $(0, 110, 0)$ in some contingencies. That is, in terms of preference $\succ_{LH}$, the DM may have preference for flexibility:

$$
\{(100, 0, 0), (0, 110, 0)\} \succ_{LH} \{(0, 110, 0)\} \succ_{LH} \{(100, 0, 0)\}.
$$

1.3 Related Literature

1.3.1 Endogenous Time Preference

Fisher [9] discussed how time preference or discount factor depends on individual characteristics and argued that one of the determinants of time preference is “habits”. He provided intuitive arguments for the endogenous time preference depending on habits and its change over time:\footnote{See Fisher [9, pp. 337-338]. This quote is the same as in Shi and Epstein [23].}

It has been noted that a person’s rate of preference for present over future income, given a certain income stream, will be high or low according to the past habits of the individual. If he has been accustomed to simple and inexpensive ways, he finds it fairly easy to save and ultimately to accumulate a little property. The habits of thrift being transmitted to the next generation, by imitation or by heredity or both, result in still further accumulation. The foundations of some of the world’s greatest fortunes have been based upon thrift.
Reversely, if a man has been brought up in the lap of luxury, he will have a keener desire for present enjoyment than if he had been accustomed to the simple living of the poor. The children of the rich, who have been accustomed to luxurious living and who have inherited only a fraction of their parent’s means, may spend beyond their mean and thus start the process of the dissipation of their family fortune. In the next generation this retrograde movement is likely to gather headway and to continue until, with the gradual subdivision of the fortune and the reluctance of the successive generations to curtail their expenses, the third or fourth generation may come to actual poverty.

The accumulation and dissipation of wealth do sometimes occur in cycles. Thrift, ability, industry and good fortune enable a few individuals to rise to wealth from the ranks of the poor. A few thousand dollars accumulation under favorable circumstances may grow to several millions in the next generation or two. Then the unfavorable effects of luxury begin, and the cycle of poverty and wealth begins anew. The old adage, “From shirt sleeves to shirt sleeves in four generations,” has some basis in fact.

Uzawa [26] formulates a model of endogenous rate of time preference in which the discount function depends on current consumption and future consumption. The Uzawa utility function has been widely used in application where the implications of constant time preference are unappealing. Shi and Epstein [23] assume that a DM’s discount function depends not only on current consumption and future consumption, but also on an index of past consumption or habit. In a closed economy, their model generates the cyclical pattern of wealth accumulation. In our model, the discount factor is random and independent of future consumption, but depends on habits. We analyze the changes of the DM’s attitude toward flexibility over time that arises from random discounting with habits.

1.3.2 Axiomatic Models

We study preference on opportunity sets (or menus) in order to provide a foundation for random discounting with habit. Opportunity sets are first considered as choice objects by Koopmans [15] to model sequential decision making. He points out that if a DM anticipates preference change in future, she may strictly prefer to keep some flexibility in her options rather than to choose a completely specified future plan. Kreps [16, 17] derive the set of anticipated future preferences from preference over opportunity sets and call it the subjective state space. By assuming richer choice objects than in Kreps [16, 17], Dekel, Lipman, and Rustichini [5] (henceforth DLR) uniquely identify the subjective state space. Their argument on the additive representation is modified by Dekel, Lipman, Rustichini, and Sarver [6].

---

2Its axiomatic foundation is provided by Epstein [8], in which the DM’s preference is defined on lotteries over consumption stream, and her von Neumann-Morgenstern utility index follows Uzawa functional form.
Higashi, Hyogo, and Takeoka [14] provide an axiomatic model of random discounting, where DLR’s model is extended to an infinite-horizon setting and the subjective state space is specified to the set of sequences of discount factors. More formally, preference over \( Z \) having the following representation is axiomatized:

\[
U(x) = \int_{[0,1]} \max_{l \in x} \int_{C \times Z} \left( (1 - \alpha)u(c) + \alpha U(z) \right) dl(c, z) d\mu(\alpha).
\]

In their model, the belief about future discount factors, that is, \( \mu \), is uniquely identified, but is assumed to be i.i.d. and independent of past consumption. Hence the DM’s attitude toward flexibility is the same over time. In contrast, in the present paper, the belief about future discount factors depends on past consumption and hence is not restricted to be i.i.d.

Gul and Pesendorfer [11] originally consider preference over \( Z \) and model a DM who suffers from self-control problem. Their DM anticipates that she will experience temptation or craving at the moment of choice, and hence, may prefer to limit her options in advance rather than to keep flexibility. Gul and Pesendorfer [12] consider a collection of preferences \( \{Z_h\} \) on \( Z \) which depend on histories of the last \( t \)-period consumption, and extend their self-control model to accommodate history dependence of degree of self-control. They axiomatize a recursive self-control representation where the strength of temptation depends on past consumption.

Rozen [21] provides a foundation for intrinsic habit model, which has been widely applied in asset pricing models, by considering a DM who has preference over deterministic consumption streams. If the DM’s preference is affected by habit, she may not prefer a consumption stream yielding high consumption at each date to one yielding low consumption at each date. The key axiom in her paper is that there is a collection of consumption streams under which the DM behaves as if she is habit-free.

\section{Model}

\subsection{Domain}

Let \( C \) be the outcome space (consumption set), which is assumed to be compact and metric. Let \( \Delta(C) \) be the set of lotteries, that is, all Borel probability measures over \( C \). Under the weak convergence topology, \( \Delta(C) \) is also compact and metric. For any compact metric space \( X \), let \( K(X) \) be denoted by the set of all non-empty compact subsets of \( X \). The set \( K(X) \) is endowed with the Hausdorff metric. \(^3\)

Gul and Pesendorfer [11] show that there exists a compact metric space \( Z \) such that \( Z \) is homeomorphic to \( K(\Delta(C \times Z)) \). Generic elements of \( Z \) are denoted by \( x, y, z, \ldots \). Each such object is called a menu (or an opportunity set) of lotteries over pairs of current consumption and menu for the rest of the horizon.

\(^3\)Details are relegated to Appendix A.
An important subdomain of $Z$ is the set $L$ of perfect commitment menus where DM is committed in every period. Then a perfect commitment menu can be viewed as a multistage lottery, that is, $L$ is a subdomain of $Z$ satisfying $L \simeq \Delta(C \times L)$. A formal treatment is found in the Appendix of Higashi, Hyogo, and Takeoka [14].

Another special subclass is a multistage lottery where a level of consumption is i.i.d. over time. Formally, for all $\ell \in \Delta(C)$, $\otimes \ell$ denotes the perfect commitment lottery such that

$$\ell \otimes \{\ell \otimes \{\ldots \}\}.$$  

Such lotteries are called i.i.d. lotteries. Let $\otimes \Delta_c \subset L$ denote the set of all i.i.d lotteries. A consumption stream with a constant level of consumption, that is, $(c, c, c, \ldots)$, is included in this subdomain under a suitable identification.

At each time period, preference is defined on $Z$, and depends on past consumption. Let $h = (\ldots, c_t, c_1)$ be a history of past consumption and $H \equiv \{h = (\ldots, c_t, \ldots, c_2, c_1) \mid c_t \in C$ for all $t = 1, 2, \ldots\}$ be the set of all histories, which is endowed with the metric

$$\rho(h, h') = \sum_{t=1}^{\infty} \frac{1}{2^t} \cdot \frac{d(c_t, c'_t)}{1 + d(c_t, c'_t)},$$

where $h = (c_1, c_2, \ldots)$, $h' = (c'_1, c'_2, \ldots)$, and $d$ is a metric on $C$.

4 For all $c \in C$ and $h = (\ldots, c_2, c_1) \in H$, the updated history $\ldots, c_2, c_1, c)$ will be denoted by $hc$. For each history $h \in H$, let $\succ_h$ be preference over $Z$. We consider a collection of preferences $\{\succ_h\}_{h \in H}$.

### 2.2 Functional Form

Let $u : \Delta(C) \to \mathbb{R}$ be a non-constant continuous mixture linear function. The set of discount factors is denoted by $[0, 1]$ and its generic element is denoted by $\alpha$. We model a DM who is uncertain about future discount factors and whose belief depends on past consumption. Formally, for all $h \in H$, let $\mu(\cdot|h)$ be a probability measure on $[0, 1]$. Let $\overline{\alpha}_h$ be the average of $\mu(\cdot|h)$, that is, $\overline{\alpha}_h \equiv \int_{[0,1]} \alpha \ d\mu(\alpha|h)$. A set of probability measures $\{\mu(\cdot|h)\}_{h \in H}$ is called a history-dependent belief and assumed to satisfy $\sup\{\overline{\alpha}_h \mid h \in H\} < 1$ and continuity of $\mu(\cdot|h)$ in $h$.

We formulate the following functional form: For each pair of an instantaneous utility function $u$ and a history-dependent belief $\{\mu(\cdot|h)\}_{h \in H}$, a history-dependent random discounting utility $U(\cdot|h) : Z \to \mathbb{R}$, $h \in H$, is defined as

$$U(x|h) = \int_{[0,1]} \max_{z \in x} \int_{C \times Z} \left( (1 - \alpha)u(c) + \alpha U(z|hc) \right) d(c, z) d\mu(\alpha|h). \quad (1)$$

$^4$Under the metric $\rho$, $H$ is complete (see Aliprantis and Border [1, p.90, Theorem 3.37]).
In this model, the DM surely anticipates that her instantaneous utility function \( u \) will not change over time, while she is uncertain about her future discount factor \( \alpha \) and has a belief \( \mu(\alpha|h) \) on it. After observing \( \alpha \), the DM will choose a lottery \( l \in x \) so as to maximize the expected utility of the weighted average between an instantaneous utility \( u(c) \) and a continuation value \( U(z|hc) \). Notice that an immediate consumption, \( c \), has two effects: One is a direct effect on an instantaneous utility \( u(c) \), and the other is an indirect effect on a continuation utility \( U(\cdot|hc) \) (or future belief \( \mu(\cdot|hc) \)) through history dependence. A menu \( x \) is evaluated according to its expected maximum value with respect to her belief \( \mu(\cdot|h) \).

Our functional form can accommodate several specifications used in macroeconomic models. Becker and Mulligan [2] suggest a model in which a discount factor (deterministically) depends on the quantity of resources the DM invests into making future pleasures less remote. For example, investment in education or health-related decisions may focus one’s attention on the future. Notice that these activities can be included into the set \( C \) of consumptions. If the outcome of investment is uncertain, consequently, discount factors will change stochastically.

An alternative interpretation of infinite-lived agents is to think of a dynasty, in which case a discount factor is regarded as a degree of altruism. The bequest motives of the current generations may depend on whether they are married or not, or whether they have descendants or not.

If a belief \( \mu(\cdot|h) \) is independent of \( h \), (1) is reduced to the history-independent random discounting model of Higashi, Hyogo, and Takeoka [14]:

\[
U(x) = \int_{[0,1]} \max_{l \in x} \int_{C \times Z} \left( (1 - \alpha)u(c) + \alpha U(z) \right) dl(c, z) d\mu(\alpha),
\]

or equivalently,

\[
U(x) = \int_{[0,1]} \max_{l \in x} \left( (1 - \alpha) \int_{C} u(c) dl_{c} + \alpha \int_{Z} U(z) dl_{z} \right) d\mu(\alpha),
\]

where \( l_{c} \) and \( l_{z} \) denote the marginal distributions of \( l \) on \( C \) and \( Z \) respectively.

3 Foundations

3.1 Axioms

The axioms which we consider on \( \{\succsim_{h}\}_{h \in H} \) are the following. The first four axioms have counterparts in DLR. The only difference is that we impose the same axioms on preferences over infinite horizon decision problems \( Z \) rather than a two-period domain \( K(\Delta(C)) \).

**Axiom 1 (Order).** For all \( h \in H \), \( \succsim_{h} \) is complete and transitive.

**Axiom 2 (Continuity).**
(i) For all \( h \in H \) and \( x \in \mathcal{Z}, \{ z \in \mathcal{Z} | x \gtrsim_h z \} \) and \( \{ z \in \mathcal{Z} | z \gtrsim_h x \} \) are closed.

(ii) For all \( x, y \in \mathcal{Z}, \{ h \in H | x \gtrsim_h y \} \) is closed.

Part (i) is a standard continuity condition. Part (ii) requires continuity with respect to histories, and has no counterpart in DLR.

For any \( x, x' \in \mathcal{Z} \) and \( \lambda \in [0,1] \), define the mixture of two menus by considering the mixtures element by element between \( x \) and \( x' \), that is,

\[
\lambda x + (1 - \lambda) x' \equiv \{ \lambda l + (1 - \lambda) l' | l \in x, l' \in x' \} \in \mathcal{Z}.
\]

**Axiom 3 (Independence).** For all \( h \in H, x, y, z \in \mathcal{Z} \) and \( \lambda \in (0,1] \),

\[
x \succ_h y \Rightarrow \lambda x + (1 - \lambda) z \succ_h \lambda y + (1 - \lambda) z.
\]

Independence can be justified by two steps as in DLR. The first step is a standard argument of the von-Neumann and Morgenstern Independence: If \( x \) is preferred to \( y \), then the lottery yielding \( x \) and \( z \) with probabilities \( \lambda \) and \( 1 - \lambda \), denoted by \( \lambda \circ x + (1 - \lambda) \circ z \), is preferred to \( \lambda \circ y + (1 - \lambda) \circ z \).

In the next step, we argue that \( \lambda \circ x + (1 - \lambda) \circ z \) is indifferent to \( \lambda x + (1 - \lambda) z \). Notice that these two alternatives differ in timing of resolution of randomization \( \lambda \). For the former, the DM makes choice out of a menu (either \( x \) or \( z \)) after the resolution of \( \lambda \), while for the latter, this order is reversed, that is, the choice out of the menu \( \lambda x + (1 - \lambda) z \) is made before the resolution of \( \lambda \). Suppose that a DM is uncertain about future preference over \( \Delta(C \times \mathcal{Z}) \), yet she surely anticipates that it will satisfy the expected utility axioms. Let \( l \) and \( l' \) be a rational choice from \( x \) and \( z \), respectively, with respect to a future preference. Therefore, \( \lambda l + (1 - \lambda) l' \) is the expected choice from \( \lambda \circ x + (1 - \lambda) \circ z \). On the other hand, if the future preference satisfies the expected utility axioms, \( \lambda l + (1 - \lambda) l' \) is also a rational choice from \( \lambda x + (1 - \lambda) z \). Therefore, irrespective of the future preference, the two alternatives will ensure indifferent consequences.

**Axiom 4 (Monotonicity).** For all \( h \in H, x, y \in \mathcal{Z}, \) if \( y \subset x \), then \( x \gtrsim_h y \).

This axiom says that a bigger menu is always weakly preferred, and is consistent with preference for flexibility.

Say that preferences \( \gtrsim^1 \) and \( \gtrsim^2 \) on \( \mathcal{Z} \) are equivalent on \( \Delta(C) \) if, for all \( \ell, \ell' \in \Delta(C) \),

\[
\{ \otimes \ell \} \succ^1 \{ \otimes \ell' \} \iff \{ \otimes \ell \} \succ^2 \{ \otimes \ell' \}.
\]

The next three axioms concern risk preference over \( \Delta(C) \).

**Axiom 5 (History-Independent Risk-Preference).** For all \( h, h' \in H, \gtrsim_h \) and \( \gtrsim_{h'} \) are equivalent on \( \Delta(C) \).
Suppose that the DM surely expects her attitude toward static risk, that is, preference over $\Delta(C)$, to be independent of histories of past consumption. Since a compound lottery such as $\otimes \ell$ provides the same risky consumption in every period, the ranking between $\{\otimes \ell\}$ and $\{\otimes \ell'\}$ should reflect solely static risk preference and have nothing to do with expectations for future time preference.

**Axiom 6 (Nondegeneracy).** For all $h \in H$, there exist $\ell, \ell' \in \Delta(C)$ such that $\{\otimes \ell\} \succ_h \{\otimes \ell'\}$.

The ranking in the axiom states that receiving a lottery $\ell \in \Delta(C)$ in every period is preferred to receiving $\ell'$ in every period, which means that an atemporal lottery $\ell$ is better than $\ell'$. The existence of such $\ell, \ell'$ implies that the DM’s risk preference over $\Delta(C)$ is sensitive.

**Axiom 7 (IID-Separability).** For all $h \in H, \ell, \ell' \in \Delta(C)$, and $\lambda \in [0, 1]$,

$$\{\lambda \otimes \ell + (1 - \lambda) \otimes \ell'\} \sim_h \{\otimes(\lambda \ell + (1 - \lambda)\ell')\}.$$

As explained above, if the DM surely anticipates that her risk preference is history-independent, a compound lottery such as $\otimes \ell$ is viewed as a static lottery $\ell$. Thus, the DM should be indifferent between $\lambda \otimes \ell + (1 - \lambda) \otimes \ell'$ and $\otimes(\lambda \ell + (1 - \lambda)\ell')$.\(^5\)

**Axiom 8 (Best-Worst).** For all $h \in H, x \in Z$, there exist $\ell, \ell' \in \Delta(C)$ such that $\{\otimes \ell\} \succ_h x \succ_h \{\otimes \ell'\}$.

For example, let $\otimes \ell^+$ be a maximal lottery among $\otimes\Delta_c$. If the DM surely anticipates her future risk preference to be time-invariant and is only concerned about uncertainty about intertemporal trade-offs, she is willing to make a commitment to $\otimes \ell^+$. On the other hand, she will never benefit from the commitment to a minimal lottery among $\otimes\Delta_c$.

**Axiom 9 (Strong Archimedean).** For all $\ell, \ell' \in \Delta(C)$ such that $\{\otimes \ell\} \succ_h \{\otimes \ell'\}$, there exists $\ell'' \in \Delta(C)$ such that $\{\otimes \ell\} \succ_h \{\otimes \ell''\} \succ_h \{\otimes \ell \otimes \ell\}$ for all $h \in H$.

By assumption of the axiom, an atemporal lottery $\ell$ is better than $\ell'$. Thus, in $\ell' \otimes \{\otimes \ell\}$, the better option is delayed until second period, and it should be strictly worse than $\otimes \ell$. The axiom requires that irrespective of histories, it is even worse than some lottery $\otimes \ell''$ which is strictly worse than $\otimes \ell$. Intuitively, this condition states that a loss from receiving a better lottery with one-period delay is uniformly bounded across histories.

The last two axioms concern intertemporal decision making and impose some restrictions on preferences having different histories.

**Axiom 10 (Dynamic Consistency).** For all $h \in H, x, y \in Z$, and $c \in C$,

$$x \succeq^{hc} y \iff \{(c, x)\} \succeq_h \{(c, y)\}.$$

\(^5\)This axiom excludes a DM who may care about timing of resolution of multistage lottery. A distinction between early and late resolution of risk is examined in Kreps and Porteus [18].
Dynamic Consistency requires that the ranking between \( \{(c, x)\} \) and \( \{(c, y)\} \) under a history \( h \) reflects how the DM evaluates \( x \) and \( y \) under the updated history \( hc \).

Finally, we introduce a key axiom called “Period-Wise Dominance”. Consider two options \((c, x)\) and \((c, x')\). At an intuitive level, if \( c \) is preferred to \( c' \) and \( x \) is preferred to \( x' \), then \((c, x)\) is definitely better than \((c', x')\) no matter what time preference the DM has. If this is the case, \((c, x)\) is said to dominate \((c', x')\), and the DM will never choose such dominated options. It is reasonable to assume that keeping dominated options within a menu adds no value of flexibility, that is,

\[
\{(c, x), (c', x')\} \sim_h \{(c, x)\}.
\]

We use this idea in Higashi, Hyogo and Takeoka [14] to characterize the history-independent random discounting model. However, in the current setup, the immediate consumption level, \( c \), has two effects; one is a direct effect from its consumption and the other is an indirect effect on evaluation of future opportunity sets through habit formation. In order to define an appropriate dominance notion, we need to separate these two effects.

To illustrate the dominance notion, take any two options \((c, x)\) and \((c_0, x_0)\). The ranking

\[
\{\otimes c\} \succsim_h \{\otimes c'\} \quad (2)
\]

should solely reflect the direct effect of immediate consumption. On the other hand, to capture the effect of habit formation, let \( \ell, \ell' \in \Delta(C) \) satisfy\(^6\)

\[
x \sim_{hc} \{\otimes \ell\} \quad \text{and} \quad x' \sim_{hc'} \{\otimes \ell'\}.
\]

The first indifference relation says that at history \( hc \), the DM is indifferent between having \( x \) and committing \( \ell \) in every time period. We can interpret \( \otimes \ell \) as an smoothing equivalent of \( x \) at history \( hc \). Similarly, \( \otimes \ell' \) is an smoothing equivalent of \( x' \) at history \( hc' \). Since the rankings over iid lotteries are assumed to be identical across histories, if

\[
\{\otimes \ell\} \succsim_h \{\otimes \ell'\}, \quad (3)
\]

then we know that the menu \( x \) after consumption \( c \) takes place is preferred to the menu \( x' \) after consumption \( c' \) takes place. That is, \((3)\) reflects the ranking of the two menus, \( x \) and \( x' \), with taking into account habit formation from difference in immediate consumption level, \( c \) and \( c' \). We say that \((c, x)\) dominates \((c', x')\) if conditions \((2)\) and \((3)\) hold.

Now we extend the above dominance notion to lotteries with finite supports. Fix a history \( h \). Take two lotteries \( l \) and \( l' \) in \( \Delta(C \times \mathcal{Z}) \) with finite supports, which can be denoted by \( l = ((c_i, x_i), \lambda_i)_{i=1}^n \) and \( l' = ((c'_i, x'_i), \lambda'_i)_{i=1}^m \). Let \( l_c \) and \( l'_c \) denote the marginal distributions of \( l \) and \( l' \) on \( C \), respectively. Since a comparison between \( \otimes l_c \) and \( \otimes l'_c \) solely reflects static risk attitude, the ranking \( \{\otimes l_c\} \succsim_h \{\otimes l'_c\} \) presumably means that \( l \) is preferred to \( l' \) in terms of current consumption.

\(^6\)Under the other axioms, each menu \( x \in \mathcal{Z} \) admits a smoothing equivalent. See Appendix B for details.
On the other hand, let \( \otimes l_i \) (resp. \( \otimes l'_i \)) be a smoothing equivalent of \( x_i \) (resp. \( x'_i \)) at history \( hc_i \) (resp. \( hc'_i \)). Since the DM receives \( x_i \) at history \( hc_i \) with probability \( \lambda_i \), by definition of smoothing equivalent, she should be indifferent between this situation and instead receiving \( \otimes l_i \) with probability \( \lambda_i \). Moreover, under Independence and IID-Separability, the latter would be indifferent to receive its reduced lottery \( \sum_{i=1}^{n} \lambda_i l_i \) every time period. Hence, the ranking \( \otimes(\sum_{i=1}^{n} \lambda_i l_i) \geq_h \{ \otimes(\sum_{i=1}^{n} \lambda_i l'_i) \} \) reveals that \( l \) is preferred to \( l' \) in terms of future opportunity sets with taking into account the effect of habit formation from current consumption.

**Definition 3.1.** For all \( h \) and \( l, l' \in \Delta(C \times Z) \) with finite supports, say that \( l \) dominates \( l' \) at history \( h \) if \( \otimes l_c \geq_h \{ \otimes l'_c \} \) and \( \otimes(\sum_{i=1}^{n} \lambda_i l_i) \geq_h \{ \otimes(\sum_{i=1}^{n} \lambda_i l'_i) \} \).

Let \( \Delta_s(C \times Z) \) be the set of lotteries with finite supports.

**Axiom 11 (Period-Wise Dominance).** For all \( h \in H, x \in Z \), and \( l, l' \in \Delta_s(C \times Z) \), if \( l \in x \) and \( l \) dominates \( l' \) at history \( h \), then

\[
x \cup \{l'\} \sim_h x.
\]

Period-Wise Dominance states that the DM should not care about dominated lotteries. Since dominated lotteries give less utilities in the future, both immediate and remote, they are useless for the DM who has in mind subjective contingencies concerning intertemporal trade-offs, and hence never chosen over a dominant lottery. As stated in the axiom, adding a dominated lottery to a menu does not change its ranking.

### 3.2 Representation Results

Now we state our main representation theorem. A proof is relegated to Appendix D.

**Theorem 3.1.** If a set of preferences \( \{ \geq_h \}_{h \in H} \) satisfies Order, Continuity, Independence, Monotonicity, History-Independent Risk-Preference, Nondegeneracy, IID-Separability, Best-Worst, Strong Archimedean, Dynamic Consistency, and Period-Wise Dominance, then it admits a history-dependent random discounting representation \( (u, \{ \mu(\cdot|h) \}_{h \in H}) \).

Conversely, for all history-dependent random discounting utilities \( (u, \{ \mu(\cdot|h) \}_{h \in H}) \), there exists a unique functional form \( U(\cdot|\cdot) \) that satisfies functional equation (1) and the preference it represents satisfies all the axioms.

The sufficiency is closely related to DLR’s study, and the role of the axioms may be well understood when compared with their axioms. DLR show that preference over menus of lotteries admits the additive representation of the form that

\[
U(x) = \int_S \max_{l \in x} V(l, s) d\mu(s),
\]
where $S$ is a state space, $\mu$ is a non-negative measure on $S$, and $V(\cdot, s)$ is a state-dependent expected utility function if and only if preference satisfies Order, Continuity-(i), Independence, and Monotonicity.\footnote{Dekel, Lipman, Rustichini, and Sarver \cite{6} fill a gap in DLR surrounding this representation result.} Thus, the other six axioms mainly play a role to specify a general state space $S$ to the set $[0, 1]$ of discount factors, and to convert $V(l, s)$ to the desired form. By History-Independent Risk-Preference, $u$ is independent of $h$, and by IID-Separability, it satisfies mixture linearity. Best-Worst, Dynamic Consistency, and Period-Wise Dominance jointly deliver the desired recursive structure of $V$ across different histories. By Nodegenacity, $u$ is non-constant, and Strong Archimedean implies that $\alpha_h$ is uniformly bounded from one. Continuity-(ii) ensures that the representation is continuous in histories.

The next result concerns uniqueness of the representation. A proof can be found in Appendix E.

**Theorem 3.2.** If history-dependent random discounting representations, $U$ and $U'$, with components $(u, \{\mu(\cdot|h)\}_{h \in H})$ and $(u', \{\mu'(\cdot|h)\}_{h \in H})$ respectively, represent the same set of preferences $\succsim_h$ for all $h \in H$, then:

(i) $u$ and $u'$ are cardinally equivalent; and

(ii) $\mu(\cdot|h) = \mu'(\cdot|h)$ for all $h \in H$.

Theorem 3.2 pins down a subjective probability measure $\mu(\cdot|h)$ over the set of future discount factors, which is interpreted as a belief over subjective states. Thus, our result is in contrast to Kreps \cite{16, 17} and DLR, where probability measures over subjective states are not identified because the ex post utility functions are state-dependent, and hence, probabilities assigned to those states can be manipulated arbitrarily. Here, it is a combination of the additive recursive structure and the normalization of discount factors that ensures uniqueness of subjective beliefs. The argument is the same as in Higashi, Hyogo, and Takeoka \cite{14}.

### 3.3 Special Case: Dependence on Average Consumption

In the history-dependent random discounting model, beliefs over future time preference depend on histories of past consumption in an arbitrary way. In the existing habit models, this history dependence has been often specified to dependence on the average consumption in a history. In this subsection, we provide such a specification.

From now on, let $C = [0, M]$ for some $M > 0$. The metric on $C$ is understood to be $d(c, c') = |c - c'|$. For some discount factor $\delta \in (0, 1)$, define the average consumption at history $h = (\ldots, c_2, c_1)$ by

$$\overline{c}_h = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} c_t.$$ 

If there exists $\delta \in (0, 1)$ such that for all $h, h' \in H$,

$$\overline{c}_h = \overline{c}'_h \implies \succ_h = \succ_{h'},$$
then \( \{ \succeq_h \}_{h \in H} \) is said to be an average-dependent preference.

For all \( c \in C \), let \( c \) denote the history where \( c \) has been consumed in every time period, that is, \( c = (\ldots, c, c, c) \). For all \( h = (c_t)_{t=1}^\infty, h' = (c'_t)_{t=1}^\infty \in H \), the \( \lambda \)-mixture between \( h \) and \( h' \) is defined as

\[
\lambda h + (1 - \lambda)h' \equiv (\lambda c_t + (1 - \lambda)c'_t)_{t=1}^\infty.
\]

For all \( h \in H \), define its equivalence class by

\[
[h] \equiv \{ h' \in H \mid h \succeq_h h' \}.
\]  

These equivalence classes define an equivalence relation on \( H \). We impose the following conditions on this equivalence relation.

**Axiom 12 (Sensitivity).** For all \( c, c' \in C \) with \( c \neq c', c' \notin [c] \).

This axiom requires that preferences differ between at histories \( c = (\ldots, c, c) \) and \( c' = (\ldots, c', c') \).

**Axiom 13 (Smoothing).** For all \( h \in H \), there exists \( c \in C \) such that \( c \in [h] \).

If \( \succeq_h \) is the same as \( \succeq_c \) for some \( c \in C \), \( c \) is interpreted as a “smoothed” history of \( h \), which induces the same habit as at the original history \( h \). Thus, the Smoothing axiom requires that each history admits its smoothed history.

**Axiom 14 (Continuity).** If \( h^n \to h \) and \( c^n \to c \) with \( c^n \in [h^n] \), then \( c \in [h] \).

This is a technical requirement, and says that the equivalence relation satisfies upper hemi-continuity.

**Axiom 15 (Linearity).** For all \( h, h' \in H, c, c' \in C \), and \( \lambda \in [0,1] \),

\[
c \in [h] \text{ and } c' \in [h'] \quad \Rightarrow \quad \lambda c + (1 - \lambda)c' \in [\lambda h + (1 - \lambda)h'].
\]

This axiom states that a mixed history between two histories \( h \) and \( h' \) admits a smoothed history which is the mixture between their own smoothed histories.

**Axiom 16.** There exists \( \hat{c} \in C \) such that if \( c \in [\hat{c}c] \) and \( c' \in [\hat{c}'c] \), then

\[
c > c' \iff \hat{c} > \hat{c}'.
\]

**Axiom 17.** For all \( h \in H \) and \( c, c_1 \in C \),

\[
c \in [h] \iff cc_1 \in [hc_1].
\]

The above two axioms capture “stationarity”. Since \( \hat{c}c \) and \( \hat{c}'c \) share the same consumption level in the immediate past, the difference between their smoothed histories should reflect the difference between \( \hat{c} \) and \( \hat{c}' \). A similar explanation can be applied to Axiom 17.
Axiom 18 (Separability). Suppose that \( c \in [hc_1], c' \in [hc'_1], \bar{c} \in [h'c_1], \) and \( \bar{c}' \in [h'c'_1]. \) Then, (i) \( c > c' \implies \bar{c} > \bar{c}', \) and (ii) \( c > \bar{c} \implies c' > \bar{c}'. \)

To interpret part (i), let \( c > c' \). By assumption, \( c \) (resp. \( c' \)) is a smoothed history of \( hc_1 \) (resp. \( hc'_1 \)). Since \( hc_1 \) and \( hc'_1 \) differ only in the immediate past, \( c > c' \) suggests \( c_1 > c'_1 \), which in turn implies that \( h'c_1 \) corresponds to a higher level of past consumption than that of \( h'c'_1 \). Thus, their smoothed histories, \( \bar{c} \) and \( \bar{c}' \) should satisfy \( \bar{c} > \bar{c}' \). A similar argument can be applied to part (ii).

Theorem 3.3. If the equivalence relation given as in (4) satisfy Axioms 12-18 if and only if \( \{\succeq_h\} \) is an average-dependent preference.

This result is obtained by translating the axioms for stationary cardinal utilities of Epstein [8] into the equivalence relation on histories. See Appendix F for a formal proof.

4 Comparative Impatience across Histories

We would like to analyze the situation where a DM is more patient at some history than at another. In case of deterministic discounting, the more patient the DM is, the closer to one of her discount factor. We provide a behavioral comparison about attitudes toward flexibility and characterize intuitive properties of subjective beliefs.

Consider a DM having history-dependent preference \( \{\succeq_h\}_{h \in H} \). Suppose that \( \{\succeq_h\}_{h \in H} \) admits a history-dependent random discounting representation. Let

\[
\begin{align*}
\mathcal{L}_P &\equiv \{ \ell \otimes \{ \otimes \ell' \} \in \mathcal{L} \mid \{ \otimes \ell' \} \succeq_h \{ \otimes \ell \} \text{ for some } \ell, \ell' \in \Delta(C) \}, \\
\mathcal{L}_I &\equiv \{ \ell \otimes \{ \otimes \ell' \} \in \mathcal{L} \mid \{ \otimes \ell \} \succeq_h \{ \otimes \ell' \} \text{ for some } \ell, \ell' \in \Delta(C) \}.
\end{align*}
\]

A multi-stage lottery \( l \in \mathcal{L}_P (\mathcal{L}_I) \) gives the DM a “worse (better)” lottery in the immediate future and a “better (worse)” lottery for the rest of the horizon. Since \( \{\succeq_h\}_{h \in H} \) satisfies History-Independent Risk Preference, \( \mathcal{L}_P \) and \( \mathcal{L}_I \) are independent of \( h \). Define

\[
Z_P \equiv \mathcal{K}(\mathcal{L}_P) \subset Z \text{ and } Z_I \equiv \mathcal{K}(\mathcal{L}_I) \subset Z.
\]

Definition 4.1. A DM is more patient at history \( h_1 \) than at history \( h_2 \) if the following two conditions hold:

(P1) For all \( x \in Z_P \) and \( l \in \mathcal{L}_I \),

\[
x \succ_{h_2} \{l\} \implies x \succ_{h_1} \{l\}.
\]

(P2) For all \( l \in \mathcal{L}_P \) and \( x \in Z_I \),

\[
\{l\} \succ_{h_2} x \implies \{l\} \succ_{h_1} x.
\]
(P1) states that if the DM at history $h_2$ strictly prefers a menu $x$ consisting of lotteries with later but larger rewards to a commitment lottery $f$ yielding the immediate but smaller reward, then so does she at history $h_1$. (P2) states that if the DM at history $h_2$ strictly prefers a commitment lottery $f$ yielding a later but larger reward to a menu $x$ consisting of lotteries with the immediate but smaller rewards, then so does she at history $h_1$.

We consider the implication of the behavioral comparison expressed by Definition 4.1. We show that Definition 4.1 characterizes the increasing convex and concave order on subjective beliefs.

**Definition 4.2.** Consider probability measures $\mu^1$ and $\mu^2$ over $[0,1]$. Say that $\mu^2$ is smaller than $\mu^1$ in the increasing convex [concave] order if, for all continuous, increasing and convex [concave] functions $v : [0,1] \rightarrow \mathbb{R}$,

$$\int_{[0,1]} v(\alpha) \, d\mu^2(\alpha) \leq \int_{[0,1]} v(\alpha) \, d\mu^1(\alpha).$$

One immediate observation is that $E_{\mu^2}[\alpha] \leq E_{\mu^1}[\alpha]$ if $\mu^2$ is smaller than $\mu^1$ in the increasing convex [concave] order because $v(\alpha) = \alpha$ is increasing and convex [concave]. Thus, these orders imply that the belief $\mu_1$ is more patient on average than the belief $\mu_2$.

The increasing convex and concave orders have a close relation to other stochastic orders used in economic analysis. A probability measure $\mu^1$ on $[0,1]$ is said to dominate $\mu^2$ in the first-order stochastic dominance if the inequality (5) holds for all increasing functions. Similarly, $\mu^1$ is said to dominate $\mu^2$ in the second-order stochastic dominance if the inequality (5) holds for all continuous concave functions. By definition, the domination according to the increasing convex [concave] order is a generalization of the first-order stochastic dominance. Moreover, if $\mu^1$ dominates $\mu^2$ in the second-order stochastic dominance, then $\mu^2$ is smaller (resp. larger) than $\mu^1$ in the increasing concave (resp. convex) order.

In the literature of stochastic orders, some equivalent conditions of Definition 4.2 are investigated. Let $F_i$ be the cumulative distribution function of a probability measure $\mu_i$ on $[0,1]$ and $F_i(\alpha) = 1 - F_i(\alpha)$ be its survival distribution function. From Shaked and Shanthikumar [22, p.182], the condition that $\mu_2$ is smaller than $\mu_1$ in the increasing convex order is equivalent to the following condition: for all $t \in [0,1]$,

$$\int_t^1 F_2(\alpha) \, d\alpha \leq \int_t^1 F_1(\alpha) \, d\alpha.$$ 

On the other hand, the condition that $\mu_2$ is smaller than $\mu_1$ in the increasing concave order is equivalent to the following: for all $t \in [0,1]$,

$$\int_0^t F_2(\alpha) \, d\alpha \geq \int_0^t F_1(\alpha) \, d\alpha.$$ 

---

8Notice that continuity is not redundant because a concave or convex function is continuous in the interior of the domain. In Hadar and Russell [13] and Rothschild and Stiglitz [20], continuity is not imposed.
Furthermore, from Shaked and Shanthikumar [22, p.197], if $E^1[\alpha] = E^2[\alpha]$, then the increasing convex [concave] order reduces to the convex [concave] order.\footnote{We say that $\mu^2$ is smaller than $\mu^1$ in the convex [concave] order if the inequality (5) holds for all continuous convex [concave] functions.}

The increasing concave order has been applied in economics. In choice under risk, the preferences of risk averse DMs can be expressed in terms of the increasing concave order. See Hadar and Russell [13], and Rothschild and Stiglitz [20]. Moreover, in Shorrocks [24], the increasing concave order is shown to be equivalent to the generalized Lorenz order which has been used in ranking income distributions; $\mu_2$ is smaller than $\mu_1$ in the generalized Lorenz order: for all $k \in [0,1],$

\[
\int_0^k F_2^{-1}(t) dt \leq \int_0^k F_1^{-1}(t) dt,
\]

where $F_i$ is the distribution function of $\mu_i$ and $F_i^{-1} = \sup \{ x \mid F_i(x) \leq k \}$.

As an illustration of the increasing concave order, following Ramos, Ollero, and Sordo [19], consider the Gamma distribution with density

\[
f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0,
\]

where $\lambda > 0$, $\alpha > 0$, and $\Gamma(\cdot)$ denotes the complete Gamma function. If $\lambda_1 > \lambda_2$ and $\frac{\alpha_1}{\lambda_1} \geq \frac{\alpha_2}{\lambda_2}$, $F_2$ is smaller than $F_1$ in the generalized Lorenz order, or equivalently in the increasing concave order. This example shows that unlike the first-order stochastic dominance, two distribution functions with single crossing property may be ranked by the increasing concave order.

We may now provide a characterization result.

**Theorem 4.1.** Assume that $\{x_{\cdot h}\}_{h \in H}$ satisfies all the axioms of Theorem 3.1 and admits a history-dependent random discounting representation $(u, \{\mu(\cdot|h)\}_{h \in H})$. Then the following conditions are equivalent:

(a) The DM is more patient at history $h_1$ than at history $h_2$.

(b) $\mu(\cdot|h_2)$ is smaller than $\mu(\cdot|h_1)$ in the increasing convex and concave orders.

A proof is relegated to Appendix G. The intuition behind (a) $\Rightarrow$ (b) in case of the increasing convex order is as follows: (P1) implies that $U(x|h_1) \geq U(x|h_2)$ for all $x \in Z_P$. Notice that for all $x \in Z_P$, the function

\[
\max_{l \in \mathcal{X}} \int_{C \times Z} \left( (1 - \alpha) u(c) + \alpha U(z|h_c) \right) dl(c, z) = \max_{l \in \mathcal{X}} \left( (1 - \alpha) u(\ell') + \alpha u(\ell'') \right),
\]

where $l = \ell' \otimes \{\otimes \ell''\}$, is increasing and convex in $\alpha$ because of $\{\otimes \ell''\} \succeq_h \{\otimes \ell'\}$. Hence the ranking $U(x|h_1) \geq U(x|h_2)$ means that the integral of an increasing convex function
of the form (6) with respect to $\mu_1$ is always bigger than that corresponding to $\mu_2$. As the last step, we show that any continuous, increasing, and convex function $v$ on $[0, 1]$ can be arbitrarily approximated by a function of the form (6) if an affine transformation of $u$ is chosen appropriately.

The following are some examples of Theorem 4.1.

**Example 1** (Fisher (1930)). Let $C$ be a compact interval of $\mathbb{R}_+$. Suppose that $\{\succeq_h\}_{h \in H}$ is an average-dependent preference with a discount factor $\delta \in (0, 1)$. Define the average consumption of history $h = (\cdots, c_2, c_1)$ by

$$\bar{c}_h = (1 - \delta)^{\sum_{t=1}^{\infty} \delta^t c_t}.$$ 

Fisher’s discussion corresponds to the case that the DM is more patient at $h$ than at history $h'$ if $\bar{c}_h < \bar{c}_{h'}$. That is, the DM tends to be more patient if the average consumption up to that time period is lower maybe because of the habit of thrift.

**Example 2** (Becker and Mulligan (1997)). Let $C = A \times B$, where $A \subset \mathbb{R}_+$ is a set of regular consumptions and $B \subset \mathbb{R}_+$ is a set of investment actions. Their assumption corresponds to the case that the DM is more patient at $h = (\cdots, (a_1, b_1))$ than at history $h' = (\cdots, (a'_1, b'_1))$ if $b_1 > b'_1$.

In the history-independent random discounting model, Higashi, Hyogo, and Takeoka [14] provide an inter-personal comparison toward flexibility which is associated with the second-order stochastic dominance on subjective beliefs over discount factors. Their condition is stronger than Definition 4.1 (P1): agent 1 is said to be more averse to commitment than agent 2 if for all $x \in K(\mathcal{L})$ and $l \in \mathcal{L}$,

$$x >^2 \{l\} \implies x >^1 \{l\}. \quad (7)$$

Unlike Definition 4.1, this comparison is not restricted to the trade-offs between sooner but smaller rewards and later but larger rewards. Condition (7) is shown to be equivalent to the condition that $\mu^2$ dominates $\mu^1$ in the second-order stochastic dominance, which is in turn equivalent to saying that $\mu^2$ is smaller than $\mu^1$ in the convex order. Theorem 4.1 shows that (P1) is equivalent to the condition that $\mu(\cdot|h_2)$ is smaller than $\mu(\cdot|h_1)$ in the increasing convex order.

The notion of comparative impatience has been already examined in a different framework from the current paper. Benoit and Ok [3] provide a comparison of the attitude towards time delay under the assumptions that agents have additively separable utility functions over streams of deterministic outcomes and that their discount functions are not necessarily exponential. Consider two agents, agent 1 and agent 2, facing the same endowment stream. In Benoit and Ok [3], agent 1 is more delay averse than agent 2 if agent 2 prefers receiving small additional consumption at an earlier period to large additional consumption at a later period, so does agent 1.
Finally, one might wonder what is the gap between Theorem 4.1 and characterizing the first-order stochastic dominance on subjective beliefs. A difficulty of behavioral characterization of the first order stochastic dominance arises from the fact that an increasing continuous function may not be approximated by the sum of an increasing convex and an increasing concave function. See Thon and Thorlund-Petersen [25, Lemma 1] for details.

Appendices

A Hausdorff Metric

Let $X$ be a compact metric space with a metric $d$. Let $\mathcal{K}(X)$ be the set of all non-empty compact subsets of $X$. For $x \in X$ and $A, B \in \mathcal{K}(X)$, let

$$d(x, B) \equiv \min_{x' \in B} d(x, x'), \quad d(A, B) \equiv \max_{x \in A} d(x, B).$$

For all $A, B \in \mathcal{K}(X)$, define the Hausdorff metric $d_H$ by

$$d_H(A, B) \equiv \max\{d(A, B), d(B, A)\}.$$

Then, $d_H$ satisfies (i) $d_H(A, B) \geq 0$, (ii) $A = B \iff d_H(A, B) = 0$, (iii) $d_H(A, B) = d_H(B, A)$, and (iv) $d_H(A, B) \leq d_H(A, C) + d_H(C, B)$. Moreover, $\mathcal{K}(X)$ is compact under the Hausdorff metric.

B Some Implications of Period-Wise Dominance

B.1 Dominance Operator for Finite Menus

Fix any history $h$. If $\succsim_h$ is restricted on iid lotteries, it induces a preference over $\Delta(C)$, denoted by $\succsim^*$. By History-Independent Risk Preference, $\succsim^*$ is independent of $h$. By Order, Continuity, Independence, and IID-separability, $\succsim^*$ satisfies all the VNM axioms. Thus, there exists a continuous mixture linear representation $v : \Delta(C) \to \mathbb{R}$. Let $\ell^+$ and $\ell^-$ be a maximal and a minimal lottery with respect to $v$. We can assume that $v(\ell^-) = 0$.

Take any $c \in C$ and $x \in Z$. By Best-Worst, $\otimes \ell^+ \succsim_{hc} x \succsim_{hc} \otimes \ell^-$. By Continuity, Independence and IID-Separability, there exists a unique $\lambda(x, hc) \in [0, 1]$ such that

$$x \sim_{hc} \{\otimes(\lambda(x, hc)\ell^+ + (1 - \lambda(x, hc))\ell^-)\}. \quad (8)$$

Hence, all $x \in Z$ admits a smoothing equivalent at history $hc$.

Now we can define a dominance operation for all $l \in \Delta_s(C \times Z)$ and all finite subsets $x \subset \Delta_s(C \times Z)$ as follows:

$$D_h^*(l) \equiv \{l' \in \Delta_s(C \times Z) | l \text{ dominates } l' \text{ at history } h\},$$

$$D_h^*(x) \equiv \cup_{l \in x} d(D_h^*(l)).$$

19
where \( \text{cl}(\cdot) \) is the closure operation in \( \Delta(C \times Z) \). That is, \( D_h^s(x) \) is the set of all lotteries dominated by some lottery in \( x \) at history \( h \) (and their accumulation points). Since \( D_h^s(x) \) is a nonempty compact set in \( \Delta(C \times Z) \), it is a well-defined choice object.

The following lemma shows that the DM does not care about dominated lotteries. Since \( x \subset D_h^s(x) \) when \( \succeq_h \) satisfies Order, the DM having preference for flexibility weakly prefers \( D_h^s(x) \) to \( x \). Thus, the lemma is a counterpoint to Monotonicity, and shows that it is not useful to keep dominated lotteries (and their accumulation points) within the menu, that is, \( x \succeq_h D_h^s(x) \).

**Lemma B.1.** For all \( h \in H \) and all finite menus \( x \subset \Delta_s(C \times Z) \), \( x \sim_h D_h^s(x) \).

**Proof.** First, take any menu \( x \subset Z \) such that there exists \( l \in Z \) with finite support. We will claim that \( x \cup \text{cl}(D_h^s(l)) \sim_h x \). As a first step, take any finite menu \( y \subset D_h^s(l) \). By applying Period-Wise Dominance finite times, we have \( x \cup y \sim_h x \). Next, since \( \text{cl}(D_h^s(l)) \) is compact, by Lemma 0 of Gul and Pesendorfer [10, p.1421], there exists a sequence \( \tilde{y}^n \subset \Delta(C \times Z) \) such that each \( \tilde{y}^n \) is finite with \( \tilde{y}^n \subset \text{cl}(D_h^s(l)) \) and \( \tilde{y}^n \to \text{cl}(D_h^s(l)) \). Since \( \tilde{y}^n \) is finite, it can be denoted by \( \tilde{y}^n = \{l_{1}^n, \ldots, l_{K_n}^n\} \). By definition of closure, we can find \( y^n = \{l_{1}^n, \ldots, l_{K_n}^n\} \subset D_h^s(l) \) such that \( d(l_{i}^n, l_{i}^n) < \frac{1}{n} \) for all \( i = 1, \ldots, K_n \). By definition of Hausdorff metric \( d_H, d_H(y^n, \tilde{y}^n) \to 0 \) as \( n \to 0 \). By the triangle inequality,

\[
d_H(y^n, \text{cl}(D_h^s(l))) \leq d_H(y^n, \tilde{y}^n) + d_H(\tilde{y}^n, \text{cl}(D_h^s(l))).
\]

Thus, \( d_H(y^n, \text{cl}(D_h^s(l))) \to 0 \) as \( n \to 0 \). Since the first step implies \( x \cup y^n \sim_h x \), by Continuity, \( x \cup \text{cl}(D_h^s(l)) \sim_h x \) as desired.

Now take any finite menu \( x \subset \Delta_s(C \times Z) \). By applying the above claim finite times,

\[
D_h^s(x) = \cup_{l \in x} \text{cl}(D_h^s(l)) = x \cup (\cup_{l \in x} \text{cl}(D_h^s(l))) \sim_h x.
\]

\( \square \)

### B.2 Extension of Dominance Operator to General Menus

We consider a dominance operator for a general menu \( x \in Z \) as we now describe below. Fix any history \( h \in H \). Recall a smoothing equivalent of \( x \) as given by (8). Define \( f_h : C \times Z \to \Delta(C) \) by

\[
f_h(c, z) \equiv \lambda(x, hc)\ell^+ + (1 - \lambda(x, hc))\ell^-
\]

and \( F_h : \Delta(C \times Z) \to \Delta(\Delta(C)) \) by

\[
F_h(l) \equiv l \circ f_h^{-1},
\]

that is, \( F_h(l)(G) = l(f_h^{-1}(G)) \) for all Borel subsets \( G \) of \( \Delta(C) \). Lemma B.2 (i) and (ii) below show that \( f_h \) is continuous, and \( F_h \) is continuous and mixture linear.
Finally, let $E : \Delta(\Delta(C)) \rightarrow \Delta(C)$ be the reduction operator, defined as the function assigning $\eta \in \Delta(\Delta(C))$ with its reduced lottery in $\Delta(C)$, that is, $E(\eta) \in \Delta(C)$ satisfies
\[
\int_C g(c) \, dE(\eta)(c) = \int_{\Delta(C)} g(c) \, d\ell(c) \, d\eta(\ell)
\]
for all continuous functions $g : C \rightarrow \mathbb{R}$. For example, if $\eta = \sum_i \lambda_i \ell_i \in \Delta(\Delta(C))$ denotes the lottery yielding $\ell_i$ with probability $\lambda_i$, then $E(\eta) = \sum_i \lambda_i \ell_i \in \Delta(C)$. See Lemma B.2 (iii) and (iv) for properties of $E$.

For all $h$ and $l \in \Delta(C \times Z)$, define $l_h \equiv E(F_h(l)) \in \Delta(C)$.

**Definition B.1.** For all $h$ and $l, l' \in \Delta(C \times Z)$, say that $l$ dominates $l'$ at history $h$ if \{\otimes_l \geq_h \otimes_{l'}\} and \{\otimes_{l_h} \geq_h \otimes_{l'_h}\}.

For all $h \in H, l \in \Delta(C \times Z)$, and $x \in Z$, let
\[
D_h(l) \equiv \{l' \in \Delta(C \times Z) \mid l \text{ dominates } l' \text{ at history } h\},
\]
\[
D_h(x) \equiv \cup_{l \in x} D_h(l).
\]
That is, $D_h(x)$ is the set of all lotteries dominated by some lottery in $x$ at history $h$. As shown in Lemma B.2 (viii), $D_h(x)$ coincides with $D^*_h(x)$ if $x \subset \Delta_s(C \times Z)$ is finite. Thus, $D_h$ is regarded as an extension of the dominance operator to general menus.

**B.3 Properties of Dominance Operator**

**Lemma B.2.** Fix an arbitrary $h \in H$.

(i) $f_h : C \times Z \rightarrow \Delta(C)$ is continuous.

(ii) $F_h : \Delta(C \times Z) \rightarrow \Delta(\Delta(C))$ is continuous and mixture linear.

(iii) $E : \Delta(\Delta(C)) \rightarrow \Delta(C)$ is well-defined, continuous, and mixture linear.

(iv) For all $\ell_i \in \Delta(C)$ and $\lambda_i \in [0, 1], i = 1, \ldots, n$, with $\sum_i \lambda_i = 1$, $E(\sum_i \lambda_i \ell_i) = \sum_i \lambda_i \ell_i$.

(v) For all $x \in Z$, $D_h(x) \in Z$.

(vi) For all $x \in Z$, if $x$ is convex, so is $D_h(x)$.

(vii) $D_h : Z \rightarrow Z$ is Hausdorff continuous.

(viii) For all finite menus $x \subset \Delta_s(C \times Z)$, $D_h(x) = D^*_h(x)$.

(ix) For all $x \in Z$, $x \sim_h D_h(x)$.
Proof. (i) Let \((e^n, z^n) \to (c, z)\). We want to show that \(f_h(e^n, z^n) \to f_h(c, z)\). Suppose otherwise. Then, there exists an open neighborhood \(O\) of \(f_h(c, z)\) such that \(f_h(e^n, z^n) \notin O\) for infinitely many \(n\). Let \(\{f_h(e^n, z^n)\}\) be a subsequence satisfying \(f_h(e^n, z^n) \notin O\). Since \(\Delta(C)\) is compact, we can assume that \(f_h(e^n, z^n) \to \ell\) for some \(\ell \in \Delta(C)\). Since \(z^n \sim_{hc} (e^n, z^n)\), Dynamic Consistency implies that \(\{(e^n, z^n)\} \sim_h \{(e^n, \otimes f_h(e^n, z^n))\}\). By Continuity, \(\{(c, z)\} \sim_h \{(c, \otimes \ell)\}\). Again, by Dynamic Consistency, \(z \sim_{hc} \{\otimes \ell\}\). Since \(\lambda(z, hc)\) is unique, we have \(f_h(c, z) = \ell\), which contradicts the fact that \(\ell \notin O\).

(ii) From (i), \(f_h\) is continuous. Continuity of \(F_h\) follows from Billingsley [4, p.20]. To show that \(F_h\) is mixture linear, take any lotteries \(l, l' \in \Delta(C \times \mathcal{Z})\) and \(\lambda \in [0, 1]\). Define, for all Borel subsets \(G\) of \(\Delta(C)\),
\[
F_h(\lambda l + (1 - \lambda) l')(G) = (\lambda l + (1 - \lambda) l') (f_h^{-1}(G)) = \lambda l (f_h^{-1}(G)) + (1 - \lambda) l' (f_h^{-1}(G)) = \lambda F_h(l)(G) + (1 - \lambda) F_h(l')(G).
\]

(iii) For all \(\eta \in \Delta(\Delta(C))\) and real-valued continuous functions \(g\) on \(C\), define
\[
T(g) \equiv \int_{\Delta(C)} \int_C g \, d\ell \, d\eta.
\]
Since \(T\) is a bounded linear functional on the set of all real-valued continuous functions on \(C\), the Riesz representation theorem ensures that there exists a unique \(E(\eta) \in \mathcal{L}\) such that
\[
T(g) = \int_C g \, dE(\eta).
\]
Hence \(E\) is well-defined.

To show continuity, let \(\eta^n \to \eta\). We want to show that \(E(\eta^n) \to E(\eta)\). Suppose otherwise. Then, there exists an open neighborhood \(O\) of \(E(\eta)\) such that \(E(\eta^n) \notin O\) for infinitely many \(n\). Let \(\{E(\eta^m)\}\) be a subsequence satisfying \(E(\eta^m) \notin O\) for all \(m\). Since \(\Delta(C)\) is compact, we can assume that \(\{E(\eta^m)\}\) converges to a limit \(\ell^* \in \Delta(C)\). By definition, for all real-valued continuous functions \(g\) on \(C\),
\[
\int_C g \, dE(\eta^m) = \int_{\Delta(C)} \int_C g \, d\ell \, d\eta^m.
\]
As \(m \to \infty\), we have
\[
\int_C g \, d\ell^* = \int_{\Delta(C)} \int_C g \, d\ell \, d\eta.
\]
Since \(E\) is well-defined, \(E(\eta) = \ell^*\), which contradicts the fact that \(\ell^* \notin O\).

Finally, we show mixture linearity of \(E\). Take any \(\eta, \eta' \in \Delta(\Delta(C))\) and \(\lambda \in [0, 1]\). Define,
\[
\int_C g \, dE(\lambda \eta + (1 - \lambda) \eta') = \int_{\Delta(C)} \int_C g \, d\ell \, d(\lambda \eta + (1 - \lambda) \eta')(\ell).
\]
Thus, for all continuous functions $g$ on $C$,
\[
\int_C g \, dE(\lambda \eta + (1 - \lambda) \eta') = \int_C g \, d(\lambda E(\eta) + (1 - \lambda) E(\eta')).
\]

Since $E$ is well-defined, $E(\lambda \eta + (1 - \lambda) \eta') = \lambda E(\eta) + (1 - \lambda) E(\eta')$.

(iv) For all continuous functions $g$ on $C$,
\[
\int_C g \, dE(\sum_i \lambda_i \circ \ell_i) = \int_{\Delta(C)} \int_C g \, d(\sum_i \lambda_i \circ \ell_i) = \sum_i \lambda_i \int_C g \, d\ell_i = \int_C g \, d(\sum_i \lambda_i \ell_i).
\]
Thus, $E(\sum_i \lambda_i \circ \ell_i) = \sum_i \lambda_i \ell_i$.

(v) Since $\Delta(C \times Z)$ is compact, it is enough to show that $D_h(x)$ is closed. Take any sequence $l^n \in D_h(x)$ converging to a limit $l \in \Delta(C \times Z)$. There exists $\overline{l}^n \in x$ such that
\[
\{ \otimes l^n_c \} \succ_h \{ \otimes l^n_c \} \quad \text{and} \quad \{ \otimes E(F_h(l^n)) \} \succ_h \{ \otimes E(F_h(\overline{l})) \}.
\]

Since $x$ is compact, without loss of generality we assume that $\overline{l}^n \to \overline{l}$ for some limit $\overline{l} \in x$. Since $\otimes l^n_c \to \otimes l_c$, $\otimes \overline{l}^n_c \to \otimes \overline{l}_c$ and since $F_h$ and $E$ are continuous, the Continuity axiom implies that
\[
\{ \otimes \overline{l}_c \} \succ_h \{ \otimes l_c \} \quad \text{and} \quad \{ \otimes E(F_h(\overline{l})) \} \succ_h \{ \otimes E(F_h(l)) \}.
\]
Thus $\overline{l} \in x$ dominates $l$ at history $h$ and, hence, $l \in D_h(x)$.

(vi) Take any two points $l, l' \in D_h(x)$ and $\lambda \in [0,1]$. There exist $l, l' \in x$ such that $l$ dominates $l$, and $l'$ dominates $l'$. Let $l^\lambda \equiv \lambda l + (1 - \lambda) l'$ and $\overline{l}^\lambda \equiv \lambda \overline{l} + (1 - \lambda) \overline{l}'$. Since $x$ is convex, $\overline{l}^\lambda \in x$. We show that $\overline{l}^\lambda$ dominates $l^\lambda$, which results in $l^\lambda \in D_h(x)$. Since $\{ \otimes l_c \} \succ_h \{ \otimes l_c \}$ and $\{ \otimes \overline{l}_c \} \succ_h \{ \otimes \overline{l}_c \}$, Independence and IID-Separability imply that
\[
\{ \otimes \overline{l}_c \} \sim_h \{ \lambda \otimes l_c + (1 - \lambda) \otimes \overline{l}_c \} \succ_h \{ \lambda \otimes l_c + (1 - \lambda) \otimes l'_c \} \sim_h \{ \otimes l^\lambda \}.
\]
Moreover, since $F_h$ and $E$ are mixture linear, Independence and IID-Separability imply that
\[
\{ \otimes E(F_h(l^\lambda)) \} = \{ \otimes (\lambda E(F_h(\overline{l})) + (1 - \lambda) E(F_h(\overline{l}'))) \}
\sim_h \{ \lambda \otimes E(F_h(l)) + (1 - \lambda) \otimes E(F_h(l')) \}
\succ_h \{ \lambda \otimes E(F_h(l)) + (1 - \lambda) \otimes E(F_h(l')) \}
\sim_h \{ \otimes (\lambda E(F_h(l)) + (1 - \lambda) E(F_h(l')) \} = \{ \otimes E(F_h(l^\lambda)) \}.
\]
Thus \(l^k\) dominates \(l^\lambda\).

(vii) Let \(x^n \to x\). We want to show \(D_h(x^n) \to D_h(x)\). By contradiction, suppose otherwise. Then, there exists a neighborhood \(U\) of \(D_h(x)\) such that \(D_h(x^n) \notin U\) for infinitely many \(n\). Let \(\{x^n\}_{n=1}^{\infty}\) be the corresponding subsequence of \(\{x^n\}_{n=1}^{\infty}\). Since \(x^n \to x\), \(\{x^n\}_{n=1}^{\infty}\) also converges to \(x\). Since \(\{D_h(x^n)\}_{n=1}^{\infty}\) is a sequence in a compact metric space \(Z\), there exists a convergent subsequence \(\{D_h(x^{m_n})\}_{m=1}^{\infty}\) with a limit \(y \neq D_h(x)\). As a result, now we have \(x^{m_n} \to x\) and \(D_h(x^{m_n}) \to y\). In the following argument, we will show that \(y = D_h(x)\), which is a contradiction.

**Step 1:** \(D_h(x) \subset y\).

Take any \(l \in D_h(x)\). Then, there exists \(\bar{l} \in x\) such that \(\{\otimes l\} \triangleright_h \{\otimes l\} \ni \{\otimes l\}.\) Since \(x^n \to x\), we can find a sequence \(\{\otimes l\}_{m=1}^{\infty}\) such that \(\otimes l \in x^n\) and \(\otimes l \to l\).

Now we will construct a sequence \(\{\otimes l\}_{m=1}^{\infty}\) with \(\otimes l \in D_h(x^n)\) satisfying \(\otimes l \to l\). For all sufficiently large \(k\), let \(B_{1/k}(l)\) be the \(1/k\)-neighborhood of \(l\) with respect to the weak convergence topology. There exists \(0 < \lambda^k < 1\) such that \(l^k \equiv \lambda^k l + (1 - \lambda^k) \otimes e \in B_{1/k}(l)\).

By construction, \(l^k \to l\).

By Independence and IID-Separability, \(\{\otimes l\} \triangleright_h \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) In the case of former, since \(\otimes l \to l\), by Continuity, there exists \(m_1^k\) such that for all \(m \geq m_1^k\), \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) In the case of latter, for all \(m\), \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) In both cases, we have \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) On the other hand, by Lemma B.2 (ii) and (iii), \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) Similarly, since \(\otimes l \to l\), by Lemma B.2 (ii) and (iii), there exists \(m_2^k\) such that for all \(m \geq m_2^k\), \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\)

Therefore, for all \(m \geq m_2^k \equiv \max\{m_1^k, m_2^k\}\),

\[
\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\},
\]

that is, \(l^k \in D_h(x^n) \subset D_h(x^n)\) for all \(m \geq m_2^k\). Since we have \(m_2^k+1 \geq m_2^k\) for all \(k\), define \(l^m \equiv l^k\) for all \(m\) satisfying \(m \leq m_2^k\). Then, \(\{l_m\}_{m=1}^{\infty}\) is a required sequence.

Since \(l \to l\) and \(D_h(x^n) \to y\) with \(l_m \in D_h(x^n)\), we have \(l \in y\). Thus, \(D_h(x) \subset y\).

**Step 2:** \(y \subset D_h(x)\).

Take any \(l \in y\). Since \(D_h(x^n) \to y\), we can find a sequence \(l_m \in D_h(x^n)\) with \(l_m \to l\).

By definition, there exists \(l_m \in x^n\) such that \(\{\otimes l_m\} \ni \{\otimes l_m\} \ni \{\otimes l_m\} \ni \{\otimes l_m\}.\) Since \(\Delta(C \times Z)\) is compact, we can assume that \(\{l_m\}\) converges to a limit \(l \in \Delta(C \times Z)\).

Since \(l \to l\) and \(x^n \to x\) with \(l_m \in x^n\), we have \(l \in x^n\). From Continuity and Lemma B.2 (ii) and (iii), \(\{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\} \ni \{\otimes l\}.\) Thus, \(l \in D_h(x)\).

(viii) It suffices to show that for all \(l \in \Delta(C \times Z)\), \(D_h(l) = cl(D_h(l))\) because then

\[
D_h(l) = \cup_{l \in \Delta(C \times Z)} cl(D_h(l)) = \cup_{l \in \Delta(C \times Z)} D_h(l) = D_h(x).
\]
Since $\otimes f_h(c_i, x_i)$ is also a smoothing equivalent of $x_i$ at history $h c_i$, $\{\otimes \ell_i\} \sim_h \{\otimes f_h(c_i, x_i)\}$. By Independence and IID-Separability, $\{\otimes(\sum_i \lambda_i \ell_i)\} \sim_h \{\otimes(\sum_i \lambda_i f_h(c_i, x_i))\}$. Thus, by part (iv),

$$\{\otimes(\sum_i \lambda_i \ell_i)\} \sim_h \{\otimes(\sum_i \lambda_i f_h(c_i, x_i))\} = \{\otimes E(\sum_i \lambda_i \circ f_h(c_i, x_i))\} = \{\otimes E(F_h(l))\}.$$

**Step 2:** $\text{cl}(D^*_h(l)) \subset D_h(l)$.

Take any $l' \in D^*_h(l)$. Since $l$ has a finite support, it can be rewritten as $l = ((c_i, x_i); \lambda_i)^n_{i=1}$. By definition, there exists a smoothing equivalent $\otimes \ell_i$ of $x_i$ at history $h c_i$. Similarly, $l'$ is written as $l' = ((c'_i, x'_i); \lambda'_i)^m_{i=1}$, and let $\otimes \ell'_i$ be a smoothing equivalent of $x'_i$ at history $h c'_i$. Then, by Step 1, $\{\otimes(\sum_i \lambda_i \ell_i)\} \supset_h \{\otimes(\sum_i \lambda'_i \ell'_i)\}$ is equivalent to $\{\otimes E(F_h(l'))\} \supset_h \{\otimes E(F_h(l))\}$, and hence $l' \in D_h(l)$. Since $D_h(l)$ is closed as shown in part (v), $\text{cl}(D^*_h(l)) \subset D_h(l)$ as desired.

**Step 3:** $D_h(l) \subset \text{cl}(D^*_h(l))$.

As a preliminary, we first show the following:

**Step 3-1:** For all $l \in \Delta(C \times Z)$, there exists a sequence $l^n \in \Delta_s(C \times Z)$ converging to $l$ such that the support of $l^n$ is included in that of $l$.

Let $E$ denote the support of $l$. Since $C \times Z$ is compact, $E$ is also compact. For all $n$, there exists a finite partition $\{E^n_i\}_{i=1}^{M_n}$ of $E$ such that the diameter of $E^n_i$ is smaller than $1/n$. Take $(c, x)^n_i \in E^n_i$ arbitrarily. Let $l^n$ be the probability measure with probability $l(E^n_i)$ at the point $(c, x)^n_i$. Take any uniformly continuous function $g : C \times Z \to \mathbb{R}$. Let $\alpha^n_i = \inf_{E^n_i} g$ and $\beta^n_i = \sup_{E^n_i} g$. Since $g$ is uniformly continuous and since the diameter of $E^n_i$ converges to zero as $n \to \infty$ uniformly in $i$, $\sup_i (\beta^n_i - \alpha^n_i) \to 0$ as $n \to \infty$. Thus, as $n \to \infty$,

$$\left| \int g \, dl^n - \int g \, dl \right| = \left| \sum_i \int_{E^n_i} (g - g((c, x)^n_i)) \, dl \right| \leq \sup_i (\beta^n_i - \alpha^n_i) \to 0.$$

Hence, $l^n \to l$ as desired.

To show that $D_h(l) \subset \text{cl}(D^*_h(l))$, take any $l' \in D_h(l)$. By definition, $\{\otimes \ell_c\} \supset_h \{\otimes \ell'_c\}$ and $\{\otimes \ell_h\} \supset_h \{\otimes \ell'_h\}$. Define $\ell^n = \frac{n-1}{n} l' + \frac{1}{n} \otimes \ell_h$ for each $n$. By Step 3-1, we can find $l^n \in \Delta_s(C \times Z)$ such that $d(l^n, l^n) < \frac{1}{n}$. Then, $l^n \to l'$ because $d(l^n, l') \leq d(l^n, l^n) + d(l^n, l')$ and $d(l^n, l^n) \to 0$ and $d(l^n, l') \to 0$.

To show that $\{\otimes \ell_c\} \supset_h \{\otimes l^n_c\}$, we consider several cases as below.

**Case 1:** $\{\otimes \ell_c\} \supset_h \{\otimes l^n_c\}$. Since $l^n \to l'$, by Continuity, we have $\{\otimes \ell_c\} \supset_h \{\otimes l^n_c\}$.

**Case 2:** $\{\otimes l^n_c\} \supset_h \{\otimes \ell_l\}$. We have $\ell^n_c = \frac{n-1}{n} l^n_c + \frac{1}{n} \ell_l$. By Independence and IID-Separability, $\{\otimes l^n_c\} \supset_h \{\otimes l^n_c\}$. Thus, by Continuity, $\{\otimes \ell_c\} \supset_h \{\otimes l^n_c\}$. 

25
Case 3: \( \{ \otimes l_c \} \sim_h \{ \otimes l' \} \sim_h \{ \otimes \ell^- \} \). By assumption, \( v(l'_c) = v(\ell^-) = 0 \). Let \( f_c : C \times Z \to C \) denote the projection mapping on \( C \). Since

\[
0 = v(l'_c) = \int v(c) \, dl'_c = \int_C v(c) \, d(l' \circ f_c^{-1})(c)
\]

\[
= \int_{C \times Z} v(f_c(c, x)) \, dl'(c, x) = \int_{C \times Z} v(c) \, dl'(c, x),
\]

by Theorem 11.16 (3) of Aliprantis and Border [1, p.412], \( v(c) = 0 \) almost surely with respect to \( l' \). That is, for all \( (c, x) \) within the support of \( l' \), \( \{ \otimes c \} \sim_h \{ \otimes \ell^- \} \). Since \( l^n \) has a finite support, \( l^n \) is written as \( l^n = ((c^n_i, x^n_i), \lambda^n)_{i=1}^n \). By Independence and IID-Separability,

\[
\{ \otimes l^n \} = \{ \otimes \sum_i \lambda_i f_c(c^n_i, x^n_i) \} \sim_h \{ \otimes \ell^- \}.
\]

Hence, for all \( n, \{ \otimes l_c \} \sim_h \{ \otimes l'_n \} \).

Next, we show \( \{ \otimes l_h \} \sim_h \{ \otimes l^n \} \) by a case-by-case argument.

Case 1: \( \{ \otimes l_h \} \succ_h \{ \otimes l'_h \} \). Since \( l^n \to l' \), by the Continuity axiom and continuity of \( E \) and \( F_h \), we have \( \{ \otimes l_h \} \succ_h \{ \otimes l^n \} \).

Case 2: \( \{ \otimes l'_h \} \succ_h \{ \otimes \ell^- \} \). Since \( E \) and \( F_h \) are mixture linear, \( \delta^n_h = E(F_h(l^n)) = \frac{n-1}{n} l'_h + \frac{1}{n} \ell^- \). By Independence and IID-Separability, \( \{ \otimes l'_h \} \succ_h \{ \otimes l^n \} \). Thus, by the Continuity axiom and continuity of \( E \) and \( F_h \), \( \{ \otimes l_h \} \succ_h \{ \otimes l^n \} \).

Case 3: \( \{ \otimes l_h \} \sim_h \{ \otimes l'_h \} \sim_h \{ \otimes \ell^- \} \). By assumption, \( v(l'_h) = v(\ell^-) = 0 \). Since

\[
0 = v(l'_h) = \int v(c) \, dE(F_h(l'_c)) = \int_{\Delta(C)} \int_C v(c) \, d\ell \, d(l' \circ f_h^{-1})(\ell)
\]

\[
= \int_{C \times Z} \int_C v(c) \, df_h(c, x) \, dl'(c, x) = \int_{C \times Z} v(f_h(c, x)) \, dl'(c, x),
\]

by Theorem 11.16 (3) of Aliprantis and Border [1, p.412], \( v(f_h(c, x)) = 0 \) almost surely with respect to \( l' \). That is, for all \( (c, x) \) within the support of \( l' \), \( \{ \otimes f_h(c, x) \} \sim_h \{ \otimes \ell^- \} \). Since \( l^n \) has a finite support, \( l^n \) is written as \( l^n = ((c^n_i, x^n_i), \lambda^n)_{i=1}^n \). By Independence and IID-Separability,

\[
\{ \otimes l^n \} = \{ \otimes \sum_i \lambda_i f_h(c^n_i, x^n_i) \} \sim_h \{ \otimes \ell^- \}.
\]

Hence, for all \( n, \{ \otimes l_h \} \sim_h \{ \otimes l^n \} \).

Therefore, \( l^n \in D^\ast_h(l) \), and hence \( l' \in cl(D^\ast_h(l)) \).

(ix) First we will claim that for all \( x \in Z \), there exists a sequence \( x^n \subset \Delta(C \times Z) \) such that \( x^n \) is finite and \( x^n \to x \). By Lemma 0 of Gul and Pesendorfer [10, p.1421], there exists a sequence \( \hat{x}^n \subset \Delta(C \times Z) \) such that each \( \hat{x}^n \) is finite and \( \hat{x}^n \to x \). Since \( \hat{x}^n \) is finite, it can be denoted by \( \hat{x}^n = \{ l^n_1, \ldots, l^n_{K_n} \} \). Since the set of lotteries with finite supports is dense in \( \Delta(C \times Z) \) (see Aliprantis and Border [1, p.513, Theorem 15.10]), we can find
Thus, $d_H(x^n, x) \to 0$ as $n \to 0$.

Now we show that for all $x \in Z$, $x \sim_h D_h(x)$. Take any $x \in Z$. By the above claim, there exists a sequence $x^n \subset \Delta_v(C \times Z)$ such that $x^n$ is finite and $x^n \to x$. By Lemma B.1 and property (viii), $D_h(x^n) = D_h^n(x^n) \sim_h x^n$. By the Continuity axiom and continuity of $D_h, D_h(x) \sim_h x$ as desired. □

C Preliminaries

Lemma C.1. Let $X$ be a compact metric space. Let $f^n, f : X \to \mathbb{R}$ be continuous functions and $p^n, p$ be Borel probability measures on $X$. If $f^n \to f$ with respect to the sup-norm and $p^n \to p$ with respect to the weak convergence topology, then $\int f^n \, dp^n \to \int f \, dp$.

Proof. For all $n$, we have

$$
\left| \int f^n \, dp^n - \int f \, dp \right| \leq \left| \int f^n \, dp^n - \int f \, dp^n \right| + \left| \int f \, dp^n - \int f \, dp \right|
$$

$$
\leq \left| f^n - f \right| \, dp^n + \left| \int f \, dp^n - \int f \, dp \right|
$$

$$
\leq \sup_{x \in X} |f^n(x) - f(x)| + \left| \int f \, dp^n - \int f \, dp \right|. \quad (11)
$$

By assumption, (11) converges to zero as $n \to \infty$. □

Lemma C.2. Let $L$ be a compact metric space and $f : L \to \mathbb{R}^N$ be a continuous function. For all $A \in \mathcal{K}(L)$, define $f(A) \equiv \{f(l) \in \mathbb{R}^N \mid l \in A\}$. If $A^n \to A$ with $A^n, A \in \mathcal{K}(L)$, then $f(A^n) \to f(A)$ with respect to the Hausdorff topology.

Proof. First of all, since $L$ is compact and $f$ is continuous, $f(L)$ is a compact subset of $\mathbb{R}^N$. Since $f(A)$ is a compact subset of $f(L)$, we have $f(A) \in \mathcal{K}(f(L))$ for all $A \in \mathcal{K}(L)$.

To show that $f(A^n) \to f(A)$ with respect to the Hausdorff metric, by contradiction, suppose otherwise. Then, there exists a neighborhood $\mathcal{O}$ of $f(A)$ such that $f(A^m) \notin \mathcal{O}$ for infinitely many $m$. Let $\{A^m\}_{m=1}^{\infty}$ be the corresponding subsequence of $\{A^n\}_{n=1}^{\infty}$. Since $A^n \to A$, $\{A^m\}_{m=1}^{\infty}$ also converges to $A$. Since $\{f(A^m)\}_{m=1}^{\infty}$ is a sequence in a compact metric space $\mathcal{K}(f(L))$, there exists a convergent subsequence $\{f(A^\ell)\}_{\ell=1}^{\infty}$ with a limit $V \neq f(A)$. As a result, now we have $A^\ell \to A$ and $f(A^\ell) \to V$.

In the following argument, we will show that $f(A) = V$, which is a contradiction. To show $f(A) \subset V$, take any $a \in f(A)$. There exists $l \in A$ such that $a = f(l)$. Since $A^\ell \to A$, we can find $\{l^\ell\}_{\ell=1}^{\infty}$ such that $l^\ell \to l$ with $l^\ell \in A^\ell$. Let $a^\ell \equiv f(l^\ell)$. Since $a^\ell \to a$ and $f(A^\ell) \to V$ with $a^\ell \in f(A^\ell)$, we have $a \in V$. Thus, $f(A) \subset V$. 

27
To show $V \subset f(A)$, take any $a \in V$. Since $f(A^t) \to V$, we can find $\{a^t\}_{t=1}^\infty$ such that $a^t \to a$ with $a^t \in f(A^t)$. There exists $l^t \in A^t$ satisfying $a^t = f(l^t)$. Since $L$ is compact, there exists a convergent subsequence $\{l^k\}_{k=1}^\infty$ with a limit $l$. By continuity of $f$, $f(l) = a$. Moreover, since $l^k \to l$, $A^k \to A$ with $l^k \in A^k$, we have $l \in A$. Thus $a \in f(A)$, which implies $V \subset f(A)$. \hfill \Box

\section{Proof of Theorem 3.1}

\subsection{Necessity}

\textbf{History-Independent Risk Preference:} First notice that for all $h$,

$$U(\{\otimes\ell\}|h) = \int \left( (1 - \alpha)u(c) + \alpha U(\{\otimes\ell\}|hc) \right) dc d\mu(\alpha|h)$$

$$= (1 - \overline{\alpha_h})u(\ell) + \overline{\alpha_h} \int U(\{\otimes\ell\}|hc) dc,$$

where $u(\ell) = \int_c u(c) dc$. Hence, for all $h$,

$$U(\{\otimes\ell\}|h) = u(\ell) \left\{ 1 - \overline{\alpha_h} + \overline{\alpha_h} \int 1d\ell + \overline{\alpha_h} \int \left( -\overline{\alpha_{hc}} + \overline{\alpha_{hc'}} \int 1d\ell' \right) dc \right\}$$

$$+ \overline{\alpha_h} \int \overline{\alpha_{hc}} \left( \int \left( -\overline{\alpha_{hc}} + \overline{\alpha_{hc'}} \int 1d\ell' \right) dc' \right) dc + \cdots$$

$$= u(\ell).$$

Thus, for all $\ell, \ell'$ and all histories $h, h'$,

$$U(\{\otimes\ell\}|h) > U(\{\otimes\ell'\}|h) \iff u(\ell) > u(\ell') \iff U(\{\otimes\ell\}|h') > U(\{\otimes\ell'\}|h').$$

We show that for all $(u, \{\mu(\cdot|h)\})$, there exists $U(\cdot|h)$ satisfying the functional equation. Let $W$ be the Banach space of all real-valued continuous functions on $Z \times H$ with the sup-norm metric. For all $(x, h) \in Z \times H$ and $W \in W$, define

$$T(W)(x, h) \equiv \int_{[0, 1]} \max_{l \in x} \int_{C \times Z} \left( (1 - \alpha)u(c) + \alpha W(z, hc) \right) dc dz d\mu(\alpha|h). \quad (12)$$

\textbf{Lemma D.1.} $T(W) \in W$ for all $W \in W$.

\textbf{Proof.} We want to show that if $(x^n, h^n) \to (x^0, h^0)$, then $T(W)(x^n, h^n) \to T(W)(x^0, h^0)$. For all $(l, h) \in \Delta(C \times Z) \times H$, define $f(l, h) \in \mathbb{R}^2$ by $f_1(l, h) \equiv u(l, c)$ and $f_2(l, h) \equiv \int_{C \times Z} W(z, hc) dl$. It is easy to see that $f_1 : \Delta(C \times Z) \times H \to \mathbb{R}$ is continuous.

\textbf{Step 1:} $f_2 : \Delta(C \times Z) \times H \to \mathbb{R}$ is continuous.
First, we will claim that if $h^n \to h$, then \( \sup_{(c,z) \in C \times Z} |W(z, h^n c) - W(z, hc)| \to 0 \). Suppose otherwise. Then, there exists \( \varepsilon > 0 \) such that \( \sup_{(c,z) \in C \times Z} |W(z, h^n c) - W(z, hc)| \geq \varepsilon \) for infinitely many \( m \). There exists a corresponding sequence \( \{(c^n, z^n)\} \) satisfying \( |W(z^n, h^n c^n) - W(z^n, hc^n)| \geq \frac{\varepsilon}{2} \) for all \( m \). Since \( C \times Z \) is compact, we can assume that \( \{(c^n, z^n)\} \) converges to a limit \( (c, z) \). Since \( W \) is continuous, \( |W(z^n, h^n c^n) - W(z, hc^n)| \to |W(z, hc) - W(z, hc)| = 0 \), which is a contradiction. By Lemma C.1, \( f_2 \) is continuous.

**Step 2:** For all \( (x, h) \in Z \times H \), let \( f(x, h) \equiv \{f(l, h) \in \mathbb{R}^2 | l \in X \} \). Then, \( f(x^n, h^n) \to f(x^0, h^0) \) with respect to the Hausdorff metric.

Let \( H^* \equiv \{h^n\}_{n=1}^\infty \cup \{h^0\} \). Since \( h^n \to h^0 \), \( H^* \) is a compact subset of \( H \). Since \( f = (f_1, f_2) : \Delta(C \times Z) \times H \to \mathbb{R}^2 \) is a continuous function, the desired result follows from Lemma C.2.

Notice that
\[
\sigma(\alpha; x^n, h^n) \equiv \max_{l \in x^n} \int_{C \times Z} \left((1 - \alpha)u(c) + \alpha W(z, h^n c)\right) dl(c, z) \\
= \max_{l \in x^n} (1 - \alpha)u(l_c) + \alpha \int_{C \times Z} W(z, h^n c) dl = \max_{(a_1, a_2) \in f(x^n, h^n)} (1 - \alpha)a_1 + \alpha a_2.
\]

By Lemma C.2 and the same argument of HHT[Lemma C.4 (i), Step 2], we can show that
\[
\sup_{\alpha \in [0, 1]} |\sigma(\alpha; x^n, h^n) - \sigma(\alpha; x^0, h^0)| \to 0
\]
as \( n \to \infty \). Thus, by Lemma C.1, \( T(W)(x^n, h^n) \to T(W)(x^0, h^0) \).

By Lemma D.1, the operator \( T : W \to W \) is well-defined. To show that \( T \) is a contraction mapping, it suffices to verify that (i) \( T \) is monotonic, that is, \( T(W) \geq T(V) \) whenever \( W \geq V \), and (ii) \( T \) satisfies the discounting property, that is, there exists \( \delta \in [0, 1] \) such that for any \( W \) and \( \beta \geq 0 \), \( T(W + \beta) \leq T(W) + \delta \beta \) (see Aliprantis and Border [1, p.97, Theorem 3.53]).

**Step 1:** \( T \) is monotonic.

Take any \( W, V \in W \) with \( W \geq V \). Since \( (1 - \alpha)u(c) + \alpha W(z, hc) \geq (1 - \alpha)u(c) + \alpha V(z, hc) \) for all \( (c, z, h) \in C \times Z \times H \) and and \( \alpha \in [0, 1] \), we have
\[
\int_{C \times Z} ((1 - \alpha)u(c) + \alpha W(z, hc)) dl(c, z) \geq \int_{C \times Z} ((1 - \alpha)u(c) + \alpha V(z, hc)) dl(c, z)
\]
for all \( l \), and hence, \( T(W)(x, h) \geq T(V)(x, h) \).

**Step 2:** \( T \) satisfies the discounting property.

Let \( \delta \equiv \sup \{\overline{\mu}_h | h \in H\} < 1 \). For all \( W \in W \) and \( \beta \geq 0 \),
\[
T(W + \beta)(x, h) = \int_{[0, 1]} \max_{l \in x} \int_{C \times Z} ((1 - \alpha)u(c) + \alpha (W(z, h\alpha c) + \beta)) dl(c, z) d\mu(\alpha|h) \\
= T(W)(x, h) + \overline{\mu}_h \beta \leq T(W)(x, h) + \delta \beta.
\]

29
By Steps 1 and 2, $T$ is a contraction mapping. Thus, the fixed point theorem (See Gul and Pesendorfer [12, Lemma 6]) ensures that there exists a unique $U \in \mathcal{W}$ satisfying $U = T(U)$. This $U$ satisfies the equation (1).

### D.2 Sufficiency

Let $\overline{x}(x)$ denote the closed convex hull of $x$. As in DLR, Order, Continuity, and Independence imply $x \sim_{\overline{h}} \overline{x}(x)$. Hence we can pay attention to the sub-domain

$$Z_c \equiv \{ x \in Z | x = \overline{x}(x) \}.$$ 

Since $Z_c$ is a mixture space, Order, Continuity, and Independence ensure that $\succeq_h$ can be represented by a continuous mixture linear function $U(\cdot|h) : Z_c \rightarrow \mathbb{R}$.

For all $h$ and $\ell \in \Delta(C)$, let $u(\ell|h) \equiv U(\{\otimes \ell\}|h)$. By definition, $u(\cdot|h)$ is continuous and represents $\succeq_h$ on $\otimes \Delta_c$.

**Lemma D.2.** $u(\cdot|h)$ is mixture linear.

*Proof.* Take all $\ell, \ell' \in \Delta(C)$ and $\lambda \in [0, 1]$. Since $U(\cdot|h)$ is mixture linear and $\succeq_h$ satisfies IID-Separability,

$$u(\lambda \ell + (1 - \lambda)\ell'|h) = U(\{\otimes (\lambda \ell + (1 - \lambda)\ell')\}|h) = U(\{\lambda \otimes \ell + (1 - \lambda) \otimes \ell'\}|h) = \lambda U(\{\otimes \ell\}|h) + (1 - \lambda)U(\{\otimes \ell'\}|h) = \lambda u(\ell|h) + (1 - \lambda)u(\ell'|h).$$

\[\square\]

**Lemma D.3.** History-dependent continuous mixture linear representations $\{U(\cdot|h)\}_{h \in H}$ can be taken to satisfy

$$U(\{\otimes \ell\}|h) = U(\{\otimes \ell\}|h')$$

for all $h, h' \in H$.

*Proof.* Fix a history $\bar{h}$ arbitrarily. By History Independent Risk Preference, for all $h, \succeq_h$ and $\succeq_{\bar{h}}$ are equivalent on $\otimes \Delta_c$. By Lemma D.2, $u(\cdot|h)$ and $u(\cdot|\bar{h})$ are continuous mixture linear functions that represent the same preference over $\Delta(C)$. Thus there exist $\beta(h) > 0$ and $\gamma(h) \in \mathbb{R}$ satisfying $u(\ell|h) = \beta(h)u(\ell|\bar{h}) + \gamma(h)$. Let $U'(x|h) \equiv (U(x|h) - \gamma(h))/\beta(h)$. Then, $U'(\cdot|h)$ is a continuous mixture linear function representing $\succeq_h$. Moreover, by definition, for all $\ell \in \Delta(C)$, $U'(\otimes \ell|h) = u(\ell|h)$. Redefine $U'(\cdot|h)$ as $U(\cdot|h)$. \[\square\]

Fix a history $\bar{h}$ arbitrarily. Let $u(\ell) \equiv u(\ell|\bar{h})$. By Lemma D.3, $u(\ell) = u(\ell|h)$ for all $h$. By Nondegeneracy, $u$ is not constant. Since $u$ is continuous and $\Delta(C)$ is compact, there exist a maximal and a minimal lottery $\ell^+, \ell^- \in \Delta(C)$ with respect to $u$. We can normalize $u(\ell^-) = 0$.

**Lemma D.4.** For all $\eta \in \Delta(\Delta(C))$,

$$U(\{\otimes E(\eta)\}|h) = \int_{\Delta(C)} U(\{\otimes \ell\}|h) \, d\eta(\ell).$$

30
Proof. For all \( \eta \in \Delta(\Delta(C)) \), there exists a sequence \( \{\eta^n\} \) converging to \( \eta \) such that each \( \eta \) has a finite support. Thus let \( \eta^n \) be denoted by \( (\lambda^n_1; \ell^n_1, \ldots, \lambda^n_m; \ell^n_m) \) with \( \sum_i \lambda^n_i = 1 \). Since \( U(\{\otimes\})|h\) is mixture linear,

\[
U(\{\otimes E(\eta^n)\}|h) = U(\{\otimes(\sum_i \lambda^n_i \ell^n_i)\}|h) = \sum_i \lambda^n_i U(\{\otimes \ell^n_i\}|h) = \int_{\ell} U(\{\otimes \ell\}|h) d\eta^n.
\]

Since \( U(\{\otimes\})|h\) and \( E \) are continuous, we have the desired result as \( \eta^n \to \eta \). \( \square \)

Lemma D.5. (i) For all \( l \in \Delta(C \times Z) \) and \( h \),

\[
U(\{\otimes l_h\}|h) = \int_{C \times Z} U(z|h) d\ell(c,z).
\]

(ii) For all \( l \in \Delta(C \times Z) \) and \( h, h' \in H \),

\[
U(\{\otimes l_h\}|h) = U(\{\otimes l\}|h').
\]

Proof. (i) By definition, \( z \sim_{hc} \{\otimes f_h(c,z)\} \). Since \( U(\{\otimes \ell\}|h) = U(\{\otimes \ell\}|h') = u(\ell) \) for all histories \( h, h' \in H \) and \( \ell \in \Delta(C) \), we have

\[
U(z|h_c) = U(\{\otimes f_h(c,z)\}|h')
\]

for all histories \( h' \). By Lemma D.4,

\[
U(\{\otimes l_h\}|h) = U(\{\otimes E(F_h(l))\}|h) = \int_{\Delta(C)} U(\{\otimes \ell'\}|h) dF_h(l)(\ell') = \int_{\Delta(C)} U(\{\otimes l_h\}|h) d\ell \circ f_h^{-1}(\ell') = \int_{C \times Z} U(\{\otimes f_h(c,z)\}|h) d\ell(c,z)
\]

\[
= \int_{C \times Z} U(z|h_c) d\ell(c,z).
\]

(ii) Since \( f_h(c,z) \in \Delta(C) \), \( U(\{\otimes f_h(c,z)\}|h) = U(\{\otimes f_h(c,z)\}|h') \) for all \( h, h' \in H \). Hence

\[
U(\{\otimes E(F_h(l))\}|h) = \int_{C \times Z} U(\{\otimes f_h(c,z)\}|h) d\ell(c,z)
\]

\[
= \int_{C \times Z} U(\{\otimes f_h(c,z)\}|h') d\ell(c,z) = U(\{\otimes E(F_h(l))\}|h').
\]

Let

\[
Z_h \equiv \{x \in Z_e | x = D_h(x)\}.
\]

31
From Lemma B.2 (vi), $Z_h$ is compact. Moreover, Lemma B.2 (iv) and (v) imply that all $x \in Z_h$ are compact and convex.

Let $C([0, 1])$ be the set of all continuous functions on $[0, 1]$ with the sup-norm. For all $x \in Z_h$, define the function $\sigma : Z_h \to C([0, 1])$ by

$$\sigma_x(\alpha) \equiv \max_{l \in x}(1 - \alpha)u(l_c) + \alpha U(\{\otimes l_h\}|h)$$

for all $\alpha \in [0, 1]$. By Lemma D.5 (i), (13) can be rewritten as

$$\max_{l \in x} \int_{C \times Z} \left((1 - \alpha)u(c) + \alpha U(z|h)c\right)dl(c, z).$$

Let

$$V_h(x) \equiv \{(u, w) \in \mathbb{R}_+^2 \mid u = u(l_c), w = U(\{\otimes l_h\}|h) \text{ for some } l \in x\}.$$ 

**Lemma D.6.** $V_h(x)$ is compact and convex.

**Proof.** Notice that $u$ is continuous and mixture linear. Moreover, by Lemma B.2 (ii) and (iii), $U(\{\otimes E(F_h(\cdot))\}|h)$ is continuous and mixture linear. Since $V_h(x)$ is the image of a compact convex set under a continuous and mixture linear function, $V_h(x)$ is compact and convex. \qed

Let

$$u^M \equiv \max\{u(l_c) \in \mathbb{R}_+ \mid l \in x\}, \quad w^M \equiv \max\{U(\{\otimes l_h\}|h) \in \mathbb{R}_+ \mid l \in x\}.$$

**Lemma D.7.** (i) $(0, 0) \in V_h(x)$, (ii) $(u^M, 0) \in V_h(x)$, and (iii) $(0, w^M) \in V_h(x)$.

**Proof.** (i) Since any lottery dominates $\{\otimes \ell^-\}$, $\{\otimes \ell^-\} \in D_h(x) = x$. Notice that $u(\ell^-) = 0$ and $U(\{\otimes E(F_h(\ell^-))\}|h) = U(\{\otimes \ell^-\}|h) = u(\ell^-) = 0$. That is, $(0, 0) \in V_h(x)$.

(ii) There exists $l \in x$ with $u(l_c) = u^M$. Take the lottery $l^* \equiv l_c \otimes \{\otimes \ell^-\} \in \Delta(C \times C)$. Then, $u(l^*_c) = u(l_c) = u^M$ and

$$U(\{\otimes E(F_h(l^*))\}|h) \equiv \int_{C \times Z} U(\{\otimes f_h(c, z)\}|h) dl(c) \otimes \{\otimes \ell^-\} = \int_C U(\{\otimes f_h(c, z)\}|h) dl_c.$$ 

Since $u(l_c) = u(l^*_c)$ and $U(\{\otimes E(F_h(l))\}|h) \geq U(\{\otimes E(F_h(l^*))\}|h)$, $l$ dominates $l^*$, and hence $l^* \in D_h(x) = x$. Therefore, $(u^M, 0) \in V_h(x)$.

(iii) There exists $l \in x$ satisfying $U(\{\otimes l_h\}|h) = w^M$. Take the lottery $l^* \equiv \ell^- \otimes \{\otimes l_h\} \in \Delta(C \otimes C)$. Then, $u(l^*_c) = u(\ell^-) = 0$. Moreover, by Lemma D.5 (ii),

$$U(\{\otimes E(F_h(l^*))\}|h) \equiv \int_{C \times Z} U(\{\otimes f_h(c, z)\}|h) dl(\ell^- \otimes \{\otimes l_h\})$$

$$= \int_C U(\{\otimes f_h(c, \otimes l_h)\}|h) dl^- - \int_C U(\{\otimes l_h\}|h) dl^-$$

$$= \int_C U(\{\otimes l_h\}|h) dl^- = U(\{\otimes l_h\}|h) = U(\{\otimes E(F_h(l))\}|h).$$
Since \( u(l_c) \geq u(l^*_c) = 0 \) and \( U(\{ \otimes E(F_h(l)) \}|h) = U(\{ \otimes E(F_h(l^*)) \}|h) \), \( l \) dominates \( l^* \), and hence \( l^* \in D_h(x) = x \). Therefore, \((0, w^M) \in V_h(x)\). 

\[ \text{Lemma D.8.} \quad \text{Take any point } (u, w) \in V_h(x). \text{ If } u \geq u' \geq 0 \text{ and } w \geq w' \geq 0, \text{ then } (u', w') \in V_h(x). \]

\[ \text{Proof.} \quad \text{Since } u^M \geq u \text{ and } w^M \geq w, \text{ Lemmas D.6 and D.7 imply that } (u, 0), (0, w) \in V_h(x). \text{ Again, by Lemmas D.6 and D.7,} \]

\[
\frac{w'}{w}(u, w) + \left( 1 - \frac{w'}{w} \right) (u, 0) = (u, w') \in V_h(x), \\
\frac{w'}{w}(0, w) + \left( 1 - \frac{w'}{w} \right) (0, 0) = (0, w') \in V_h(x).
\]

Hence

\[
\frac{u'}{u}(u, w') + \left( 1 - \frac{u'}{u} \right) (0, w') = (u', w') \in V_h(x).
\] 

\[ \text{Lemma D.9.} \quad \sigma : Z_h \to C([0, 1]) \text{ is injective.} \]

\[ \text{Proof.} \quad \text{Take } x, x' \in Z_h \text{ with } x \neq x'. \text{ Without loss of generality, assume } x \notin x'. \text{ Take } \tilde{l} \in x \setminus x'. \text{ Let } \tilde{u} = u(\tilde{l}_c) \text{ and } \tilde{w} = U(\{ \otimes \tilde{l}_h \}|h). \]

We will claim that \((\{ \tilde{u}, \tilde{w} \}) + \mathbb{R}^2_+ \cap V_h(x') = \emptyset\). Suppose otherwise. Then, there exists \( l' \in x' \) such that \( u(l'_c) \geq \tilde{u} \) and \( U(\{ \otimes l'_h \}|h) \geq \tilde{w} \). By Lemma D.8, \((\tilde{u}, \tilde{w}) \in V_h(x')\), that is, \( \tilde{l} \in D_h(x') = x' \). This is a contradiction.

By Lemma D.6 and the above claim, the separating hyperplane theorem ensures that there exists \( \alpha \in [0, 1] \) and \( \gamma \in \mathbb{R} \) such that \((1 - \alpha)\tilde{u} + \alpha\tilde{w} > \gamma > (1 - \alpha)u' + \alpha w' \) for all \((u', w') \in V_h(x')\). Equivalently,

\[
(1 - \alpha)u(\tilde{l}_c) + \alpha U(\{ \otimes \tilde{l}_h \}|h) > \gamma > (1 - \alpha)u(l'_c) + \alpha U(\{ \otimes \tilde{l}_h \}|h),
\]

for all \( l' \in x' \). Hence,

\[
\sigma_x(\alpha) = \max_{l \in x} (1 - \alpha)u(l_c) + \alpha U(\{ \otimes l_h \}|h) \geq (1 - \alpha)u(\tilde{l}_c) + \alpha U(\{ \otimes \tilde{l}_h \}|h)
\]

\[
> \max_{l' \in x'} (1 - \alpha)u(l'_c) + \alpha U(\{ \otimes \tilde{l}_h \}|h) = \sigma_{x'}(\alpha).
\]

Therefore, \( \sigma_x \neq \sigma_{x'} \).

\[ \text{Lemma D.10.} \quad (i) \sigma \text{ is continuous.} \]

\[ (ii) \text{ For all } x, y \in Z_h \text{ and } \lambda \in [0, 1], \lambda \sigma_x + (1 - \lambda)\sigma_y = \sigma_{D_h(x+\gamma, y)} \]
Lemma D.11. (i) \( C_h \) is convex.

(ii) The zero function belongs to \( C_h \).

(iii) The constant function equal to a positive number \( c > 0 \) belongs to \( C_h \).

(iv) The supremum of any two points \( f, f' \in C_h \) belongs to \( C_h \). That is, \( \max[f(\alpha), f'(\alpha)] \) belongs to \( C_h \).

(v) For all \( f \in C_h \), \( f \geq 0 \).

Proof. (i) Take any \( f, f' \in C_h \) and \( \lambda \in [0,1] \). There are \( x, x' \in \mathcal{Z}_2 \) satisfying \( f = \sigma_x \) and \( f' = \sigma_{x'} \). From Lemma D.10 (ii),

\[
\lambda f + (1 - \lambda)f' = \lambda \sigma_x + (1 - \lambda)\sigma_{x'} = \sigma_{D_h(\lambda x + (1 - \lambda)x')} \in \mathcal{Z}_h.
\]

Hence, \( C_h \) is convex.

(ii) Let \( x \equiv D_h(\{\ell^-\}) \in \mathcal{Z}_h \). Then, for all \( \alpha \),

\[
\sigma_x(\alpha) = \max_{l \in D_h(\{\ell^-\})} (1 - \alpha)u(l_c) + \alpha U(\{\otimes E(F_h(l))\}|h) = (1 - \alpha)u(l^-) + \alpha U(\{\otimes l^-\}|h) = 0.
\]

(iii) Take a lottery \( \ell \in \Delta(C) \) such that \( u(\ell) > 0 \). Let \( c \equiv u(\ell) \). Since \( U(\{\otimes E(F_h(\otimes \ell^+))\}|h) = u(\ell^+) \) and \( U(\{\otimes E(F_h(\otimes \ell^-))\}|h) = (1 - \alpha)u(\ell^-) + \alpha U(\{\otimes \ell^-\}|h) = 0 \), Lemma D.8 ensures that \( (c,c) \in V_h(D_h(\{\otimes \ell^+\})) \). Hence, there exists \( l^* \in D_h(\{\otimes \ell^+\}) \) such that \( U(l^*|c) = c \) and \( U(\{\otimes E(F_h(l^*))\}|h) = c \). Let \( x \equiv D_h(\{l^*\}) \in \mathcal{Z}_h \). Then, for all \( \alpha \),

\[
\sigma_x(\alpha) = \max_{l \in l^*} (1 - \alpha)u(l_c) + \alpha U(\{\otimes E(F_h(l))\}|h) = (1 - \alpha)u(l^*_c) + \alpha U(\{\otimes E(F_h(l^*))\}|h) = c.
\]

(iv) There exist \( x', x \in \mathcal{Z}_h \) such that \( f = \sigma_x \) and \( f' = \sigma_{x'} \). Let \( f'' \equiv \sigma_{D_h(\{\otimes l^-\})} \in C_h \). Then, \( f''(\alpha) = \max[\sigma_x(\alpha), \sigma_{x'}(\alpha)] \).

(v) There exists \( x \in \mathcal{Z}_h \) such that \( f = \sigma_x \). Since \( D_h(\{\otimes l^-\}) \subset x \), Lemma D.11 (ii) implies \( f(\alpha) = \sigma_x(\alpha) \geq \sigma_{D_h(\{\otimes l^-\})}(\alpha) = 0 \) for any \( \alpha \).

Define \( T_h : C_h \to \mathbb{R} \) by \( T_h(f) = U(\sigma^{-1}(f)|h) \). By Lemma D.9, \( T_h \) is well-defined. Notice that \( T_h(0) = 0 \) and \( T_h(c) = c \), where 0 and \( c \) are identified with the zero function and the constant function equal to \( c > 0 \), respectively. Since \( U(\cdot|h) \) and \( \sigma \) is continuous and mixture linear, so is \( T_h \).
By the same argument as in DLR and Dekel, Lipman, Rustichini, and Sarver [6], we will extend \( T \) to \( C([0, 1]) \) step by step. For any \( r \geq 0 \), let \( rC_h \equiv \{ rf \mid f \in C_h \} \) and \( H_h \equiv \bigcup_{r \geq 0} rC_h \). For any \( f \in H_h \setminus 0 \), there is \( r > 0 \) satisfying \((1/r) f \in C_h \). Define \( T_h(f) \equiv rT_h((1/r)f) \).

By the same argument of C.6 of HHT, we can show that \( T_h \) is linear on \( C_h \), which implies that \( T_h(f) \) is well-defined. It is easy to see that \( T_h \) on \( H_h \) is mixture linear. By the same argument of C.6 of HHT, \( T_h \) is also linear.

Let

\[
H_h^* \equiv H_h - H_h = \{ f_1 - f_2 \in C([0, 1]) \mid f_1, f_2 \in H_h \}.
\]

For any \( f \in H_h^* \), there are \( f_1, f_2 \in H_h \) satisfying \( f = f_1 - f_2 \). Define \( T_h(f) \equiv T_h(f_1) - T_h(f_2) \). We can verify that \( T_h : H_h^* \to \mathbb{R} \) is well-defined because of linearity of \( T_h \) on \( H_h \).

**Lemma D.12.** \( H_h^* \) is dense in \( C([0, 1]) \).

**Proof.** From the Stone-Weierstrass theorem, it is enough to show that (i) \( H_h^* \) is a vector sublattice, (ii) \( H_h^* \) separates the points of \([0, 1] \); that is, for any two distinct points \( \alpha, \alpha' \in [0, 1] \), there exists \( f \in H_h^* \) with \( f(\alpha) \neq f(\alpha') \), and (iii) \( H_h^* \) contains the constant functions equal to one. By the exactly same argument as Lemma 11 (p.928) in DLR, (i) holds.

To verify condition (ii), take \( \alpha, \alpha' \in [0, 1] \) with \( \alpha \neq \alpha' \). Without loss of generality, \( \alpha > \alpha' \). Take a lottery \( \ell \in \Delta(C) \) such that \( u(\ell) > 0 \). Since \( U(\{ \otimes E(F_h(\ell^+)) \}) h = u(\ell^+) \), \( u(\ell^+) \in V_h(D_h(\{ \otimes \ell^+ \})) \). Since \( u(\ell^+) \geq u(\ell) \), Lemma D.8 ensures that \( (u(\ell), 0) \in V_h(D_h(\{ \otimes \ell^+ \})) \). Hence, there exists \( l^* \in D_h(\{ \otimes \ell^+ \}) \) such that \( U(\{ \otimes E(F_h(l^*)) \}) h = 0 \) and \( u(l^*_c) = u(\ell) \). Let \( x \equiv D_h(\{ l^* \}) \in \mathcal{Z}_h \). Then, \( \sigma_x \in C_h \subset H_h^* \) and

\[
\sigma_x(\alpha) = (1 - \alpha)u(l^*_c) + \alpha U(\{ \otimes E(F_h(l^*)) \} h) = (1 - \alpha)u(l^*_c) \\
> (1 - \alpha')u(l^*_c) = (1 - \alpha')u(l^*_c) + \alpha' U(\{ \otimes E(F_h(l^*)) \} h) = \sigma_x(\alpha').
\]

Finally, condition (iii) directly follows from Lemma D.11 (iii) and the definition of \( H_h \). \( \square \)

By adopting the same argument as in Theorem 2 of Dekel, Lipman, Rustichini, and Sarver [6], we can show that there exists a constant \( K > 0 \) such that \( T_h(f) \leq K \| f \| \) for any \( f \in H_h^* \). See Lemma C.8 of HHT for more details.

By the Hahn-Banach theorem, we can extend \( T_h : H_h^* \to \mathbb{R} \) to \( \overline{T}_h : C([0, 1]) \to \mathbb{R} \) in a linear, continuous and increasing way. Since \( H_h^* \) is dense in \( C([0, 1]) \) by Lemma D.12, this extension is unique.

Since \( \overline{T}_h \) is a positive linear functional on \( C([0, 1]) \), the Riesz representation theorem ensures that there exists a unique countably additive probability measure \( \mu(\cdot | h) \) on \([0, 1] \) satisfying

\[
\overline{T}_h(f) = \int_{[0, 1]} f(\alpha) \, d\mu(\alpha | h),
\]

for all \( f \in C([0, 1]) \). Thus, by Lemma D.5 (i), we have

\[
U(x | h) = \overline{T}_h(\sigma(x)) = \int_{[0, 1]} \max_{l \in x} ((1 - \alpha)u(l_c) + \alpha U(\{ l_h \} | h)) \, d\mu(\alpha | h) \\
= \int_{[0, 1]} \max_{l \in x} \int_{C \times Z} ((1 - \alpha)u(c) + \alpha U(z | h)) \, dl(c, z) \, d\mu(\alpha | h).
\]

35
Lemma D.13. \( \sup \{ \overline{\alpha} \mid h \in H \} < 1 \).

Proof. Take any \( \ell, \ell' \) with \( U(\{ \ell \}) > U(\{ \ell' \}) \). By Strong Archimedean, there exists \( \ell'' \in \Delta(C) \) such that \( U(\{ \ell \}) > U(\{ \ell'' \}) \geq U(\{ \ell' \}) \). Since \( U(\{ \ell \otimes \ell'' \} | h) = (1 - \overline{\alpha}h)u(\ell') + \overline{\alpha}u(\ell') = U(\{ \ell' \} | h) \), we have \( u(\ell) > u(\ell'') > u(\ell') \). There exists \( \lambda \in (0, 1) \) such that \( \lambda u(\ell) + (1 - \lambda)u(\ell') > u(\ell') \), and hence, \( \lambda u(\ell) + (1 - \lambda)u(\ell') > (1 - \overline{\alpha}h)u(\ell') + \overline{\alpha}u(\ell') \). The latter condition implies \( (u(\ell) - u(\ell'))(\lambda - \overline{\alpha}h) > 0 \). Since \( u(\ell) - u(\ell') > 0 \), we have \( \overline{\alpha}h < \lambda \). This property holds for all \( h \), and hence, we have \( \sup \{ \overline{\alpha} \mid h \in H \} \leq \lambda < 1 \).

\[ \square \]

Lemma D.14. For all \( x \in Z, U(x) \colon H \rightarrow [0, 1] \) is continuous.

Proof. Since \( U(\{ \ell \}) = u(\ell) \) for all \( \ell \in \Delta(C) \), by Continuity-(ii), \( \{ h \in H \mid U(x|h) \geq u(\ell) \} \) and \( \{ h \in H \mid u(\ell) \geq U(x|h) \} \) are closed. Since \( \succ_h \) satisfies Order, this is equivalent to say that \( \{ h \in H \mid U(x|h) < u(\ell) \} \) and \( \{ h \in H \mid u(\ell) < U(x|h) \} \) are open for all \( \ell \in \Delta(C) \). Since \( u(\ell) \) varies over \([0, 1] \), this property says that the inverse images of all subbasic open sets of \([0, 1] \) are open in \( H \). By Dugundji [7, p.83, Theorem 10.1], \( U(x) \) is continuous.

\[ \square \]

Let \( L_1 = \{ \ell \otimes \{ \ell' \} \in L \mid \ell, \ell' \in \Delta(C) \} \). For \( x \in K(L_1) \subset Z \), define the function \( \sigma : K(L_1) \rightarrow C([0, 1]) \) by

\[ \sigma_x(\alpha) = \max_{\ell \otimes \{ \ell' \} \in x} (1 - \alpha)u(\ell) + \alpha u(\ell'). \]

Denote \( C_1 = \{ \sigma_x \mid x \in K(L_1) \} \), \( H_1 = \cup_{r \geq 0} r C_1 \), and \( H_1^* = H_1 - H_1 \).

Lemma D.15. For all \( f \in H_1^* \), \( \int f(\alpha) \, d\mu(\alpha \mid \cdot) : H \rightarrow R \) is continuous.

Proof. Since \( u(\ell) = U(\{ \ell \mid h \}) \) for all \( h \in H \), we have for all \( x \in K(L_1) \),

\[ \int \sigma_x(\alpha) \, d\mu(\alpha \mid h) = \int \max_{\ell \otimes \{ \ell' \} \in x} (1 - \alpha)u(\ell) + \alpha U(\{ \ell' \} \mid h) \, d\mu(\alpha \mid h) = U(x \mid h). \]

Let \( f \in H_1^* \). Then there exists \( r, r' \geq 0 \) and \( x, x' \in K(L_1) \) such that \( f = r \sigma_x - r' \sigma_{x'} \). Thus, \( \int f(\alpha) \, d\mu(\alpha \mid \cdot) = r U(x \mid \cdot) - r' U(x' \mid \cdot) \), and the result directly follows from Lemma D.14.

\[ \square \]

Lemma D.16. \( H_1^* \) is dense in \( C([0, 1]) \).

Proof. The same argument as the proof of Lemma D.12 applies. We show only that (ii) \( H_1^* \) separates the points of \([0, 1] \), and (iii) \( H_1^* \) contains the constant functions equal to one.

For (ii), let \( x = \{ \ell^+ \otimes \{ \ell^- \} \} \), where \( \ell^+ \) and \( \ell^- \) are a maximal and a minimal lottery with respect to \( u \). Notice that \( u(\ell^+) = 1 \) and \( u(\ell^-) = 0 \) by normalization. Then \( \sigma_x \in H_1^* \) and \( \sigma_x(\alpha) = (1 - \alpha)u(\ell^+) \) separates the points. For (iii), \( \sigma_{\{ \ell^+ \}} \) is the required function.

\[ \square \]

Lemma D.17. \( \mu(\cdot \mid h) \) is continuous with respect to \( h \).
Proof. Suppose $h^n, h \in H$ and $h^n \to h$. Let $f \in C([0,1])$. For abbreviation, denote $\mu^n = \mu(\cdot | h^n)$ and $\mu = \mu(\cdot | h)$. We show that $\int f \, d\mu^n \to \int f \, d\mu$.

Let $\epsilon > 0$. From Lemma D.16, we can take $f_\epsilon \in H^*_1$ satisfying $\sup |f_\epsilon - f| < \frac{\epsilon}{3}$, and thus

$$\left| \int f_\epsilon \, d\mu^n - \int f \, d\mu^n \right| \leq \int |f_\epsilon - f| \, d\mu^n < \frac{\epsilon}{3}, \text{ and}$$

$$\left| \int f_\epsilon \, d\mu - \int f \, d\mu \right| \leq \int |f_\epsilon - f| \, d\mu < \frac{\epsilon}{3}.$$ 

Moreover, from Lemma D.15, there exists $n$ such that for all $n \geq \bar{n}$,

$$\left| \int f_\epsilon \, d\mu^n - \int f_\epsilon \, d\mu \right| < \frac{\epsilon}{3}.$$

Therefore, we have for all $n \geq \bar{n}$,

$$\left| \int f \, d\mu^n - \int f \, d\mu \right|$$

$$< \left| \int f_\epsilon \, d\mu^n - \int f_\epsilon \, d\mu \right| + \left| \int f_\epsilon \, d\mu^n - \int f_\epsilon \, d\mu \right| + \left| \int f_\epsilon \, d\mu - \int f \, d\mu \right|$$

$$< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.$$

\[ \square \]

E Proof of Theorem 3.2

(i) Since mixture linear functions $u$ and $u'$ represent the same preference over $\Delta(C)$, by the standard argument, $u'$ is rewritten as an affine transformation of $u$. That is, $u$ and $u'$ are cardinally equivalent.

(ii) From (i), there exist $\gamma > 0$ and $\zeta \in \mathbb{R}$ such that $u' = \gamma u + \zeta$. Since $U(\cdot | h)$ and $U'(\cdot | h)$ are mixture linear functions representing the same preference, there exist $\gamma_h > 0$ and $\zeta_h \in \mathbb{R}$ such that $U'(\cdot | h) = \gamma_h U(\cdot | h) + \zeta_h$. Let $x_c$ be the perfect commitment menu to $c$, that is, $x_c \equiv \{(c, \{(c, \{\cdots \})\})\}$. Since $U(x_c | h) = u(c)$ and $U'(x_c | h) = u'(c)$, we have $U'(x_c | h) = \gamma U(x_c | h) + \zeta$, which implies $\gamma_h = \gamma$ and $\zeta_h = \zeta$. Now we have

$$U'(x | h) = \int_{[0,1]} \max_{l \in c} \int_{C \times Z} ((1 - \alpha)u'(c) + \alpha U'(z|h)c) \, dl(c,z) \, d\mu'(\alpha | h)$$

$$= \int_{[0,1]} \max_{l \in c} \int_{C \times Z} ((1 - \alpha)(\gamma u(c) + \zeta) + \alpha(\gamma U(z|h)c + \zeta)) \, dl(c,z) \, d\mu'(\alpha | h)$$

$$= \gamma \int_{[0,1]} \max_{l \in c} \int_{C \times Z} ((1 - \alpha)u(c) + \alpha U(z|h)c) \, dl(c,z) \, d\mu'(\alpha | h) + \zeta.$$
Hence,
\[ U'''(x|h) \equiv \int_{[0,1]} \max_{l \in x} \int_{C \times Z} ((1 - \alpha)u(c) + \alpha U(z|hc)) \, dl(c, z) \, d\mu'(\alpha|h) \]
also represents the same preference. Since \( U'(\cdot|h) = \gamma U(\cdot|h) + \xi \) and \( U''(\cdot|h) = \gamma U''(\cdot|h) + \xi \), we must have \( U(x|h) = U'''(x|h) \) for all \( x \). For all \( x \in \mathcal{Z} \) and \( \alpha \in [0, 1] \), let
\[ \sigma_x(\alpha) \equiv \max_{l \in x} \int_{C \times Z} ((1 - \alpha)u(c) + \alpha U(z|hc)) \, dl(c, z). \]
Then,
\[ U(x|h) = \int \sigma_x(\alpha) \, d\mu(\alpha|h) = \int \sigma_x(\alpha) \, d\mu'(\alpha|h) = U'''(x|h). \quad (14) \]
If \( x \) is convex, \( \sigma_x \) is its support function. Equation (14) holds also when \( \sigma_x \) is replaced with \( a\sigma_x - b\sigma_y \) for any convex menus \( x, y \) and \( a, b \geq 0 \). From Lemma D.12, the set of all such functions is a dense subset of the set of real-valued continuous functions over \([0, 1] \). Hence, equation (14) holds when \( \sigma_x \) is replaced with any real-valued continuous function. Hence, the Riesz representation theorem implies \( \mu(\cdot|h) = \mu'(\cdot|h) \).

F Proof of Theorem 3.3

Define the binary relation \( \succeq^* \) on \( H \) by
\[ h \succeq^* h' \iff \exists c \geq c' \text{ for some } c \in [h] \text{ and } c' \in [h']. \]
For all \( h \in H \), Sensitivity and Smoothing ensure that there exists a unique \( c \in C \) such that \( c \in [h] \). Define \( V: H \to \mathbb{R}_+ \) by \( V(h) = c \) according to this property. Then, \( V \) is a representation of \( \succeq^* \). It is enough to show that \( V(h) \) is rewritten as
\[ V(h) = (1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} c_t \]
for some \( \delta \in (0, 1) \). Indeed, under this claim, for all \( h, h' \in H \),
\[ \bar{c}_h = \bar{c}_{h'} = c \implies V(h) = V(h') = c \implies c \in [h], [h'] \implies c \succeq^* c \implies c \succeq^* h'. \]

First of all, \( V \) is a functional defined on a mixture space \( H \). We will claim that \( V \) satisfies all the assumptions of Epstein [8, Theorem 2]. By the Linearity axiom, for all \( h, h' \) and \( \lambda \in [0, 1] \),
\[ V(\lambda h + (1 - \lambda)h') = \lambda c + (1 - \lambda)c' = \lambda V(h) + (1 - \lambda)V(h'), \]
where \( c \in [h] \) and \( c' \in [h'] \). Thus, \( V \) is mixture linear.

Let \( h^n \to h \). there exist \( c, c^n \in C \) such that \( c \in [h] \) and \( c^n \in [h^n] \). We want to show that \( V(h^n) \to V(h) \). Suppose otherwise. Then, there exists \( \varepsilon > 0 \) such that \( |V(h^n) - V(h)| > \varepsilon \).
for infinitely many $m$, or equivalently, $|c^m - c| \geq \varepsilon$. Let $\{h^n\}$ denote a corresponding subsequence of $\{h^n\}$. Since $\{c^n\}$ is a sequence in $C = [0, M]$, there exists a subsequence $\{c^i\}$ converging to a point $c^* \in C$ with $c^* \neq c$. Let $\{h^i\}$ denote the corresponding subsequence of $\{h^m\}$. Notice that $\{h^i\}$ also converges to $h$. Since $h^i \to h$ and $c^i \to c^*$, Continuity implies that $c^* \in [h]$, or $V(h) = c^*$. This contradicts $V(h) = c \neq c^*$. Thus, $V$ is continuous.

Take any $c, c' \in C$ with $c \neq c'$. By Sensitivity, $V(c) = c \neq c' = V(c')$. Thus, Assumption 1 holds.

Take any $h, h' \in H$. Let $c \in [h] \cap [c]$, $c' \in [h'] \cap [c']$. Suppose $V(h) = c \geq c' = V(h')$. By Axiom, $c \in [hc] ^{\varepsilon} \cap [c']$. Let $\bar{c} \in [cc] ^{\varepsilon} \cap [c']$. Since $c \geq c'$, Axiom implies that $\bar{c} \geq c'$. Since $\bar{c} \in [hc] ^{\varepsilon} \cap [c']$, $V(hc) = c \geq c' = V(h')$. Thus, Assumption 2 holds.

Take any $c_1, c_1' \in C$ and $h, h' \in H$. Let $V(hc_1) \geq V(h'c_1')$. There exist $c, c' \in C$ such that $c \in [hc_1]$ and $c' \in [h'c_1]$ with $c \geq c'$. Let $\bar{c}, \bar{c}'$ satisfy $\bar{c} \in [hc_1']$ and $\bar{c}' \in [h'c_1']$. By Separability (ii), $\bar{c} \geq \bar{c}'$, that is, $V(hc_1) \geq V(h'c_1')$. Thus, Assumption 3 holds.

Take any $c_1, c_1' \in C$ and $h, h' \in H$. Let $V(hc_1) \geq V(h'c_1')$. There exist $c, c' \in C$ such that $c \in [hc_1]$ and $c' \in [h'c_1']$ with $c \geq c'$. Let $\bar{c}, \bar{c}'$ satisfy $\bar{c} \in [hc_1']$ and $\bar{c}' \in [h'c_1']$. By Separability (i), $\bar{c} \geq \bar{c}'$, that is, $V(hc_1) \geq V(h'c_1')$. Thus, Assumption 5 holds.

By Theorem 2 of Epstein [8], there exists $\delta \in (0, 1)$ and $v : C \to \mathbb{R}$ such that $V(h) = \sum_{t=1}^{\infty} \delta^{t-1} v(c_t)$. For all $c \in C$,

$$ c = V(c) = \frac{v(c)}{1 - \delta}. $$

Thus, $v(c) = (1 - \delta)c$ as desired.

### G Proof of Theorem 4.1

**Lemma G.1.** (i) (P1) implies that, for all $x \in Z_P$ and $l \in L_I$, $x \succ_{h_1} \{l\} \Rightarrow x \succ_{h_1} \{l\}$.

(ii) (P2) implies that, for all $l \in L_P$ and $x \in Z_I$, $\{l\} \succ_{h_2} x \Rightarrow \{l\} \succ_{h_2} x$.

**Proof.** (i) It suffices to show that $x \sim_{h_2} \{l\} \Rightarrow x \succ_{h_1} \{l\}$. By assumption, $l$ is denoted by $l \otimes \{l\}$ for some $\ell, \ell' \in \{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\}$. Let $\ell^+$ and $\ell^-$ be a maximal and minimal lottery in $\Delta(C)$ with respect to $u$. If either $\{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\}$ or $\{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\}$, we have $(1 - \pi_{h_2})u(\ell) + \pi_{h_2}u(\ell') > (1 - \pi_{h_2})u(\lambda + (1 - \lambda)\ell^-) + \pi_{h_2}u(\lambda) + (1 - \lambda)\ell^-)$ for all $\lambda \in (0, 1)$, that is, $x \sim_{h_2} \{\ell \otimes \ell'\} \succ_{h_2} \{(\lambda + (1 - \lambda)\ell^-) \otimes \{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\} \succ_{h_2} \{(\lambda + (1 - \lambda)\ell^-) \otimes \{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\} \}$ for all $\lambda \in (0, 1)$. By assumption, $x \succ_{h_1} \{(\lambda + (1 - \lambda)\ell^-) \otimes \{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\} \succ_{h_2} \{(\lambda + (1 - \lambda)\ell^-) \otimes \{\ell \otimes \ell'\} \succ_{h_2} \{\ell \otimes \ell'\} \}$ for all $\lambda \in (0, 1)$. Thus Continuity implies $x \succ_{h_1} \{l\}$ as $\lambda \to 1$. If both $\{\ell \otimes \ell'\}$ and $\{\ell \otimes \ell'\}$ are indifferent to $\{\ell \otimes \ell'\}$, define $x^\lambda \equiv \{(\lambda + (1 - \lambda)\ell^-) \otimes \{\ell \otimes \ell'\} \in x \in Z_P$. From the representation, $U(x^\lambda|h_2) > U(x|h_2) = U(\{l\}|h_2)$ for all $\lambda \in (0, 1)$. By assumption, $x^\lambda \succ_{h_1} \{l\}$. Thus by Continuity, $x \succ_{h_1} \{l\}$.

(ii) The result follows from the same argument as above.

**Lemma G.2.** (i) (P1) implies that, for all $x \in Z_P$, $U(x|h_2) \leq U(x|h_1)$.
(ii) (P2) implies that, for all \( x \in \mathcal{Z}_1 \), \( U(x|h_1) \leq U(x|h_2) \).

Proof. (i) Notice that for all \( \ell \in \Delta(C) \), \( U(\{\otimes\ell\}|h_1) = u(\ell) = U(\{\otimes\ell\}|h_2) \). Since \( \{\succ_h\}_{h \in H} \) satisfies Best-Worst, there exist \( l, l' \in \mathcal{L} \) such that \( \{l\} \succ_{h_2} x \succ_{h_2} \{l'\} \). Let \( \ell^+ \) and \( \ell^- \) be a maximal and minimal lottery in \( \Delta(C) \) with respect to \( u \). From the representation, \( \{\otimes\ell\} \succ_{h_2} \{l\} \) and \( \{\otimes\ell\} \succ_{h_2} \{\otimes\ell^-\} \). Thus, \( \{\otimes\ell^+\} \succ_{h_2} x \succ_{h_2} \{\otimes\ell^-\} \), or equivalently, \( u(\ell^+) \geq U(x|h_2) \geq u(\ell^-) \). By continuity of \( u \), there exists \( \lambda \in [0,1] \) such that \( U(x|h_2) = u(\lambda \ell^+ + (1-\lambda)\ell^-) \), or \( x \sim_{h_2} \{\otimes(\lambda \ell^+ + (1-\lambda)\ell^-)\} \). Since \( l \equiv \otimes(\lambda \ell^+ + (1-\lambda)\ell^-) \in \mathcal{L}_I \), \( x \sim_{h_2} \{l\} \) implies that \( x \succ_{h_1} \{l\} \). Thus, \( U(x|h_1) \geq U(\{l\}|h_1) = U(\{l\}|h_2) = U(x|h_2) \).

(ii) By the same argument as above, we have \( l \in \mathcal{L}_P \) such that \( U(\{l\}|h_1) = U(\{l\}|h_2) \) and \( x \sim_{h_2} \{l\} \). Then, (P2) implies that \( \{l\} \succ_{h_1} x \). Thus, \( U(x|h_1) \leq U(\{l\}|h_1) = U(\{l\}|h_2) = U(x|h_2) \).

Lemma G.3. Assume that \( \{\succ_h\}_{h \in H} \) satisfies all the axioms of Theorem 3.1 and admits a history-dependent random discounting representation \((u, \{\mu(\cdot|h)\}_{h \in H})\). Then (P1) holds if and only if (ICX) \( \mu(\cdot|h_2) \) is smaller than \( \mu(\cdot|h_1) \) in the increasing convex order.

Proof. (P1) \( \implies \) (ICX). We show that, for all continuous, increasing and convex functions \( v \) of \( \alpha \), there is a sequence \( \{v_n\} \) of functions such that \( v \geq v_n \),

\[
\sup_{\alpha} |v(\alpha) - v_n(\alpha)| < \frac{1}{n}, \quad \text{and} \quad \int v_n(\alpha) \, d\mu(\alpha|h_2) \leq \int v_n(\alpha) \, d\mu(\alpha|h_1)
\]

for all \( n = 1, 2, \ldots \). Then the result follows from the dominated convergence theorem.

Let \( v : [0,1] \to \mathbb{R} \) be a continuous increasing convex function. Then, for every \( \hat{\alpha} \in [0,1] \), there exists a vector \( (p_{\hat{\alpha},1}, p_{\hat{\alpha},2}) \in \mathbb{R}^2 \) such that \( p_{\hat{\alpha},2} \geq p_{\hat{\alpha},1} \) and for all \( \alpha \in [0,1] \),

\[
v(\alpha) \geq (1-\alpha)p_{\hat{\alpha},1} + \alpha p_{\hat{\alpha},2}
\]

with equality for \( \hat{\alpha} \).

Fix \( n \). Since \( v(\alpha) - \{(1-\alpha)p_{\hat{\alpha},1} + \alpha p_{\hat{\alpha},2}\} \) is continuous with respect to \( \alpha \), there exists an open neighborhood \( B(\hat{\alpha}) \) of \( \hat{\alpha} \) such that for every \( \alpha \in B(\hat{\alpha}) \)

\[
0 \leq v(\alpha) - \{(1-\alpha)p_{\hat{\alpha},1} + \alpha p_{\hat{\alpha},2}\} < \frac{1}{n}.
\]

It follows from the compactness of \([0,1]\) that there exists a finite set \( \{\hat{\alpha}_i\}_{i=1}^M \subset [0,1] \) such that \( \{B(\hat{\alpha}_i)\}_{i=1}^M \) is a covering of \([0,1]\).

We define \( v_n : [0,1] \to \mathbb{R} \) by

\[
v_n(\alpha) = \max_{1 \leq i \leq M} [(1-\alpha)p_{\hat{\alpha}_i,1} + \alpha p_{\hat{\alpha}_i,2}].
\]

Then it is straightforward that \( v(\alpha) \geq v_n(\alpha) \) for every \( \alpha \in [0,1] \). Moreover, we see that

\[
\sup_{\alpha} |v(\alpha) - v_n(\alpha)| < \frac{1}{n}.
\]
In fact, pick an arbitrary $\alpha \in [0, 1]$. Then there is $j \in M$ such that $\alpha \in B(\hat{\alpha}_j)$. This implies

$$0 \leq v(\alpha) - v_n(\alpha) \leq v(\alpha) - \{(1 - \alpha)p_{\hat{\alpha}_j,1} + \alpha p_{\hat{\alpha}_j,2}\} < \frac{1}{n}.$$ 

Finally we will see that

$$\int v_n(\alpha) \, d\mu(\alpha|h_2) \leq \int v_n(\alpha) \, d\mu(\alpha|h_1).$$

Since $u(\Delta(C))$ is a closed nondegenerate interval, we can assume, without loss of generality, that there exist $\ell_i, \ell'_i \in \Delta(C)$, $i = 1, \cdots, M$, satisfying $u(\ell_i) = p_{\hat{\alpha}_i,1}$ and $u(\ell'_i) = p_{\hat{\alpha}_i,2}$. Thus $v_n$ is rewritten as

$$v_n(\alpha) = \max_i (1 - \alpha)u(\ell_i) + \alpha u(\ell'_i).$$

Now consider the menu $x^n = \{\ell_i \otimes \{\ell'_i\}| i = 1, \cdots, M\}$. Since $u(\ell'_i) = p_{\hat{\alpha}_i,2} \geq p_{\hat{\alpha}_i,1} = u(\ell_i)$, $\{\ell'_i\} \gtrsim_h \{\ell_i\}$ for all $h$. Thus, by definition, $x^n \in \mathcal{P}$. Then it follows from Lemma G.2 that

$$\int v_n(\alpha) \, d\mu(\alpha|h_2) = U(x^n|h_2) \leq U(x^n|h_1) = \int v_n(\alpha) \, d\mu(\alpha|h_1),$$

which completes the proof.

$(ICX) \implies (P1)$. First, we will claim that for all $x \in \mathcal{P}$, $U(x|h_2) \leq U(x|h_1)$. Let

$$v_x(\alpha) \equiv \max_{\ell \otimes (\otimes \ell') \in x} (1 - \alpha)u(\ell) + \alpha u(\ell').$$

Then, $v_x$ is convex. Moreover, since $u(\ell') \geq u(\ell)$, $v_x$ is increasing. Since $\mu(\cdot|h_2)$ is smaller than $\mu(\cdot|h_1)$ in the increasing convex order, we have

$$U(x|h_2) = \int v_x(\alpha) \, d\mu(\alpha|h_2) \leq \int v_x(\alpha) \, d\mu(\alpha|h_1) = U(x|h_1).$$

Now assume $U(x|h_2) > U(\{l\}|h_2)$ for some $x \in \mathcal{P}$ and $l \in \mathcal{L}_I$. By definition, $l$ is written as $\ell \otimes \{\otimes \ell'\}$ for some $\ell, \ell'$ with $\{\otimes \ell\} \gtrsim \{\otimes \ell'\}$. Notice that $\overline{\alpha}_{h_2} \leq \overline{\alpha}_{h_1}$ because of the increasing convex order. From the above claim and this observation,

$$U(x|h_1) \geq U(x|h_2) > U(\{l\}|h_2) = (1 - \overline{\alpha}_{h_2})u(\ell) + \overline{\alpha}_{h_2}u(\ell') \geq (1 - \overline{\alpha}_{h_1})u(\ell) + \overline{\alpha}_{h_1}u(\ell') = U(\{l\}|h_1),$$

as desired. \hfill \Box

**Lemma G.4.** Assume that $\{\gtrsim_h\}_{h \in H}$ satisfies all the axioms of Theorem 3.1 and admits a history-dependent random discounting representation $(u, \{\mu(\cdot|h)\}_{h \in H})$. Then (P2) holds if and only if $(ICV)$ $\mu(\cdot|h_2)$ is smaller than $\mu(\cdot|h_1)$ in the increasing concave order.

41
Proof. \((P2) \implies (ICV)\). We show that, for all continuous, decreasing and convex functions \(w\) of \(\alpha\), there is a sequence \(\{w_n\}\) of functions such that \(w \geq w_n\),

\[
\sup_\alpha |w(\alpha) - w_n(\alpha)| < \frac{1}{n}, \quad \text{and} \quad \int w_n(\alpha) \, d\mu(\alpha|h_1) \leq \int w_n(\alpha) \, d\mu(\alpha|h_2)
\]

for all \(n = 1, 2, \ldots\). Then the result follows from the dominated convergence theorem.

Let \(w : [0, 1] \to \mathbb{R}\) be a continuous decreasing convex function. Then, for every \(\hat{\alpha} \in [0, 1]\), there exists a vector \((p_{\hat{\alpha}, 1}, p_{\hat{\alpha}, 2}) \in \mathbb{R}^2\) such that \(p_{\hat{\alpha}, 1} \geq p_{\hat{\alpha}, 2}\) and for all \(\alpha \in [0, 1]\),

\[
w(\alpha) \geq (1 - \alpha)p_{\hat{\alpha}, 1} + \alpha p_{\hat{\alpha}, 2}
\]

with equality for \(\hat{\alpha}\).

By the same argument as Lemma G.3, there exists a finite set \(\{\hat{\alpha}_i\}_{i=1}^M \subset [0, 1]\) such that

\[
w_n(\alpha) = \max_{1 \leq i \leq M} [(1 - \alpha)p_{\hat{\alpha}_i, 1} + \alpha p_{\hat{\alpha}_i, 2}], \quad \text{and} \quad \sup_\alpha |w(\alpha) - w_n(\alpha)| < \frac{1}{n}.
\]

Finally we will see that

\[
\int w_n(\alpha) \, d\mu(\alpha|h_1) \leq \int w_n(\alpha) \, d\mu(\alpha|h_2).
\]

By the same argument as Lemma G.3, there exists a menu \(x^n = \{\ell_i \otimes \{\otimes \ell'_i\}|i = 1, \ldots, M\}\) such that \(u(\ell_i) = p_{\hat{\alpha}_{i, 1}} \geq p_{\hat{\alpha}_{i, 2}} = u(\ell'_i)\), or \(\otimes \ell_i \succeq_h \{\otimes \ell'_i\}\) for all \(h\). Thus, by definition, \(x^n \in Z_I\). Then it follows from Lemma G.2 that

\[
\int w_n(\alpha) \, d\mu(\alpha|h_1) = U(x^n|h_1) \leq U(x^n|h_2) = \int w_n(\alpha) \, d\mu(\alpha|h_2),
\]

which completes the proof.

\((ICV) \implies (P2)\). The result follows from the same argument as in \((ICX) \implies (P1)\). \(\Box\)

References


