Abstract

We axiomatize a class of recursive dynamic models that capture subjective constraints on the amount of information a decision maker can obtain, pay attention to, or absorb, via a Markov Decision Process for Information Choice ( MIC ). An MIC is a subjective decision process that specifies what type of information about the payoff relevant state is affordable in the current period, and how the choice of what to learn now affects what can be learned in the future. The constraint imposed by the MIC is identified from choice behavior up to a recursive extension of the Blackwell order. All the other parameters of the model, namely the anticipated evolution of the payoff relevant state, state dependent consumption utilities, and the discount factor are also uniquely identified.

Key Words: Dynamic Preferences, Recursive Information Constraints, Recursive Blackwell Dominance, Rational Inattention, Subjective Markov Decision Process, Familiarity Bias.

JEL Classification: D80, D81

1. Introduction

People acquire and react to information, and they often face constraints on the amount of information they can obtain, pay attention to, or simply absorb. For example, consumers cannot at all times be aware of relevant prices at all possible retailers (as is evident from

(*) We thank Attila Ambrus, Alexander Bloedel, Andrew Ellis, Itzhak Gilboa, Jay Lu, George Mailath, Pietro Ortoleva, Wolfgang Pesendorfer, Todd Sarver, and Ran Spiegler for helpful comments, and the National Science Foundation for support under awards SES 1461469 and SES 1461502. This work began while Dillenberger and Krishna were visiting the Economics departments at NYU and Duke University, respectively. They are most grateful to these institutions for their hospitality.

(†) University of Pennsylvania <dill@sas.upenn.edu>

(‡) Florida State University <rvk3570@gmail.com>

(§) Duke University <p.sadowski@duke.edu>
the proliferation of online comparison shopping engines) and firms have limited human
resources they can expend on market analysis. While accounting for such information
constraints can significantly change theoretical predictions (see, for instance, Stigler (1961),
Persico (2000), and the literature on rational inattention pioneered by Sims (1998, 2003)),
an inherent difficulty in modeling them, as well as the actual choice of information, is that
they are often private and unobservable to outsiders.

In this paper, we provide a dynamic model that incorporates intertemporal information
constraints and allows us to identify and quantify them from observable choice behavior.¹
Our framework unifies behavioral phenomena that arise in the presence of such constraints,
independently of their nature. For example, it applies whether (as is common in the literature
on rational inattention) the constraints are cognitive, so that even when information appears
to be readily available, individuals may have limited ability to take it into account; or
physical, where the constraint reflects the cost of acquiring information.

We axiomatize the behavior of a Decision Maker (henceforth $\text{DM}$) whose choice of
information is constrained by a subjective Markov Decision Process ($\text{MDP}$), which specifies
how future constraints depend on unobservable current and past choices of information.
$\text{MDPs}$ are the fundamental building blocks for dynamic programming. We focus on learning
via partitions of the space $S$ of payoff relevant states, where the state changes over time.
Formally, an $\text{MDP}$ for Information Choice ($\text{MIC}$) is parametrized by an $\text{MDP}$ state $\theta$, a function
$\Gamma(\theta)$ that determines the set of feasible partitions of $S$, and an operator $\tau$ that governs the
transition of $\theta$ in response to the choice of partition and the realization of $s \in S$. Examples
and a discussion of such $\text{MICs}$ are at the end of this section.

We show that from observing $\text{DM}$’s choice between appropriate infinite horizon deci-
sion problems, one can infer the entire set of parameters governing his preferences, namely
(i) state dependent utilities; (ii) (time varying) beliefs about the state $s \in S$; (iii) the dis-
count factor; and (iv) the $\text{MIC}$ up to a recursive extension of Blackwell dominance. Here,
identifying a subjective $\text{MDP}$ from behavior is our main conceptual contribution.

The domain on which choice is observable is the space of Recursive Anscombe-
Aumann Choice Problems ($\text{RACP}$s) that consists of menus of acts on $S$, where each act is a
state-contingent lottery that yields current consumption and a continuation $\text{RACP}$ (a new menu of acts for the next period).² Our representation suggests the following timing of
events and decisions, as illustrated in Figure 1. $\text{DM}$ enters a period facing an $\text{RACP}$ $x$ with a
prior belief $\pi_s$ over $S$ and an $\text{MDP}$ state $\theta$. He first chooses a partition $P \in \Gamma(\theta)$. For any
realization of a cell $I \in P$, $\text{DM}$ updates his beliefs using Bayes’ rule and chooses an act

---

¹ Recent literature on rational inattention has demonstrated how to identify information constraints from
observed choice data in static settings (see De Oliveira et al. (2016) and Ellis (2016)). We discuss the
relation with these papers and others in Section 5.

² For example, our domain can accommodate the standard consumption-investment problem, wherein $\text{DM}$
simultaneously chooses what to consume and how to invest his residual wealth, thereby determining the
consumption-investment choices available in the next period, contingent on the evolution of the stock
market.
f from the RACP x he faces. At the end of the period, the true state s' is revealed and DM
receives the lottery f (s'), which determines the current consumption c and an RACP y for
the next period. At the same time, a new MDP state θ' = τ(θ, P, s') and a new belief πs'
is determined for next period, where the latter is based on another subjectively perceived
Markov process that governs the transition of the state, starting from s'.

\[ V(x, θ, s) = \max_{P ∈ Γ(θ)} \sum_{I ∈ P} \left[ \max_{f ∈ x} \sum_{s' ∈ I} \pi_s(s') \left[ E^{f(s')} \left[ u_{s'}(c) + \delta V(y, τ(θ, P, s'), s') \right] \right] \right] \pi_s(I) \]

Figure 1: Timeline

DM’s objective is then to maximize the expected utility which consists of state-
dependent consumption utilities, \((u_s)\), and the discounted continuation value:

Theorem 1 establishes that \((u_s, π_s)\)s∈S and δ are essentially uniquely identified, and
that the remaining preference parameter, the mic, is identified up to the addition or deletion
of information choices that are dominated in terms of a recursive extension of Blackwell’s
comparison of informativeness.

Indeed, because the mic is subjective, all that matters for behavior are the information
choices it permits. Section 2.4 constructs a space of Recursive Information Constraints
(RICs), and shows that it is canonical in the sense that every constraint that an mic generates
corresponds to a unique RIC. The proof of Theorem 1 relies on a notion of duality — which
we term strong alignment — between the space of RICs and our domain of RACPs.3

Theorem 2 provides an axiomatic foundation for our model. In the sequel, we shall call
a preference over dynamic choice problems self-generating if it has the same properties as
each of the preferences over continuation problems that together generate it. Self-Generation
is satisfied in any recursive model precisely because it embodies the dynamic programming
principle. While our value function depends on the MDP state θ — and so is not stationary
— it is nonetheless recursive, meaning that preferences should satisfy the weaker notion of
Self-Generation. Because θ is subjective, observed behavior cannot condition on it. This

(3) The proof intuition provided in Section 2.5 clarifies how particular types of RICs rely on different
structural features of our domain to achieve alignment. These insights should be helpful when applying
our identification approach to other instances where agents face unobservable subjective decision processes,
such as in health or labor economics where agents control health or human capital stocks via often
unobservable actions.
necessitates, first, a statement of Self-Generation in terms of ex-ante preferences and, second, an investigation of other standard properties such as Independence and Time Separability, that are central to virtually all existing axiomatic models of dynamic choice. In addition to Self-Generation, our key (and novel) axioms formalize aspects of the standard properties that can be maintained even when dynamic choice depends on simultaneous information choice. That is, while our results establish subjective mics as a versatile tool that provides a unified view of many seemingly complicated preference patterns, the structure of our main axioms resembles that of standard assumptions, thereby facilitating the comparison between our model and existing dynamic choice models.

An example of a preference pattern that can be accommodated by our model and is difficult to reconcile in the absence of dynamic information constraints is familiarity bias, according to which individuals are reluctant to switch away from choice problems they are familiar with. In Section 4.1 we formally explore how a subjective information constraint can explain familiarity bias by allowing $\text{dm}$ to develop expertise in discerning specific events. As another example, Section 4.2 studies a simple search problem, where in each period an unemployed worker draws a wage from an iid distribution and needs to decide whether to accept the offer (and work forever at the accepted wage) or to keep searching. Unlike the fixed reservation wage prediction of the standard model, our model can accommodate a reservation wage that decreases over time, because the expected value of continuing the search decreases as the information constraint tightens, perhaps due to search-fatigue.

We now present some examples of mics.

**Example 1.1.** Let $S$ be a set of states. $\text{dm}$ receives in each period an attention ‘income’ $\kappa \geq 0$. Any stock of attention not used in the current period can be carried over to the next one at a decay rate of $\beta$. Let $K$ denote the attention stock in the beginning of a period. Learning the partition $P$ costs $c(\ P\ )$, for some cost function $c$ (measured in units of ‘attention’ and not utils), which is increasing in the fineness of the partition. For example, as is common in the rational inattention literature, $c(\ P\ )$ can be the entropy of $P$ calculated using some probability distribution over $S$. Formally, for attention stock $K$, any partition $P \in \Gamma(K) = \{P : c(\ P\ ) \leq K + \kappa\}$ can be chosen, whereupon the stock transitions to $\tau(K, P) = \beta[K + \kappa - c(\ P\ )]$ to determine the continuation constraint.

**Example 1.2** (Example 1.1 continued). The cost of learning a partition depends on past choices. In particular, if partition $Q$ was chosen yesterday, then the cost of learning $P$ today is $c(P \mid Q) = (1 - b)H_\mu(P) + bH_\mu(P \mid Q)$, where, given a probability $\mu$ over $S$, $H_\mu(P)$ is the entropy of $P$ and $H_\mu(P \mid Q)$ is the relative entropy of $P$ with respect to $Q$. Note that $H_\mu(P \mid P) = 0$ and hence $c(P, P) = (1 - b)H_\mu(P)$. That is, while learning $P$ initially costs $H_\mu(P)$, learning $P$ again in the subsequent period costs only a fraction $(1 - b)$ thereof. The parameter $b$ measures the degree to which $\text{dm}$ can gain expertise. (Note that if $\mu$ evolves over time it becomes part of the $\text{MDP}$ state space.)
Example 1.3 (Expertise). This mic also captures expertise. From a set of possible experiments, each of which corresponds to a different partition of $S$, $dm$ can set up at most $k$ at a rate of one new experiment per period. Once an experiment is set up, it can be carried out every period. Formally, let $\mathcal{P}_{\text{exp}} \subset P$ be the space of all partitions of $S$ that correspond to a possible experiment. If $dm$ has chosen partitions $P_1, \ldots, P_n$ in the past, his current access to information is given by the partition $P := P_1 \land \cdots \land P_n$ (where $P \land Q$ denotes the coarsest refinement of $P$ and $Q$). Then, using $(P, n)$ as an MDP state, the constraints on information choice and the subsequent transitions are given by

$$
\Gamma(P, n) := \begin{cases}
\{P \land Q : Q \in \mathcal{P}_{\text{exp}}\} & n < k \\
\{P\} & n = k
\end{cases}
$$

and

$$
\tau((P, n), Q, s) := \begin{cases}
(P \land Q, n + 1) & Q \neq P \\
(P, n) & Q = P
\end{cases}
$$

where the initial MDP state is $\{(S), 0\}$.

Example 1.4. $dm$ cannot acquire information in two consecutive periods. If he has learned a non-trivial partition of $S$ in the previous period, he cannot afford to learn anything (i.e., he can only learn the trivial partition of $S$) today.

Example 1.5 (State dependence). The feasible set of partitions at any period solely depends on the realization of the state in the previous period.

Example 1.6 (Resource exhaustion). $dm$ is endowed with an initial attention stock $K$, which he draws down every time he chooses to learn.

The rest of the paper is organized as follows. In Section 2 we introduce the analytical framework, state our utility representation and establish our identification result. Section 3 provides a behavioral foundation, namely the axioms and representation theorem. Section 4 gives examples of the patterns of behavior that our model can address. Section 5 surveys related decision-theoretic literature, while other related literature is discussed in the relevant sections. Section 6 discusses some conceptual issues and concludes. Proofs can be found in the Appendix, while additional technical details are in the Supplementary Appendix.

---

(4) For example, a growing start-up may be able to slowly hire up to $k$ experts with different specialized understanding of relevant markets.

(5) For example, paying attention may cause fatigue, which diminishes the ability to pay further attention. Therefore, periods in which individuals pay careful attention are usually followed by periods in which they should rest. Similarly, acquiring information may consume time or physical resources and thus crowd out the completion of other essential tasks; those tasks then have to be performed in consecutive periods, when they, in turn, crowd out further acquisition of information.

(6) This type of mic is reminiscent of the ‘willpower depletion’ model of Ozdenoren, Salant, and Silverman (2012) in which $dm$ is initially endowed with a willpower stock and depletes his willpower whenever he limits his rate of consumption.

(7) The Supplementary Appendix, Dillenberger, Krishna, and Sadowski (2016b), is available online at <...>.
2. Representation with Subjective Information Constraints

2.1. Domain

Let $S$ be a finite set of objective or observable states, and let $C$ be a compact metric space, representing consumption. For any compact metric space $Y$, we denote by $\mathcal{K}(Y)$ the space of closed and non-empty subsets of $Y$. The space of Recursive Anscombe-Aumann Choice Problems (racps) is $X \simeq \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$. On the space $X$, we let $tx + (1-t)y := \{tf + (1-t)g : f \in x, g \in y\}$.

Intuitively, each $x \in X$ is a menu of acts on $S$, where each act yields a state-dependent lottery over instantaneous consumption and continuation racps (ie, a new $y \in X$) for the next period. A consumption stream is a degenerate racp in that it does not involve choice at any point in time. The space $L$ of all consumption streams can be written recursively as $L \simeq \mathcal{F}(\Delta(C \times L))$. There is a natural embedding of $L$ in $X$. We analyze a preference, $\succ$, which is a binary relation on $X$. We denote its restriction to $L$ by $\succ |_L$.

The space $X$ of racps subsumes some domains previously studied in the literature. For instance, if $S$ is a singleton, $X$ reduces to the domain considered by Gul and Pesendorfer (2004). Furthermore, if the horizon is also finite, it reduces to the domain in Kreps and Porteus (1978). The space $X$ of racps is therefore descriptively appealing, as they argue. The subspace $L$ of consumption streams is also a subspace of the domain in Krishna and Sadowski (2014), although the latter is distinct from the space of racps considered here.

2.2. mic-Representation

When faced with an racp, dm chooses a partition in every period. Let $\mathcal{P}$ be the space of all partitions of $S$. dm’s choice of partition is constrained by an mic. Formally, an mic is a tuple $\mathcal{M} = (\Theta, \theta_0, \Gamma, \tau)$, where $\Theta$ is a set of mdp states; $\theta_0$ is the initial state; $\Gamma : \Theta \to 2^\Theta \setminus \emptyset$ is a set of feasible partitions in a given mdp state $\theta$; and $\tau : \mathcal{P} \times \Theta \times S \to \Theta$ is a transition operator that denotes the consequences of a particular choice of partition in the mdp state $\theta$ and objective state $s$. Let $\mathcal{M}$ be the space of mics.

In addition, let $(u_s)_{s \in S}$ be a collection of (real-valued) continuous functions on $C$ such that at least one $u_s$ is non-trivial (ie, non-constant), and let $\delta \in (0, 1)$ be a discount factor. Let $\Pi$ be a fully connected transition operator\(^9\) for a Markov process on the space $S$, where $\Pi(s, s') =: \pi_s(s')$ is the probability of transitioning from state $s$ to state $s'$. Let $s_0 \notin S$ be an auxiliary state, and denote by $\pi_{s_0}$ the unique invariant measure of $\Pi$.

We consider the following utility representation of $\succ$ on the space of racps.

---

(8) Formally, $X$ is linearly homeomorphic to $\mathcal{K}(\mathcal{F}(\Delta(C \times X)))$. See Appendix A.2 for details.

(9) The transition operator $\Pi$ is fully connected if $\Pi(s, s') > 0$ for all $s, s' \in S$. 

---
Definition 2.1. A preference $\succeq$ on $X$ has a mic-representation $((u_s)_{s \in S}, \delta, \Pi, \mathcal{M})$ if the function $V(\cdot, \theta_0, s_0) : X \to \mathbb{R}$ represents $\succeq$, where $V : X \times \Theta \times (S \cup \{s_0\}) \to \mathbb{R}$ satisfies

$$V(x, \theta, s) = \max_{P \in \Gamma(\theta)} \sum_{I \in P} \left[ \max_{s' \in I} \sum_{f \in x} \mathbb{E}^{f(s')} \left[ u_{s'}(c) + \delta V(y, \tau(P, \theta, s'), s') \right] \pi_{s'}(s' | I) \right] \pi_{s}(I) \tag{Val}$$

In the representation above, for each $s' \in S$, $f(s') \in \Delta(C \times X)$ is a probability measure over $C \times X$ (with the Borel $\sigma$-algebra), so that $\mathbb{E}^{f(s')}$ is the expectation with respect to this probability measure.\textsuperscript{10}

A dynamic information plan prescribes a choice of $P \in \Gamma(\theta)$ for each tuple $(x, \theta, s)$. Thus, an mic describes the set of feasible information plans available to $\text{dm}$. The next proposition ensures the existence of the value function and an optimal dynamic information plan.

Proposition 2.2. Each mic-representation $((u_s)_{s \in S}, \delta, \Pi, \mathcal{M})$ induces a unique function $V : X \times \Theta \times S \cup \{s_0\} \to \mathbb{R}$ that is continuous on $X$ and satisfies [Val]. Moreover, an optimal dynamic information plan exists.

A proof is in Appendix A.4.

2.3. Identification

The space of mics has a natural order. To see this, consider first two sets of partitions $\{P_1, \ldots, P_m\}$ and $\{Q_1, \ldots, Q_n\}$. If for every $Q_i$ there is a $P_j$ that is finer than it, we say that $\{P_1, \ldots, P_m\}$ setwise Blackwell dominates $\{Q_1, \ldots, Q_n\}$. In this sense, the finest partition setwise Blackwell dominates every other set of partitions. The same notion can be extended to multiple periods and, in a natural way, to the space of all mics.

Because mics have a recursive structure, so too does our definition of the recursive Blackwell order, which is the largest order that satisfies the following: For all $\mathcal{M}, \mathcal{M}^\dagger \in \mathcal{M}$, $\mathcal{M}$ dominates $\mathcal{M}^\dagger$ if for all $P^\dagger \in \Gamma^\dagger(\theta_0^\dagger)$ there is $P \in \Gamma(\theta_0)$ such that (i) $P$ is finer than $P^\dagger$, and (ii) $(\Theta, \tau(P, \theta_0, s), \Gamma, \tau)$ dominates $(\Theta^\dagger, \tau^\dagger(P^\dagger, \theta_0^\dagger, s), \Gamma^\dagger, \tau^\dagger)$ for all $s \in S$. Propositions A.4 and A.11 in the appendices imply that the recursive Blackwell order is a well defined preorder (ie, is reflexive and transitive).

\textsuperscript{10} One of the central properties of dynamic choice is dynamic consistency, which requires $\text{dm}$’s ex post preferences to agree with his ex ante preferences over plans involving the contingency in question. Because our primitive is ex ante choice between racps, we cannot investigate dynamic consistency directly in terms of behavior. However, our representation [Val] describes behavior as the solution to a dynamic programming problem with state variables $(x, \theta, s)$, so that implied behavior is dynamically consistent contingent on those state variables. The novel aspect is that the mdp state $\theta$ is controlled by $\text{dm}$ and is not observed by the analyst.
**Theorem 1.** Let \((u_s, \delta, \Pi, \mathcal{M})\) be an \(\text{mic}\)-representation of \(\succeq\). Then, the functions \((u_s)_{s \in S}\) are unique up to the addition of constants and a common scaling, \(\delta\) and \(\Pi\) are unique, and \(\mathcal{M}\) is unique up to recursive Blackwell equivalence.\(^{(11)}\)

By the theorem, we can identify a preference \(\succeq\) that has a \(\text{mic}\)-representation with the tuple of parameters \((u_s)_{s \in S}, \delta, \Pi, \mathcal{M}\) that generates this representation. An immediate benefit of identifying all the parameters is that it allows a meaningful comparison of people.

Consider, then, two decision makers with preferences \(\succeq\) and \(\succeq'\), respectively. We say that \(\succeq\) has a greater affinity for dynamic choice than \(\succeq'\) if for all \(x \in X\) and \(\ell \in L\), \(x \succeq' \ell\) implies \(x \succeq \ell\).\(^{(12)}\) The comparison in the definition implies that \(\succeq\) and \(\succeq'\) have the same ranking over consumption streams in \(L\).\(^{(13)}\) A typical \(\text{RACP}\) \(x\) may allow \(\text{dm}\) to wait for information before making a choice. This option should be more valuable the more information plans \(\text{dm}\)'s \(\text{mic}\) renders feasible. The uniqueness established in Theorem 1 allows us to formalize this intuition.

**Proposition 2.3.** Let \((u_s, \delta, \Pi, \mathcal{M})\) and \((u_s', \delta', \Pi', \mathcal{M}')\) be \(\text{mic}\)-representations of \(\succeq\) and \(\succeq'\) respectively. The preference \(\succeq\) has a greater affinity for dynamic choice than \(\succeq'\) if, and only if, \(\Pi = \Pi', \delta = \delta', (u_s)_{s \in S}\) and \((u_s')_{s \in S}\) are identical up to the addition of constants and a common scaling, and \(\mathcal{M}\) recursively Blackwell dominates \(\mathcal{M}'\).

A proof is in Appendix B. The Proposition connects a purely behavioral comparison of \(\text{mics}\) to recursive Blackwell dominance, which is independent of preferences, and hence of utilities and beliefs. This indicates a duality between our domain of choice, and the information constraints that can be generated by \(\text{mics}\).

In the remainder of this section, we discuss the main ideas behind Theorem 1 and its proof, emphasizing aspects that might generalize beyond the context of our specific model. Towards that end, Section 2.4 introduces the space of \(\text{ric}\)s which is the canonical space of information constraints, and Section 2.5 formalizes the duality, which we call strong alignment, between the space of \(\text{ric}\)s and the domain \(X\) of \(\text{RACPs}\).

### 2.4. Recursive Information Constraints

\(\text{mics}\) are subjective, and so all that matters for behavior are the information choices they permit. This is captured by the following notion: The \(\text{mics}\) \(\mathcal{M}\) and \(\mathcal{M}'\) are indistinguishable if they have the same choices of partition in the first period and, for any choice in the first period, the values of the utility functions change by a constant for all states.

---

\(^{(11)}\) In other words, for any additional representation of \(\succeq\) with parameters \((u_s', \delta, \Pi, \mathcal{M}')\), it is the case that \(u_s' = au_s + b_s\), for some \(a > 0\) and \(b_s \in \mathbb{R}\) for each \(s \in S\), and that \(\mathcal{M}\) and \(\mathcal{M}'\) recursively Blackwell dominate each other.

\(^{(12)}\) This definition is the analogue of notions of ‘greater preference for flexibility’; see for example Dekel, Lipman, and Rustichini (2001), Krishna and Sadowski (2014), and Dillenberger et al. (2014).

\(^{(13)}\) That is, \(\ell \succeq \ell'\) if, and only if, \(\ell \succeq' \ell'\) for all \(\ell, \ell' \in L\). This is Lemma 34 in Appendix F of Krishna and Sadowski (2014), and assumes that both \(\succeq\) and \(\succeq'\) satisfy Independence on \(L\).
period, the same choices in the second period, and so on. Intuitively, indistinguishable mics differ only up to a relabeling of the mdp states, and up to the addition of mdp states that can never be reached. This definition of indistinguishability is formalized in Appendix A.6, and leads to a recursive characterization described in Lemma A.8.

It is convenient to consider canonical mics that are defined on an mdp state space which is compact and metrisable. We now describe the construction of \( \hat{\Omega} \). Suppose \( \delta m \) can choose from the set of partitions \( \{P_1^{(1)}, \ldots, P_m^{(1)}\} \) in the first period. Also suppose that upon choosing the partition \( P_j^{(1)} \) in the first period, and after the realization of the state \( s^{(1)} \), a set of partitions \( \{P_1^{(2)}, \ldots, P_k^{(2)}\} \) is available in the second period. If this process proceeds to infinity, with each history of choices and realized states determining the set of feasible partitions, we get the space \( \hat{\Omega} \).

The description above suggests a recursive way to think of \( \hat{\Omega} \): Each \( \omega \in \Omega \) describes the set of feasible partitions available for choice in the first period, and how a choice of partition \( P \) and the realized state \( s \) determine a new \( \omega' \in \Omega \) in the next period. That is, \( \omega \) is a finite collection of pairs \( (P, \omega') \), where \( \omega' = (\omega'_s)_{s \in S} \). In other words, \( \hat{\Omega} \) is isomorphic to \( K_\delta(\mathcal{P} \times \Omega^S) \). We call \( \hat{\Omega} \) the space of Recursive Information Constraints (ric) so that each \( \omega \) of \( \hat{\Omega} \) is an rric. Conversely, every rric is also an mic. (Indeed, set \( \Gamma^*(\omega) = \{P : (P, \omega') \in \omega\} \) and \( \tau^*(\omega, P, s) = \omega'_s \) to obtain the mic \( M_\omega = (\Omega, \Gamma^*, \tau^*, \omega) \) which is indistinguishable from \( \omega \).)

**Proposition 2.4.** The space \( M \) of mic is isomorphic to \( \hat{\Omega} \) in the following sense.

(a) Every \( M \in M \) is indistinguishable from a unique \( \omega \in \hat{\Omega} \).

(b) Every \( \omega \in \hat{\Omega} \) induces an \( M_\omega \in M \) that is indistinguishable from \( \omega \).

Viewing \( \hat{\Omega} \) as the canonical state space for mic implies that the recursive Blackwell order on \( \hat{\Omega} \) is the unique recursive order so defined that is continuous and reflexive (see Proposition A.4 in the appendix).

When considering a representation with a canonical mic \( \omega \) (that is, an rric), we often write [Val] as

\[
V(x, \omega, s) = \max_{(P, \omega') \in \omega} \sum_{I \in P} \left[ \max_{f \in X} \sum_{s' \in I} E^{f(s')} \left[ u_{s'}(c) + \delta V(y, \omega'_s, s') \right] \pi_{s'}(s' | I) \right] \pi_s(I)
\]

By Proposition 2.4, it is sufficient to establish the identification in Theorem 1 for rics, which we describe next.

---

(14) In particular, the isomorphism is a homeomorphism. A formal construction of \( \hat{\Omega} \) is given in Appendix A.3. For metric spaces \( X \) and \( Y \), \( K_\delta(X \times Y) \) is the space of all non-empty closed subsets of \( X \times Y \) with the property that the subset contains distinct \((x, y)\) and \((x', y')\) only if \( x \neq x' \). The metric on \( \hat{\Omega} \) can be intuitively described as follows: Consider \( \omega, \omega' \in \hat{\Omega} \) as mic. If they differ in the set of feasible partitions only after \( n \) periods, regardless of the choice and the realized state in the first \( n \) periods, then the distance between \( \omega \) and \( \omega' \) is at most \( 1/2^n \). Thus, \( \omega, \omega' \in \hat{\Omega} \) are indistinguishable if, and only if, they are identical.
2.5. Strong Alignment between RACPs and RCIs

Notice that on the subdomain \( L \), \( V \) is independent of \( \omega \) and satisfies Independence. Indeed, \( V \) is then completely characterized by the parameters \( ((u_\omega), \Pi, \delta) \), meaning that \( \succcurlyeq \mid_L \) has a Recursive Anscombe-Aumann (RAA) representation on \( L \), with these parameters. Krishna and Sadowski (2014) show that an RAA representation on \( L \) exists and is unique up to the addition of constants and a common scaling of \((u_\omega)\).\(^{15}\) What remains then is to identify the RIC \( \omega \).

We call an RIC simple if it is of the form \((P, \omega)\), i.e., if it offers no information choice in the first period. Consider RICS \( \omega, \omega' \in \Omega \). An RACP \( x \in X \) separates \( \omega \) and \( \omega' \) if \( V(x, \omega, \pi_0) > V(x, \omega', \pi_0) \). An RACP \( x \in X \) is strongly aligned with a simple RIC \((P, \omega)\) if (i) \( V(x, (P, \omega), \pi_0) \geq V(x, \omega', \pi_0) \) for all \( \omega' \) and any \( \pi_0 \), and (ii) \( \omega' \) does not recursively Blackwell dominate \( (P, \omega) \) implies \( x \) separates \((P, \omega)\) and \( \omega' \).

The notion of strong alignment can be extended to general RCIs as follows. A finite set \( F_\omega \subset X \) of RACPs is uniformly strongly aligned with \( \omega \) if (i) \( V(x, \omega, \cdot) \geq V(x, \omega', \cdot) \) for all \( x \in F_\omega \) and \( \omega' \in \Omega \), and (ii) \( \omega' \) does not recursively Blackwell dominate \( \omega \) implies there exists \( x' \in F_\omega \) such that \( x' \) separates \( \omega \) and \( \omega' \). To prove Theorem 1, it suffices to show that each \( \omega \) has a set \( F_\omega \) of RACPs that is uniformly strongly aligned with it.

Consider some \((P, \omega') \in \omega \). We will construct an RACP \( x(P, \omega') \) that is strongly aligned with it, so that \((P, \omega')\) is an optimal information choice in the first period. The collection \( F_\omega := \{x(P, \omega') : (P, \omega') \in \omega\} \) is then uniformly strongly aligned with \( \omega \). We now sketch the construction of \( x(P, \omega') \).

For every \( s \in S \), let \( c_s^+ \) and \( c_s^- \) denote, respectively, the best and worst consumption outcomes under \( u_s \), and let \( \ell^* \) and \( \ell_* \) denote the consumption streams that deliver, respectively, \( c_s^+ \) and \( c_s^- \) at every date in state \( s \).\(^{16}\) Clearly, \( \ell^* \) is the best consumption stream while \( \ell_* \) is the worst one, and for all \( x \in X \), \( V(\ell^*, \cdot, \cdot) \geq V(x, \cdot, \cdot) \geq V(\ell_*, \cdot, \cdot) \). (This reflects the fact that the value of information is purely instrumental.)

Let \( \hat{\omega} \) be the RIC that delivers the coarsest partition in every period, offering no choice at all. Every other RIC recursively Blackwell dominates \( \hat{\omega} \).\(^{17}\) Consider now an RIC \( \omega_1 \) that only has non-trivial choice of partition in the first period, and provides the continuation RIC \( \hat{\omega} \) in the beginning of the second, regardless of first period choice of partition. Then, for each \((P, \hat{\omega}) \in \omega_1 \), define the RACP \( x_1(P, \hat{\omega}) \) as

\[
\left[\bullet\right] \quad x_1(P, \hat{\omega}) := \{f_1, J : J \in P\} \quad \text{where} \quad f_1, J(s) := \begin{cases} (c_s^+, \ell^*) & s \in J \\ (c_s^-, \ell_*) & s \notin J \end{cases}
\]

\(^{15}\) More precisely, Krishna and Sadowski (2014) establish the uniqueness of an RAA representation when the consumption space \( C \) is finite. We establish in Proposition I.5 of the Supplementary Appendix that the existence and uniqueness of the RAA representation also holds when \( C \) is a compact metric space.

\(^{16}\) Recall that we only require that some \( u_s \) be non-trivial which allows for the possibility that \( c_s^+ = c_s^- \) for all but one \( s \in S \).

\(^{17}\) This follows immediately from Proposition A.4 in the Appendix.
Given \( x_1(P, \hat{\omega}) \), no choice of partition can give a higher utility than picking \( P \), that is, \( V(x_1(P, \hat{\omega}), (P, \hat{\omega}), s) = V(\ell^*, (P, \hat{\omega}), s) \geq V(x', \omega', s) \) for all \( x' \) and \( \omega' \). In particular, \( V(x_1(P, \hat{\omega}), \omega_1, s) \geq V(x_1(P, \hat{\omega}), \omega', s) \) for all \( \omega' \in \Omega \). Moreover, if \( \omega' \) does not recursively Blackwell dominate \( \omega_1 \), then it must be that there exists \( (P, \hat{\omega}) \in \omega_1 \) such that for no \( (Q, \omega) \in \omega' \) is \( Q \) finer than \( P \). (This is the only possibility because, as noted above, every ric recursively Blackwell dominates \( \hat{\omega} \).) It is now straightforward to verify that \( x_1(P, \hat{\omega}) \) separates \( \omega_1 \) and \( \omega' \), and hence \( x_1(P, \hat{\omega}) \) is strongly aligned with \((P, \hat{\omega})\). The collection of menus \( F_{\omega_1} := \{x_1(P, \hat{\omega}) : (P, \hat{\omega}) \in \omega_1\} \) is then uniformly strongly aligned with \( \omega_1 \).

Our proof builds on this idea to construct an \( F_{\omega_2} \) that is uniformly strongly aligned with the \( \text{ric} \) \( \omega_2 \), which has non-trivial choice for only two periods, ie, the first period’s choice of partition results in a one-period \( \text{ric} \) of the form considered above. Given this extension, we can then proceed inductively to achieve strong alignment for any \( \text{ric} \) where a non-trivial choice of partition is allowed for only finitely many periods. Finally, we observe that any \( \omega \in \Omega \) can be approximated by a sequence of such \( \text{ric} \)s.

In the rest of this section, we will illustrate the construction of \( F_{\omega_2} \). Readers not interested in the precise details can skip to Remark 2.5 at the end of this section, which summarizes aspects of our construction that could be useful in other settings.

Consider the following example, based on partitions \( P^a, \ldots, P^d \) and \( Q^a, \ldots, Q^e \) which are strictly ordered, in terms of fineness, as follows (and are not ordered otherwise):

\[
P^a \rightarrow P^d \rightarrow P^c \quad Q^c \rightarrow Q^a \rightarrow Q^d \rightarrow Q^e
\]

where \( P^b \rightarrow P^d \), for instance, indicates that \( P^b \) is strictly finer than \( P^d \).

Define the one-period \( \text{ric} \)s \( \omega_1^a := \{(Q^a, \hat{\omega})\} \), \( \omega_1^b := \{(Q^b, \hat{\omega})\} \), \( \omega_1^c := \{(Q^c, \hat{\omega})\} \), and \( \omega_1^d = \{(Q^d, \hat{\omega}), (Q^e, \hat{\omega})\} \) with \( \hat{\omega} \) as defined above. Define the two-period \( \text{ric} \)s \( \omega_2 := \{(P^a, \omega_1^a), (P^b, \omega_1^b), (P^d, \omega_1^d)\} \) and \( \omega_2 := \{(P^a, \omega_1^a), (P^b, \omega_1^b), (P^c, \omega_1^c), (P^d, \omega_1^d)\} \), where \( \omega_1^s \) for \( i \in \{a, \ldots, d\} \) for all \( s \in S \). These \( \text{ric} \)s are displayed in Figure 2. Notice that neither \( \omega_2 \) nor \( \hat{\omega}_2 \) recursively Blackwell dominates the other.

One might think that \( \text{DM} \) would always be better off with \( \hat{\omega}_2 \) as an \( \text{ric} \) instead of \( \omega_2 \). After all, given \( \omega_2 \), the plan that says ‘Pick \( P^d \) in the first period, and \( Q^d \) in the second’ is weakly dominated (in the sense of being less valuable for every choice problem and strictly less for some choice problem) by the plan ‘Pick \( P^a \) in the first period and \( Q^a \) in the second’, and the latter is feasible under both, \( \omega_2 \) and \( \hat{\omega}_2 \). Similarly, the plan ‘Pick \( P^d \) in the first period, and \( Q^e \) in the second’ is dominated by ‘Pick \( P^b \) in the first period and \( Q^b \) in the second’.

Nevertheless, there are \( \text{racps} \) for which \( \hat{\omega}_2 \) is not as valuable to \( \text{DM} \) as \( \omega_2 \) because, under \( \omega_2 \), choosing \( (P^d, \omega_1^d) \) in the first period allows \( \text{DM} \) to wait until the second period and for intervening uncertainty to resolve, before making a choice of partition for the second

\(^{18}\) The construction of \( x_1(P, \hat{\omega}) \) is as in Theorem 1 of Laffont (1989, p59), which concerns a static setting.
period. Define the two-period rACP $x_2(P^d, \omega_1^d)$ as

$$x_2(P^d, \omega_1^d) := \{ f_{2,J}: J \in P^d \}$$

where

$$f_{2,J}(s) := \begin{cases} (e^+ s, \text{Unif} ((x_1(Q^d, \omega), (x_1(Q^e, \omega)))) & s \in J \\ \ell_e(s) & s \notin J \end{cases}$$

where Unif $((x_1(Q^d, \omega), (x_1(Q^e, \omega)))$ is the equiprobable lottery over the rACPs $x_1(Q^d, \omega)$ and $x_1(Q^e, \omega)$ defined in [4], and similarly define the two-period rACPs $x_2(P^a, \omega_1^a)$ and $x_2(P^b, \omega_1^b)$.

We claim that the set of rACPs $F_{\omega_2} := \{ x_2(P^a, \omega_1^a), x_2(P^b, \omega_1^b), x_2(P^d, \omega_1^d) \}$ is uniformly strongly aligned with $\omega_2$. For example, $x_2(P^d, \omega_1^d)$ separates $\omega_2$ and $\tilde{\omega}_2$. To see this, note that $V(x_2(P^d, \omega_1^d), (P^d, \omega_1^d), s) = V(\ell^*, (P^d, \omega_1^d), s)$ for all $x'$ and $\omega'$. In words, the plan ‘Pick $P^d$ in the first period and then pick $Q^d$ or $Q^e$ in the second when facing $x_1(Q^d, \omega)$ or $x_1(Q^e, \omega)$, respectively’ is at least as good as any other plan.

Now compare the value of $x_2(P^d, \omega_1^d)$ for the three feasible plans under $\tilde{\omega}_2$. ‘Pick $P^c$ followed by $Q^c$’ generates lower expected first period payoffs than $\ell^*$ because $P^d$ is finer than $P^c$. Both, ‘Pick $P^a$ followed by $Q^a$’ and ‘Pick $P^b$ followed by $Q^b$’ determine the second period partition in period 1. With either of these plans, DM will end up with a partition in the second period that is not (weakly) finer than one of the partitions $Q^d$ or $Q^e$, but still receives, with probability $\frac{1}{2}$, an rACP that is strongly aligned with either $Q^d$ or $Q^e$. Our earlier (static) argument shows that this entails a loss of utility, relative to $\ell^*$.

Similarly, $F_{\tilde{\omega}_2} := \{ x_2(P^a, \omega_1^a), (P^b, \omega_1^b), (P^c, \omega_1^c) \}$ is uniformly strongly aligned with $\tilde{\omega}_2$. In particular, $x_2(P^c, \omega_1^c)$ is aligned with $(P^c, \omega_1^c)$ and thus separates $\tilde{\omega}_2$ and $\omega_2$. These arguments can be extended to arbitrary rICs. Because each rIC $\omega$ affords only finitely many choices of partitions in the first period, it follows immediately that there is a finite set $F_\omega$ of rACPs that is uniformly strongly aligned with it.

The uniform strong alignment between $\omega \in \Omega$ and the finite set of rACPs $F_\omega \subset X$ amounts to a notion of duality between $X$ and $\Omega$. The example illustrates two central features.
of this duality. First, because rics may require dm to trade off coarser partitions at one date with finer partitions at another date, it is essential that racps allow for sufficient variation in acts at different dates. Second, because rics can accommodate information plans that allow dm to delay the choice of partition until a later date, racps must feature temporal lotteries\(^\text{(19)}\) which constitute truly dynamic choice, in that dm can wait for uncertainty to resolve at a particular date before he chooses an alternative.

**Remark 2.5.** Our insights are potentially valuable for other environments where dm faces dynamic choice problems, and has a collection of subjective (unobservable) dynamic plans that an analyst would like to infer. Consider another environment where \(X\) represents the space of dynamic choice problems, each \(\omega\) represents the collection of plans available to dm, and \(V(x, \omega)\) (which may also depend on objective state variables) is a function that evaluates the choice problems under the best plan from \(\omega\). In such a setting, dominance of plans is easy to define in terms of the value function, and so then is the ordering of sets of plans (just as we defined it for rics in Section 2.4). The following observations apply in all such settings.

- The set \(\omega\) can only be inferred up to the deletion or addition of dominated plans, though there may not always be a preference independent notion of dominance, such as the recursive Blackwell dominance in our environment.
- One can define the notion of *strong alignment* between a choice problem \(x\) and a set of plans \(\omega\). Clearly, identification of the set \(\omega\) is possible if, and only if, there is a dynamic choice problem \(x\) that is strongly aligned with it.
- If the space of plans allows for subjective choice at later dates, so that plans are truly dynamic, then temporal lotteries over choice problem are necessary for identification. Put differently, if dm has the flexibility to make subjective choices after the first period, then such plans have value, and can therefore be identified only if the dynamic choice space itself consists of dynamic stochastic control problems.

### 3. Axioms

In this section we introduce our axioms on the preference \(\succeq\) over \(X\) and state our representation theorem. The axioms broadly fall into three different categories: Axioms 1 and 3–5 do not rely on the recursive structure of our domain; they simply restrict preferences on \(\mathcal{H}(\mathcal{F}(\Delta(C \times X)))\), ignoring the fact that \(X\) is itself again the domain of our preferences (that is, \(X \simeq \mathcal{H}(\mathcal{F}(\Delta(C \times X)))\)). Axiom 2 imposes assumptions on \(\succeq|_L\), the restriction of \(\succeq\) to the set of consumption streams, \(L\). The subdomain \(L\) is special because it includes no consumption choice to be made in the future, which renders information choice inconsequential. Only Axiom 6 exploits the recursive structure of \(X\).

\(^{(19)}\) In a temporal lottery, introduced in Kreps and Porteus (1978), uncertainty resolves over time.
Section 3.1 contains standard assumptions collected in Axioms 1 and 2. The motivation for our more novel axioms is based on the type of learning process we envision, where in each period, \( d_m \) is constrained in his choice of partition, and takes into account that this choice will also determine the state-dependent continuation constraint for next period. Sections 3.2 to 3.5 discuss, in the following order, to what extent the standard properties of Separability, Strategic Rationality, Independence, and Stationarity are satisfied even if the analyst is not able to condition on information choice. Section 3.6 contains the representation result. Section 3.7 investigates the implications of further strengthening our notions of Stationarity and Separability, and also shows that imposing Independence implies that information is not determined by a choice process, but instead exogenously arrives over time.

### 3.1. Standard Properties

Our first axiom collects basic properties of \( \succeq \) that are common in the menu-choice literature.

**Axiom 1 (Basic Properties).**

(a) Order: \( \succeq \) is non-trivial, complete, and transitive.

(b) Continuity: The sets \( \{y : y \succeq x\} \) and \( \{y : x \succeq y\} \) are closed for each \( x \in X \).

(c) Lipschitz Continuity: There exist \( \ell^\#, \ell_* \in L \) and \( N > 0 \) such that for all \( x, y \in X \) and \( t \in (0, 1) \) with \( t \geq Nd(x, y) \), we have \( (1 - t)x + t \ell^\# > (1 - t)y + t \ell_* \).

(d) Monotonicity: \( x \cup y \succeq x \) for all \( x, y \in X \).

(e) Aversion to Randomization: If \( x \sim y \), then \( x \succeq \frac{1}{2}x + \frac{1}{2}y \) for all \( x, y \in X \).

Items (a)–(d) are standard. Item (e) is familiar from Ergin and Sarver (2010a) and De Oliveira et al. (2016) and relaxes Independence in order to accommodate unobserved information choice: Suppose \( d_m \) is indifferent between the menus \( x \) and \( y \) on the basis of two different information plans. Choosing from \( \frac{1}{2}x + \frac{1}{2}y \) amounts to choosing an act from \( x \) and an act from \( y \) (before knowing which of the two will determine payoffs). In the presence of an information constraint, \( d_m \) may not be able to acquire (or process) both types of information at the same time, and thus would prefer to learn whether \( x \) or \( y \) is relevant before making his information choice.

The next axiom captures the special role played by consumption streams, which leave no consumption choice to be made in the future and therefore require no information (that is, all information alternatives perform equally well). The axiom thus requires \( \succeq|_L \) to satisfy additional standard assumptions. In what follows, for any \( c \in C \) and \( \ell \in L \), let \((c, \ell)\) be the constant act that yields consumption \( c \) and continuation stream \( \ell \) with probability one in every state \( s \in S \). By Continuity (Axiom 1(b)) and the compactness of \( L \), there exist best and worst consumption streams. As in Section 2.5, we denote these by \( \ell^\ast \) and \( \ell_* \).

---

(20) We remark that all our axioms except our notion of continuity can be falsified with finite data.

(21) For a discussion of (c) see Dekel et al. (2007) and for (d) see Kreps (1979).
respectively. For each \( I \subset S \), \( f \in \mathcal{F}(\Delta(C \times X)) \), \( (c, y) \in C \times X \), and \( \varepsilon \in [0, 1] \), define \( f \oplus_{\varepsilon, I} y \in \mathcal{F}(\Delta(C \times X)) \) by

\[
(f \oplus_{\varepsilon, I} y)(s) := \begin{cases} 
(1 - \varepsilon)f(s) + \varepsilon(c_s^-, y) & \text{if } s \in I \\
(f(s) & \text{otherwise}
\end{cases}
\]

That is, for any state \( s \in I \), the act \( f \oplus_{\varepsilon, I} y \) perturbs the continuation lottery with \( y \). Note that \( s \sim WD_{\ell_s} \sim WD_{\ell_s^*} \), so that we can define the induced binary relation \( \sim_s \) on \( L \) by \( \ell \sim_s \hat{\ell} \) if, and only if, \( \ell_s \sim WD_{\ell_s} \).

**Axiom 2** (Consumption Stream Properties).
(a) \( L \)-Independence: For all \( x, y \in X \), \( t \in (0, 1] \), and \( \ell \in L \), \( x \succ y \) implies \( tx + (1 - t)\ell > ty + (1 - t)\ell \).
(b) \( L \)-History Independence: For all \( \ell, \hat{\ell} \in L \), \( c \in C \), and \( s, s', s'' \in S \), \( (c, \ell_s) \succ_{s'} (c, \hat{\ell}_s) \) implies \( (c, \ell_s) \succ_{s''} (c, \hat{\ell}_s) \).
(c) \( L \)-Stationarity: For all \( \ell, \hat{\ell} \in L \) and \( c \in C \), \( \ell \succ WD_{\ell_s} \hat{\ell} \) if, and only if, \( (c, \ell) \succ WD_{\ell_s} (c, \hat{\ell}) \).
(d) \( L \)-Indifference to Timing: \( \frac{1}{2}(c, \ell) + \frac{1}{2}(c, \ell') \sim WD (c, \frac{1}{2}\ell + \frac{1}{2}\ell') \).

Axiom 2 (a) is closely related to the C-Independence axiom in Gilboa and Schmeidler (1989), and is motivated in a similar fashion: Because consumption streams require no information choice, mixing two menus with the same consumption stream should not alter the ranking between these menus. For a discussion of properties (b) through (d) see Krishna and Sadowski (2014).

### 3.2. Separability

Whether or not \( \text{dm} \) is likely to face a non-trivial decision in the future determines how much information he would like to gather about the state at that time, which, in turn, determines the expected opportunity cost of acquiring information prior to the realization of the current state. However, this expected opportunity cost depends only on the distribution over future decision problems that \( \text{dm} \) faces. That is, \( \text{dm} \)'s optimal learning will not change when substituting act \( f \) with \( g \) as long as they induce, on each state \( s \), the same marginal distributions over \( C \) and \( X \).

The following notation will be useful. For any \( f \in \mathcal{F}(\Delta(C \times X)) \), we denote by \( f_1(s) \) and \( f_2(s) \) the marginals of \( f(s) \) on \( C \) and \( X \), respectively. For any finite \( x \in X \), define the menu \( x \triangleright_f g \) obtained from \( x \) and \( f, g \in \mathcal{F}(\Delta(C \times X)) \) as

\[
x \triangleright_f g := \begin{cases} 
(x \setminus \{f\}) \cup \{g\} & f \in x \\
x & \text{otherwise}
\end{cases}
\]

(22) Since we are interested in the comparison of continuation problems, we hold the perturbation of the consumption outcome fixed across different perturbations. Fixing the perturbation of consumption to be \( c_s^- \), which is the worst possible consumption outcome in each state, will be of convenience later.
**Axiom 3 (State-Contingent Indifference to Correlation).** For a finite menu \( x \), if \( f, g \in \mathcal{F}(\mathcal{D}(C \times X)) \) are such that \( g_1(s) = f_1(s) \) and \( g_2(s) = f_2(s) \) for all \( s \in S \), then \( [x \triangleright f g] \sim x \).  

### 3.3. Strategic Rationality

Suppose that, contingent on a sequence of actions and realizations, \( \text{DM} \) is offered a chance to replace a certain continuation problem with another. \( \text{DM} \)'s attitude towards such replacements may depend on his initial information choice, which is subjective, unobserved, and menu-dependent. That said, any strategy of choice from an RACP gives rise to a consumption stream. Therefore, any continuation problem \( y \) should leave \( \text{DM} \) no worse off than receiving the worst consumption stream, \( \ell^* \). In particular, since the best consumption stream, \( \ell^* \), leaves \( \text{DM} \) strictly better off than \( \ell^*_s \) in every state, so should \( (1 - t)y + t \ell^* \) for all \( t > 0 \). Formally, with the abbreviated notation \( x \triangleright (f,\varepsilon,s) y := x \triangleright f \left( f \oplus_{(\varepsilon,s)} y \right) \), we require \([x \triangleright (f,\varepsilon,s) ((1 - t)y + t \ell^*)] \succ [x \triangleright (f,\varepsilon,s) \ell^*_s]\) for any \( f \in x, s \in S, y \in X \) and \( \varepsilon, t \in (0, 1) \). This restrictions is item (a) of Axiom 4 below. Item (b) investigates the conditions under which \( \text{DM} \) is actually indifferent to replacing continuation lotteries with the worst consumption stream.

Recall that the Ric requires \( \text{DM} \) to choose a partition of \( S \) in every period. Because partitions generate deterministic signals (each state is identified with only one cell of the partition), \( \text{DM} \)'s choice of partition determines which act he will choose from a given menu, contingent on the state. \( \text{DM} \) should then be willing to commit to this choice. In other words, there should be a contingent plan that specifies which act \( \text{DM} \) will choose for each state, contingent on his choice does not coincide with that plan. To formalize this state contingent notion of strategic rationality, we define the set of contingent plans \( \mathcal{E}_x \) to be the collection of functions \( \xi_x : S \to x \). An *Incentivized Contingent Commitment* to \( \xi_x \in \mathcal{E}_x \), is then the set

\[
\mathcal{I}(\xi_x) = \{ f \oplus_{(1,IC)} \ell^*_s : f \in x \text{ and } I = \{ s : f = \xi_x(s) \} \}
\]

which replaces the outcome of \( f \) with the worst outcome \((c^{-}_{\xi_x}, \ell^*_s)\) in any state where \( f \) should not be chosen according to \( \xi_x \). Obviously \( x \succeq \mathcal{I}(\xi_x) \) for all \( \xi_x \in \mathcal{E}_x \). However, if for no \( s \in S \) it is ever optimal to choose an act outside \( \xi_x(s) \), then \( x \sim \mathcal{I}(\xi_x) \) should hold.

**Axiom 4 (Indifference to Incentivized Contingent Commitment).** If \( x \in X \) is finite, then the following must hold.

---

(23) Axiom 3 is closely related to Axiom 5 in Krishna and Sadowski (2014), where other related notions of separability are also mentioned. The important difference is that Axiom 3 requires indifference to correlation in *any* RACP \( x \), rather than just singletons, because different information may be optimal for different RACPs.
(a) \( x \succ (f, s) ((1 - t) \ell^* + ty) \succeq x \succ (f, s) \ell^* \) for all \( \varepsilon, t \in (0, 1), f \in x, s \in S, y \in X \).

(b) There is \( \xi \in \mathcal{E}_x \) such that \( x \sim \mathcal{I} (\xi) \).\(^{24}\)

3.4. Independence

We envision information constraints where the choice of partition and the actual realization of the payoff relevant state in the initial period fully determine the available information choices in the subsequent period. We say that \( x \) and \( y \) are concordant if the same initial information choice is optimal for both \( x \) and \( y \). Note that if \( x \) and \( y \) are concordant, then both should be concordant with the convex combination \( \frac{1}{2}x + \frac{1}{2}y \). While Independence may be violated when considering \( \text{racp} \)s that lead to different optimal initial information choices, \( \succeq |_{X'} \) should satisfy Independence if \( X' \subset X \) consists only of concordant \( \text{racp} \)s.

We now introduce our behavioral notion of concordance (Definition 3.1 below).

We begin by making two observations. First, finiteness of \( S \) implies that if a partition is uniquely optimal for \( x \), then it will stay uniquely optimal for any \( \text{racp} \) in a small enough neighborhood of \( x \). Second, any one-period choice problem \( z \in \mathcal{K} (L) \) requires no choice after the initial period, so that its value depends only on the partition under which it is evaluated. In particular, for \( x_1 (P) := \{ \ell^* \oplus (1, t) \ell^*_t : I \in P \} \in \mathcal{K} (L) \), we have \( x_1 (P) \sim \ell^* \) if, and only if, \( x_1 (P) \) is evaluated under a partition that is finer than \( P \).

Given these two observations, consider two \( \text{racp} \)s \( x \) and \( y \) with \( x \sim y \), for which the unique optimal choices of partition are \( P_x \) and \( P_y \), respectively. There are two possibilities. Either (i) \( P_x = P_y \), in which case there is \( \varepsilon \in (0, 1) \) small enough, such that \( (1 - \varepsilon) x + \varepsilon z \sim (1 - \varepsilon) y + \varepsilon z \) for all \( z \in \mathcal{K} (L) \) and in particular for any \( x_1 (P), P \in \mathcal{P} ; \) or (ii) \( P_x \neq P_y \), which means that one of them, say \( P_y \), is not finer than the other and we have \( (1 - \varepsilon) x + \varepsilon x_1 (P_x) \sim (1 - \varepsilon) y + \varepsilon x_1 (P_x) \) for any \( \varepsilon \in (0, 1) \). We will say that \( x \) and \( y \) are concordant in case (i) but not in (ii).\(^{25}\) To extend this notion to \( x \) and \( y \) with \( x \sim y \), note that no choice of act is required for any \( \ell \in L \), and thus \( P_x \) must also be optimal for \( (1 - t)x + t\ell \). Therefore, if \( y \) is concordant with \( (1 - t)x + t\ell \), we will say that it is also concordant with \( x \).

**Definition 3.1.** For \( \varepsilon \in (0, 1) \), \( \text{racp} \)s \( x \) and \( y \) are \( \varepsilon \)-concordant if \( x \sim y \) and \( (1 - \varepsilon) x + \varepsilon x_1 (P) \sim (1 - \varepsilon) y + \varepsilon x_1 (P) \) for all \( P \in \mathcal{P} \). Two \( \text{racp} \)s \( x \) and \( y \) are concordant if \( (1 - t) x + t\ell \) and \( y \) are \( \varepsilon \)-concordant for some \( t \in [0, 1) \), \( \ell \in L \), and \( \varepsilon \in (0, 1) \).

**Axiom 5** (Concordant Independence). If \( x \) and \( y \) are \( \varepsilon \)-concordant, so are \( x \) and \( \frac{1}{2}x + \frac{1}{2}y \). Furthermore, if \( X' \subset X \) consists of concordant \( \text{racp} \)s, then \( \succeq |_{X'} \) satisfies Independence.\(^{26}\)

---

\(^{24}\) This is conceptually related to the Indifference to State Contingent Commitment Axiom introduced in Dillenberger et al. (2014), which also relates partitional learning to a state contingent notion of strategic rationality.

\(^{25}\) If the optimal partition for \( x \) or \( y \) is not unique, then our notion of concordance suggests that for any partition that is optimal for \( x \) there is at least as fine a partition that is optimal for \( y \), and vice versa.

\(^{26}\) If \( x, y, z, (1 - t) x + tz, (1 - t) y + tz \in X', t \in (0, 1) \), and \( x > y \), then \( (1 - t) x + tz > (1 - t) y + tz \).
3.5. Stationarity

We shall call a preference over dynamic choice problems *self-generating* if it has the same properties as each of the preferences over continuation problems that together generate it. In other words, ex ante and continuation preferences should satisfy the same set of axioms.\(^{27}\) *Self-Generation* is satisfied in any recursive model precisely because it embodies the dynamic programming principle.

Because continuation preferences are determined by the initial choice of partition \(P\) and the realized state \(s\), we will denote them by \(\succsim_{P,s}\). Self-Generation – which we state as an axiom after defining \(\succsim_{P,s}\) as a binary relation that is induced by the ex ante preference, \(\succsim\) – then requires the following:

\[ \succsim_{P,s} \text{ satisfies Axioms 1–5 and Self-Generation.} \]

It is important to note the self-referential character of Self-Generation, which is the only axiom that relies on the recursive structure of \(X\) (apart from \(L\)-Stationarity (Axiom 2(c)), which relies on the recursive structure of \(L\)); it requires today’s preferences on \(X\) and induced preferences over tomorrow’s continuation problems (again on \(X\)) to satisfy the same axioms. This includes the Self-Generation Axiom itself, thereby connecting preferences over tomorrow’s continuation problems to preferences over continuation problems two periods ahead, and so forth. This type of self-referential structure is built into the standard Stationarity axiom as well, where tomorrow’s preferences are required to coincide with today’s, and therefore those for the day after tomorrow also coincide with tomorrow’s, and so forth. One could, alternatively, write the axiom in extensive form, in which case it would simply require induced preferences in every period to satisfy Axioms 1–5.\(^{28}\)

Clearly, \(\succsim_{P,s}\) must be defined based on the initial ranking of \(\text{RACP}\)s, all of which give rise to the same optimal choice of partition, \(P\). To gain intuition for the construction below, suppose \(P\) is the unique optimal choice for the \(\text{RACP} \, x\). Because there are only finitely many partitions of \(S\), we can perturb any act \(f \in x\) by mixing it with different continuation problems, making sure to maintain the optimality of \(P\) by verifying concordance for each perturbation. Suppose further that perturbing \(f\) with the continuation problem \(y\) in state \(s\) leads to an improvement. In that case the perturbed \(f\) must be chosen optimally in state \(s\), which implies that the ranking of this perturbation and an analogous one with \(y'\) reveals the ranking of \(y\) and \(y'\) according to \(\succsim_{(P,s)}\). That is, if the perturbation by \(y'\) also leads to an improvement.

---

\(^{27}\) The bite of Self-Generation in a particular model (such as ours) will therefore depend on the axioms on ex ante choice that it perpetuates.

\(^{28}\) Another alternative formulation is that \(\succsim\) must belong to the recursively defined set \(\Psi^*\), where \(\Psi^* := \{\succsim\text{ on } X : (i)\text{ \(\succsim\) satisfies Axioms 1–5, and (ii) } \succsim_{(P,s)} \in \Psi^*\}\). Notice that the set \(\Psi^*\) is the fixed point of an operator just as the self-generating set of equilibrium payoffs in Abreu, Pearce, and Stacchetti (1990) is the fixed point of an appropriate operator. Our representation theorem, Theorem 2, characterizes the largest such set \(\Psi^*\) via a well defined recursive value function, and establishes that it is non-empty.
improvement, then we can compare the two perturbed menus, but even if the perturbation by \( y' \) does not lead to an improvement, we can nevertheless infer that \( y \succsim_{(P,s)} y' \).\(^{29}\)

Although unobservable, \( \succsim_{(P,s)} \) can be deduced from choice behavior. We now define an induced binary relation \( \succsim_{(x,s)} \) on \( X \), verify in Appendix C.4 that it is well defined, and show that if it is complete on \( X' \subset X \), then \( \succsim_{(x,s)} = \succsim_{(P,s)} \) on \( X' \), where \( P \) is an optimal information choice given \( x \). Conversely, for every \( P \) and \( s \) there are a sequence of finite \( x_n \in X \), and a sequence \( X_n \) with \( X_n \subset X \) that converges to a dense set in \( X \), such that \( \succsim_{(x_n,s)} = \succsim_{(P,s)} \) on \( X_n \).

**Definition 3.2.** For \( y, y' \in X, s \in S \) and finite \( x \), suppose there are \( \epsilon > 0 \) and \( f \in x \) such that \( x \succ (f, \epsilon, s) y, x \succ (f, \epsilon, s) y' \), and \( x \) are pairwise concordant, and \( [x \succ (f, \epsilon, s) y] > x \). For any such \( y, y' \in X \), let \( y \succsim_{(x,s)} y' \) if \([x \succ (f, \epsilon, s) y] > [x \succ (f, \epsilon, s) y']\).

**Axiom 6 (Self-Generation).** If \( \succsim_{(x,s)} \) is complete on \( X' \subset X \), then it satisfies Axioms 1–6 on \( X' \).

Axiom 6 is weaker than Stationarity (eg, as in Gul and Pesendorfer (2004)), in the sense that it only requires immediate and continuation preferences to be of the same type rather than identical, but it is stronger in the sense that it restricts contingent ex post preferences, rather than aggregated future preferences.

### 3.6. Representation Theorem

**Theorem 2.** Let \( \succsim \) be a binary relation on \( X \). Then, the following are equivalent:
(a) \( \succsim \) satisfies Axioms 1–6.
(b) There exists an \( \mu \)-representation of \( \succsim \).

The proof of Theorem 2 is quite involved. In Appendix C we establish the following representation of \( \succsim \):

\[
V(x) = \max_{P \in \mathcal{P}} \sum_{I \in \mathcal{I}} \left( \max_{f \in x} \sum_{s \in I} E^{f(s)} \left[ u_s(c) + v_s(y, P) \right] \pi(s \mid I) \right) \pi(I)
\]

where \( \mathcal{I} \subset \mathcal{P} \) is a set of partitions of \( S \), the measure \( \pi(s \mid I) \) is the probability of \( s \) conditional on the event \( I \subset S \), and utilities \( (u_s, \pi) \) over continuation problems depend only on the partition \( P \). We say that \( V \) is implemented by \( ((u_s, \pi, \mathcal{I}, (v_s(c, P)), \pi)) \). This representation already has all the features we need to establish, except that it is static; it does not exploit the recursive structure of \( X \). Correspondingly, we do not rely on Axiom 6, but only on Axioms 1–5 to derive it.

\(^{29}\) In contrast, if the perturbation by \( y \) reduces the value of \( x \), this might be because \( \text{DM} \) now chooses a different, previously inferior act in state \( s \). In this case, comparing the value of perturbed menus might not reveal \( \succsim_{(P,s)} \).
Next, adapting the terminology of Abreu, Pearce, and Stacchetti (1990), we define the set $\Phi^*$ of self-generating value functions, where each $v \in \Phi^*$ is implemented by some tuple $((u_s), \emptyset, (v_s(. , P)), \pi)$ in a way that each $v_s(., P)$ is itself in $\Phi^*$ (see Appendix A.8). We rely on Self-Generation (Axiom 6) to show that the representation $V$ of $\succsim$ can be made self-generating. Clearly, a self-generating value function may not be recursive.

The remainder of our construction has two main components. First, we construct an ríc $\omega_0$ from a self-generating representation, and argue that any other self-generating representation of $\succsim$ would yield the same $\omega_0$ up to recursive Blackwell dominance. The intuition for this construction is precisely the one in our proof of Theorem 1, where we elicit $\omega_0$ from $\succsim$ without having to elicit beliefs.

Second, we establish a recursive representation of $\succsim|_L$, which is the raa representation in Krishna and Sadowski (2014), parametrized by $((u_s), \delta, \Pi)$, and discussed in Appendix A.7. Starting from agreement of the self-generating and raa representations on $L$, we then show that we can pair the parameters $((u_s), \delta, \Pi)$ with $\omega_0$ to find the mic-representation, $((u_s), \delta, \Pi, \omega_0)$, which is recursive on all of $X$, and where $\omega_0$ is a canonical mic. Intuitively, the lack of recursivity in the self-generating representation, which conditions only on the objective state $s$, is absorbed by the evolution of the subjective state $\omega$ in our representation, so that the representation becomes recursive when conditioning on both, $s$ and $\omega$.

### 3.7. Invariant Per-Period Constraint and Fixed Arrival of Information

We now discuss two special cases of the mic-representation. In the first, dm faces the same information constraint each period. This case is of interest due to its simplicity and its frequent use in dynamic models of rational inattention, where there is a periodic time invariant upper bound on information gain, measured by the expected reduction in entropy. Recall that $x \in X$ is $\succsim$-maximal if $x \succsim y$ for all $y \in X$.

**Axiom 7 (Stationary Maximal rACP).** $x \in X$ is $\succsim$-maximal if, and only if, it is $\succsim_{(y,s)}$-maximal for all $y \in X$ and $s \in S$.

The axiom requires maximal rACPs to be stable in three ways: Stationarity, because between $\succsim$ and $\succsim_{(y,s)}$ a period has passed; Separability, through the comparison of $\succsim_{(y,s)}$ and $\succsim_{(y',s)}$; and State Independence, through the comparison of $\succsim_{(y,s)}$ and $\succsim_{(y,s')}$.

**Definition 3.3.** The mic $\mathcal{M} = (\Theta, \theta, \Gamma, \tau)$ is an invariant per-period constraint if $\Gamma(\theta)$ is constant on $\Theta$ (or, equivalently, if $\Theta$ is a singleton).

In contrast to a general mic, an invariant per-period constraint is independent of past information choice, and so does not accommodate any intertemporal trade-offs in processing information.

---

(30) See also Footnote 28.
Proposition 3.4. If \( \succcurlyeq \) has mic-representation \( ((u_\delta), \delta, \Pi, \mathcal{M}) \), then it satisfies Axiom 7 if, and only if, \( \mathcal{M} \) is an invariant per-period constraint.

To see why this must be true, note that \( \ell^* \) is both, \( \succcurlyeq \)-best and \( \succcurlyeq_{(y,s)} \)-best for all \( y \in X \) and \( s \in S \). It follows from the argument in Section 2.5 that the mic \( \mathcal{M} = (\Theta, \theta_0, \Gamma, \tau) \) is indistinguishable from the mic \( (\Theta, \tau(\theta_0, P, s), \Gamma, \tau) \) for all \( P \in \Gamma(\theta_0) \) and \( s \in S \). The other direction is immediate.

In the second special case we consider, \( \mathcal{D}m \) faces a trivial choice between information plans, that is, he can not influence the arrival of information about the state of the world.\(^{31}\)

Axiom 8 (Independence). If \( x > y \), then \( tx + (1-t)z > ty + (1-t)z \) for all \( x, y, z \in X \) and \( t \in (0, 1) \).

Definition 3.5. The mic \( \mathcal{M} = (\Theta, \theta, \Gamma, \tau) \) captures fixed arrival of information if \( \Gamma(\theta) \) is a singleton for all \( \theta \in \Theta \).

Proposition 3.6. If \( \succcurlyeq \) has mic-representation \( ((u_\delta), \delta, \Pi, \mathcal{M}) \), then it satisfies Axiom 8 if, and only if, \( \mathcal{M} \) captures fixed arrival of information.

To see why this must be true, suppose instead that \( P, P' \in \Gamma(\theta) \) where \( P \) and \( P' \) are not ranked by fineness for some \( \theta \). Then \( x_1(P) \sim x_1(P') \sim \ell^* > \frac{1}{2}x_1(P) + \frac{1}{2}x_1(P') \), contradicting Independence. This argument easily extends to mics that contain any two information plans that are not ranked by recursive Blackwell dominance.

Remark 3.7. At the end of Section 2.5 we discussed aspects of our identification strategy that might generalize to other situations where \( \mathcal{D}m \) faces an unobserved decision process. Similarly, some of our axioms should remain relevant in such a situation. We have already noted that a version of Self-Generation (Axiom 6) must hold for any recursive value function. In addition, our motivations for Axiom 3 (a notion of Separability) and Axiom 5 (which relaxes Independence) did not rely on the specifics of the mic, but only on the presence of some unobserved decision process that interacts with observable choice. The two special cases above suggest that Independence will be violated whenever \( \mathcal{D}m \) faces non-trivial unobserved choice, and Separability cannot hold if the subjective constraint is not time invariant.

4. Applications

We propose two applications that build on examples of mics outlined in the Introduction.

\(^{31}\) This parallels the representation in Krishna and Sadowski (2014), where \( \mathcal{D}m \) faces a fixed stream of information about the own taste, rather than the state of the world.
4.1. Expertise in learning and familiarity bias

Suppose \( \text{dm} \) has become familiar with a certain set of alternatives (consumption acts) and has gained expertise in learning the specific information needed to optimally choose among them. Such expertise may lead \( \text{dm} \) to be biased towards choosing between those alternatives, as he may find it too attention-intensive to optimally choose between less familiar ones. For example, investors who decide whether or not to enter new markets, or professionals who debate a career change, may find it more difficult to make decisions in the face of new and unfamiliar alternatives, relative to making more routine choices. This can lead to a ‘locked-in’ phenomenon that we term familiarity bias, according to which individuals are reluctant to switch away from familiar choice problems, even in favor of options that are deemed superior in the absence of familiarity.

Let \( \mathcal{K} \subset X \) collect all separable consumption problems, where \( (F_t) \in \mathcal{K} \) denotes the racp for which \( \text{dm} \) expects to choose from \( F_t \in \mathcal{K} (\Delta (C)) \) in period \( t \). For \( F, G \in \mathcal{K} (\Delta (C)) \), denote by \( F T G \) the problem \( (F_t) \in \mathcal{K} \) with

\[
F_t := \begin{cases} 
F & t \leq T \\
G & t > T 
\end{cases}
\]

and let \( F_\infty \) be the problem \( (F_t) \in \mathcal{K} \) with \( F_t := F \) for all \( t \).

**Definition 4.1.** \( \sim \) is familiarity biased if

(a) \( F T G \sim F_\infty \) implies \( G_\infty \sim G T F \) for all \( F, G \in \mathcal{K} (\Delta (C)) \) and \( T > 0 \).

(b) \( F_\infty \succ F T G \) and \( G_\infty \succ G T F \) for some \( F, G \in \mathcal{K} (\Delta (C)) \) and \( T > 0 \).

That is, it cannot be that replacing \( F \) with \( G \) after first choosing from \( F \), and replacing \( G \) with \( F \) after first choosing from \( G \) are both beneficial, and for some \( F \) and \( G \) both are detrimental.

To simplify the exposition, in the analysis below we allow \( \text{dm} \) to familiarize himself with any racp \( x \) before having to choose from it. For \( \ell \in L \) and \( x \in X \), denote by \( \ell \circ_{T*} x \in X \) the racp that pays according to \( \ell \) in every period until \( T* - 1 \) and then in period \( T* \) in every state delivers \( x \). For an arbitrary but fixed \( \ell \in L \), define \( \sim_{T*} \) on \( X \) by \( x \sim_{T*} y \) if, and only if, \( \ell \circ_{T*} x \sim \ell \circ_{T*} y \). That is, \( \sim_{T*} \) compares \( x \) and \( y \) after \( T* - 1 \) periods that require no choice.

Consider \( \sim \) with the mic-representation \( (u_s, \delta, \Pi, M) \). We now verify that for \( T* \) large enough, \( \sim_{T*} \) will be familiarity biased when the mic \( M = (\Theta, \theta_0, \Gamma, \tau) \) relies on expertise in the sense of the following two criteria. First, after learning \( P \) it is always possible

---

(32) Our notion of familiarity bias resembles that of status-quo bias, as in Samuelson and Zeckhauser (1988), according to which individuals often prefer to stick with their current or previous decision over switching to a new alternative. In our context, the menu \( \text{dm} \) is familiar with serves as the baseline or reference point, and switching away from it is costly. In other words, the bias is not towards the alternative ultimately consumed, but towards the choice problem from which to select this alternative. Home bias in investment choices, as in Massa and Simonov (2006), could be thought of as an instance of familiarity bias.
to learn $P$ again. Second, there are $P \neq Q$ that are maximal among those accessible under $\mathcal{M}$, and for which it is impossible to learn $P$ after $Q$ or $Q$ after $P$.

**Definition 4.2.** $\mathcal{M}$ relies on expertise if the following hold.

(a) $P \in \Gamma (\tau (P, \theta, s))$ for all $\theta \in \Theta$ and $s \in S$.

(b) There are $P, Q$ that are maximal among those that are accessible under $\mathcal{M}$ for which $Q \notin \Gamma (\tau (P, \theta, s))$ and $P \notin \Gamma (\tau (Q, \theta, s))$ for any $\theta \in \Theta$ and $s \in S$.

**Proposition 4.3.** If $\mathcal{M}$ relies on expertise, then for $T^\ast$ large enough $\preceq T^\ast$ is familiarity biased.

Intuitively, for $T^\ast$ large enough $\mathcal{M}$ can use the first $T^\ast$ periods to gain the expertise to learn any partition accessible under $\mathcal{M}$, thereby achieving the highest possible per period payoff from any $F$ once facing $F_\infty$. Therefore, it is not possible to have both $F_T G \succ T^\ast F_\infty$ and $G_T F \succ T^\ast G_\infty$. In particular, there are $F$ and $G$ for which the uniquely optimal partitions among those accessible under $\mathcal{M}$ are $P$ and $Q$ from part (b) of Definition 4.1, respectively, so that achieving the highest possible per-period payoff is not possible when choosing from $F_T G$ for any $T > 0$, and analogously for $G_T F$. If $F$ and $G$ are such that they generate sufficiently similar value under the respective optimal partitions, then switching from $F$ to $G$ (or from $G$ to $F$) after $T$ large enough will be detrimental.

It is easy to verify that the mic from Example 1.3 in the Introduction relies on expertise, provided $\mathcal{M}$ can never learn the finest partition of $S$. In that case, it can be shown that $T^\ast = 0$ in Proposition 4.3, that is, $\preceq$ itself is familiarity biased.

### 4.2. Search for Wages and Optimal Stopping Rule

Consider the following search problem. An unemployed worker seeks to maximize $\mathbb{E} \sum_{t=0}^{\infty} \delta^t a_t$, where $a_t = w$ if the worker is employed at wage $w$, $a_t = 0$ if the worker is unemployed, and $\delta \in (0, 1)$. Each period, the unemployed worker draws a wage from an iid distribution, where we denote by $\pi_s$ the probability that the wage drawn in the next period is $w_s \in W := \{w_1, \ldots, w_n\}$. Once he accepts an offer, the worker works forever at the accepted wage. There is no firing or quitting.

This optimal stopping problem can naturally be embedded in the domain of racp as follows. The stopping problem is the menu $x(w) := \{\text{accept, continue}\}$, where ‘accept’ is the act that pays a consumption stream of $w$ forever (with corresponding lifetime value of $w/(1-\delta)$) and ‘continue’ is the act with continue$(w) = (0, x)$ for all $s$.

---

(33) We say that $P$ is accessible under $\mathcal{M}$ if $P \in \Gamma (\theta)$ for some $\theta \in \Theta$ that is accessible from $\theta_0$ in finite time. We say that $P$ is maximal among those that are accessible under $\mathcal{M}$ if no strictly finer partition is accessible.

(34) The mic from Example 1.2 also relies on expertise if $\beta = 0$ and the prior on $S$ used in the calculation of $c(P \mid Q)$ is constant, as would be the case for states that are distributed independently over time.

(35) See for example Ljunqvist and Sargent (2004, p 161).
Let \( v(w) \) be the value of being offered the wage \( w \in W \) when there is no attention constraint. The corresponding Bellman equation is

\[
v(w) = \max \left\{ \frac{w}{1-\delta}, \delta \sum_{s=1}^{n} \pi_s v(w_s) \right\}
\]

It is easy to see that the optimal policy has the following stationary form: there is a reservation wage \( w^* \) such that the worker accepts an offer \( w \) if, and only if, \( w \geq w^* \).

Now suppose instead that the worker faces the following attention constraint: at the beginning of each period he can either learn nothing or pay attention to learn the wage precisely. The worker is endowed with an initial attention stock \( K \), which he draws down by \( K/k \) with \( k \in \mathbb{N} \) every time he chooses to learn the wage. That is, the worker can learn the wage at the beginning of at most \( k \) periods. The worker’s choice will now depend on the remaining attention stock, or, equivalently, the remaining number of periods for which he can learn the wage.

Suppose first that the worker is left without any attention. He then can not learn the current wage offer, and instead faces expected wage \( \bar{w} := \sum_{s=1}^{n} \pi_s w_s \). He also anticipates facing \( \bar{w} \) in each future period, and hence will accept immediately due to discounting. Again due to discounting (and the fact that the expected wage does not change across periods) \( \delta \bar{w} \) will learn for the first \( k \) consecutive periods, and accept in period \( k + 1 \) if he did not accept prior to that; hence \( v_{k+1} = \frac{\bar{w}}{1-\delta} \).

In period \( k \) after observing \( w \), the worker compares \( \frac{w}{1-\delta} \) to \( \delta v_{k+1} \) or \( w \) to \( \delta \bar{w} =: c_k \). We thus have \( v_k(w) = \frac{1}{1-\delta} \max \{ w, c_k \} \). In period \( k - 1 \), the worker compares \( w \) to \( c_{k-1} \) where \( c_{k-1} := \delta \sum_{s=1}^{n} \pi_s \max \{ w_s, c_k \} \). Similarly, in period \( k - t \), he compares \( w \) to \( c_{k-t} \) where \( c_{k-t} := \delta \sum_{s=1}^{n} \pi_s \max \{ w_s, c_{k-t+1} \} \). Importantly, \( \sum_{s=1}^{n} \pi_s \max \{ w_s, c \} \) is strictly increasing in \( c \), and hence for \( \delta \bar{w} > 0 \) we have \( c_k > 0 \) and \( c_{k-t} > c_{k-t+1} \). That is, the cutoff \( c_t \) is decreasing in \( t \) until \( t = k + 1 \), at which time any wage realization is accepted.

It is a well documented pattern that reservation wages decrease over time.\(^{36}\) Our model is consistent with this pattern and suggests that passing search time may reduce the reservation wage because the expected value of continuing the search decreases as the attention constraint tightens over time.

5. Related Literature

Kreps (1979) studies choice between menus of prizes. He explains preference for flexibility via uncertain tastes that are yet to be realized. Dekel, Lipman, and Rustichini (2001) show that by considering menus of lotteries over prizes (where lotteries live in a linear space), those tastes can be regarded as vN-M utility functions over prizes. Dillenberger et al. (2014) subsequently show that preference for flexibility over menus of acts (where the space of acts

\(^{36}\) Brown, Flinn, and Schotter (2011) discuss this evidence and also document and investigate declining reservation wages in a laboratory experiment.
is a linear space) corresponds to uncertainty about future beliefs about the objective state of the world. Ergin and Sarver (2010a) and De Oliveira et al. (2016) weaken Independence in the respective models in order to accommodate subjective uncertainty that is not fixed, but a choice variable. The former studies costly contemplation about future tastes, while the latter studies rational inattention to information about the state. In that sense, weakening Independence allows us to interpret the subjective state space as a collection of (one-period) plans.

None of the models discussed so far are recursive or truly dynamic. Krishna and Sadowski (2014) provide a dynamic extension of Dekel, Lipman, and Rustichini (2001) where the flow of information is taken as given by $\mathbf{dm}$. In particular, Krishna and Sadowski (2014) assume Independence, and so their subjective state space is the space of vN-M utility functions in each period. Their recursive domain consists of acts that yield a menu of lotteries over consumption and a new act for the next period. When all menus are degenerate, their domain reduces to the set of consumption streams $L$, as it does here. The key difference between the two domains lies in the timing of events: Instead of acts over menus of lotteries, $\text{racp}$s are menus of acts over lotteries, which is appropriate for a dynamic extension of Dillenberger et al. (2014). Our model also extends De Oliveira et al. (2016), in the sense that the arrival of information is not given, but is determined by a constrained choice process, the $\text{ric}$ (or equivalently the $\text{mic}$).

As a consequence of $\mathbf{dm}$ controlling the $\text{mic}$, his preferences will be interdependent across time, which significantly complicates our analysis, especially because we can no longer appeal to the stationarity assumptions of Krishna and Sadowski (2014). To deal with this complication, we observe that preferences over consumption streams, $\succsim_L$, should satisfy the standard axioms, including Stationarity, because future information plays no role when there is no consumption choice in the future. We then use the ranking of consumption streams to ‘calibrate’ preferences over all $\text{racp}$s, similar to the approach in Gilboa and Schmeidler (1989), where preferences over unambiguous acts (lotteries) are used to calibrate ambiguity averse preferences over all acts.

A recent paper by Piermont, Takeoka, and Teper (2015) studies a decision maker who learns about his uncertain, but time invariant, consumption taste (only) through consumption. For static choice situations, the literature based on ex post choice partly parallels the menu-choice approach. Ellis (2016) identifies a partitional information constraint from ex post choice data. Caplin and Dean (2014) use random choice data to characterize a representation of costly information acquisition with more general information structures. They then proceed to consider stochastic choice data under the assumption that attention entails entropy costs, as do Matějka and McKay (2014). To our knowledge there is no counterpart to our recursive analysis in the random-choice literature.

(37) The insight that weakening Independence is essential in order to allow unobserved actions (when the domain of choice is a linear space) originates in Gilboa and Schmeidler (1989). They consider an Anscombe-Aumann style setting, and allow for subjective beliefs that vary with the choice of act.
6. Discussion

Descriptive Interpretation of Domain and Representation

Koopmans (1964) argues that it is implausible to assume that, when planning ahead, people determine the optimal course of action for all future contingencies. Instead, the decision making process is better described as piecemeal planning, whereby people choose at each instance consumption for the current period together with a continuation problem for the next period, without already planning their optimal choice from the latter. To do so, they directly assign value to (continuation) choice problems. This is exactly the intuition behind Bellman’s dynamic programming principle, where the value \( dM \) assigns to choice problems is consistent with the value optimal (static) choice at each instance will generate over the infinite horizon.

In our model, \( dM \) faces two decisions in each period, one for information and one for consumption — with implications for the continuation \( RIC \) and \( RACP \) respectively — but the same interpretation applies to the recursive structure of our representation; \( dM \) determines a value for each pair of \( RIC \) and \( RACP \) instead of forming a plan for the choice of information and consumption for all the future. That this interpretation does not rely on cognitively very involved forward looking behavior is perhaps of particular relevance when \( dM \) faces a cognitive, rather than physical information constraint. This discussion also suggests that choice between menus is quite natural in a dynamic environment, where \( dM \) chooses a (continuation) \( RACP \) in each period. The ex ante choice between \( RACPs \) can then be viewed as a ‘snapshot’ of the ongoing process of piecemeal planning.

Costly Information Acquisition

A popular alternative model to ours is one in which limitations on information acquisition are modeled via direct information costs, measured in consumption ‘utils’ (see, for example, Woodford (2012) and Caplin and Dean (2014)). We confine our attention to information constraints rather than costs for a number of reasons.

First, in the \( RIC \)-representation (or equivalently the \( MIC \)-representation), the \( RIC \) is not measured in utils, and hence its elicitation from behavior is done independently of the elicitation of the collection of state-dependent utility functions (see Section 2.5). Second, because a constraint corresponds to a cost function that can take only two values, 0 or \( \infty \), our model has significantly fewer degrees of freedom than the alternative formulation. Third, \( RICs \) can generate opportunity costs of information acquisition via tighter future constraints. Our model allows us to focus entirely on the behavioral implications of this new type of dynamic costs. 38

38 In static settings, information constraints imply that the amount of information chosen is independent of the scaling of the payoffs involved, which stands in sharp contrast to the stake-dependency under costly information acquisition. Because \( MICs \) can generate opportunity costs of information acquisition, choice
Finally, information costs or constraints are identified only up to dominated information strategies, that is, strategies that are never optimally chosen. A model with costs raises the problem that an information strategy is undominated only if its instrumental value in terms of utilities justifies its cost. This notion of dominance amounts to a joint restriction on all the preference parameters, and will generically not have an intuitive formulation. In contrast, the \( \text{ric} \) in our model can be identified from preferences up to recursive Blackwell dominance, which ranks \( \text{rics} \) independently of preferences.\(^{39}\)

One could confine attention to representations that are minimal, in the sense that they only include undominated information strategies (see, for example, Ergin and Sarver (2010a)). Theorem 1 immediately implies that the \( \text{ric} \) can be fully identified from behavior in a minimal \( \text{ric} \)-representation.\(^{40}\) Furthermore, for a minimal representation with information costs, our proof strategy would need only minor adaptations to establish unique identification.\(^{41}\) However, most intuitive examples of information costs or constraints (including all those in Section 1) will \emph{not} give rise to a minimal representation, and because the space of possible information strategies is infinite, it is impossible to verify the minimality of a representation in finitely many steps.

Learning the Payoff-relevant State after each Period

As is apparent from Equation [Val], last period’s state of nature \( s \in S \) is a state-variable in our recursive model, that is, \( \text{dm} \) always learns the realized state of nature at the end of a period. Only the acquisition of information at the beginning of each period is constrained.

Naturally, once \( \text{dm} \) becomes aware of the continuation problem he faces, he should take into account the information about the state of nature encoded in its realization or, to be consistent, the realization of the pair \( (c, y) \) of consumption and continuation problem.\(^{42}\) But if \( \text{dm} \) learns only the information encoded in the act, then he should be willing to pay a premium (e.g., in terms of current consumption) to avoid acts with state-independent payoffs,

\(^{39}\) In a static setting, a useful way to avoid the additional lack of identification that arises in the context of information costs is to consider unbounded consumption utilities (see De Oliveira et al. (2016)). In our dynamic setting, where the space of information strategies is infinite dimensional, unbounded utilities introduce new complications; for example, even ensuring existence of a recursive value function would require additional structural assumptions.

\(^{40}\) Supplementary Appendix G provides a recursive notion of minimality for sets of \( \text{ric} \)'s that is based on recursive Blackwell dominance, and is therefore independent of preferences.

\(^{41}\) In particular, rather than mixing over continuation problems with uniform weights in the identification proof, more general mixing must be considered. We omit formal arguments for brevity.

\(^{42}\) To justify this assumption, suppose \( \text{dm} \) plans to choose an act \( f \in x \) which yields a continuation problem \( y \) if, and only if, the realized state is \( s \). When evaluating \( x \), \( \text{dm} \) calculates the continuation value of \( y \) using \( \pi_s \). Now suppose \( \text{dm} \) becomes aware that the realized continuation problem for next period is actually \( y \), but he does not take into account the information contained in this realization. In that case, he will choose from \( y \) based on a probability distribution which is \emph{less} accurate than \( \pi_s \). That \( \text{dm} \) becomes less informed after learning seems unreasonable.
leading to implausible violations of continuity. Moreover, if it is possible to place side bets with arbitrarily small stakes, DM would always do so in order to fully reveal the state, so that it is with minor loss to assume that the state becomes known for free. We, therefore, simply assume DM always learns the state at the end of the period.43

Elicitation

In Section 2.5 we discuss how to construct for any \((P, \omega')\) an RACP \(x(P, \omega')\) that is strongly aligned with it. This RACP can be used to learn about an agent’s information constraint (in practice). To illustrate, suppose we know that DM’s preference has ric-representation \(((u_s)_{s \in S}, \Pi, \delta, \omega_0)\) (or an equivalent mic-representation), but do not know the value of \(\omega_0\). Suppose also that we are only interested in finding out whether for a particular RACP \(y\) a particular dynamic information plan (or one that weakly dominates it) is feasible for DM.44 To do so, we can consider \(\omega\) that contains only said plan. Suppose the plan requires the first period choice of \((P, \omega')\), with subsequent choices in \(\omega'\) and so on. Now consider the menu \(x(P, \omega')\). If DM prefers \(x(P, \omega')\) to \(\ell^*\), then it must be that the plan, or something that weakly dominates it, is feasible under \(\omega_0\). In other words, for a particular RACP \(y\) it is possible to test whether \(\omega_0\) allows DM to follow a particular information plan with just one binary choice question.45

mics in Strategic Situations

This paper provides the first dynamic model of choice under recursive information constraints, which can naturally be used in dynamic applications of rational inattention as well as in other studies of information acquisition over time. While we focus on understanding mics in the context of single-person decision making, our model could also be applied in strategic situations. To suggest just one instance, consider a monopolistic competition setting where each firm faces an mic as in Example 1.4; in each period, it can either learn the true state of the economy or stay uninformed, but cannot learn the state in two consecutive periods. Each firm thus needs to decide when to learn the state and how to price their product conditional on being informed or uninformed. This setting raises the question of whether or

\[\text{(43) An alternative model could assume that DM learns neither state nor realized continuation problem at the end of a period. This would require modelling choice under unawareness of the available alternatives. Our assumption avoids the complications this would entail in order to focus our model on the novel feature of recursive information constraints. This tension is less severe in environments where the set of available actions remains unchanged, and at most their payoff consequences vary, as, for example, in Steiner, Stewart, and Matějka (2015).}
\]

\[\text{(44) For example, a policy maker might be interested to know whether an agent is able to follow the least demanding information plan that would allow him to optimally make a particular sequence of decisions given \(y\).}
\]

\[\text{(45) Because there are only finitely many partitions, there are only finitely many possible \(T\)-period information plans. Hence, we can learn exactly which of those \(\omega_0\) admits, based on finitely many observations of the type just discussed, and proceed to (monotonically) approximate \(\omega_0\) by increasing \(T\). Elicitation of the other parameters from \(\gtrsim |L|\) is discussed in Krishna and Sadowski (2014).}
\]
not we see coordination in the processing of information. In particular, given the attention constraint specified above, will all firms decide to learn the state and adjust their prices in the same period — thereby inducing a larger price volatility in one period than in the other — or will we observe heterogeneous behavior with constant volatility? ④6

Appendices

A. Preliminaries

A.1. Metrics on Probability Measures

Let \((Y, dy)\) be a metric space and let \(\Delta(Y)\) denote the space of probability measures defined on the Borel sigma-algebra of \(Y\). The following definitions may be found in Chapter 11 of Dudley (2002). For a function \(\varphi \in \mathbb{R}^Y\), the supremum norm is \(\|\varphi\|_\infty := \sup_y |\varphi(y)|\), and the Lipschitz seminorm is defined by \(\|\varphi\|_L := \sup_{y \neq y'} |\varphi(y) - \varphi(y')| / d_Y(y, y')\). This allows us to define the bounded Lipschitz norm \(\|\varphi\|_{BL} := \|\varphi\|_L + \|\varphi\|_\infty\). Then, \(BL(Y) := \{\varphi \in \mathbb{R}^Y : \|\varphi\|_{BL} < \infty\}\) is the space of real-valued, bounded, and Lipschitz functions on \(Y\).

Define \(d_D\) on \(\Delta(Y)\) as \(d_D(\alpha, \beta) := \frac{1}{2} \sup \{\int \varphi \, d\alpha - \int \varphi \, d\beta : \|\varphi\|_{BL} \leq 1\}\). This is the Dudley metric \(\Delta(Y)\). Theorem 11.3.3 in Dudley (2002) says that for separable \(Y\), \(d_D\) induces the topology of weak convergence on \(\Delta(Y)\). We note that the factor \(\frac{1}{2}\) is not standard. We introduce it to ensure that for all \(\alpha, \beta \in \Delta(Y)\), \(d_D(\alpha, \beta) \leq 1\).

A.2. A Recursive Domain

Let \(X_1 := \mathcal{K}(\mathcal{F}(\Delta(C)))\). For acts \(f^1, g^1 \in \mathcal{F}(\Delta(C))\), define the metric \(d^{(1)}\) on \(\mathcal{F}(\Delta(C))\) by \(d^{(1)}(f^1, g^1) := \max_s d_D(f^1(s), g^1(s))\). For any \(f^1 \in \mathcal{F}(\Delta(C))\) and \(x_1 \in X_1\), the distance of \(f^1\) from \(x_1\) is \(d^{(1)}(f^1, x_1) := \min_{g^1 \in X_1} d^{(1)}(f^1, g^1)\) (where the minimum is achieved because \(x_1\) is compact). Notice that for all acts \(f^1\) and \(g^1\), \(d^{(1)}(f^1, g^1) \leq 1\).

This allows us to define the Hausdorff metric \(d_H^{(1)}\) on \(X_1\) as

\[
d_H^{(1)}(x_1, y_1) := \max \left[ \max_{f^1 \in X_1} d^{(1)}(f^1, y_1), \max_{g^1 \in Y_1} d^{(1)}(g^1, x_1) \right]
\]

and because the distance of an act from a set is bounded above by 1, it follows that for all \(x_1, y_1 \in X_1\), \(d_H^{(1)}(x_1, y_1) \leq 1\). Intuitively, \(X_1\) consists of all one-period Anscombe-Aumann (AA) choice problems.

Now define recursively, for \(n > 1\), \(X_n := \mathcal{K}(\mathcal{F}(\Delta(C \times X_{n-1}))\). The metric on \(C \times X_{n-1}\) is the product metric; that is, \(d_{C \times X_{n-1}}((c, x_{n-1}), (c', x'_{n-1})) = \max[d_C(c, c'), d^{(n-1)}(x_{n-1}, x'_{n-1})]\). This induces the Dudley metric on \(\Delta(C \times X_{n-1})\).

(④6) We describe the problem as a firm’s decision of when to pay attention. For example, in Maćkowiak and Wiederholdt (2009), firms also need to decide what to pay attention to. Some other works, such as Myatt and Wallace (2012), study a related problem of information acquisition in coordination games.
For acts \( f^n, g^n \in \mathcal{T}(\Delta(C \times X_{n-1})) \), define the distance between them as \( d^{(n)}(f^n, g^n) := \max_s d_D(f^n(s), g^n(s)) \). As before, we may now define the Hausdorff metric \( d^{(n)}_H \) on \( X_n \) as
\[
d^{(n)}_H(x_n, y_n) := \max\left[ \max_{f^n \in x_n} d^{(n)}(f^n, y_n), \max_{g^n \in y_n} d^{(n)}(g^n, x_n) \right]
\]
which is also bounded above by 1. Here, \( X_n \) consists of all \( n \)-period AA choice problems. The agent faces a menu of acts which pay off in lotteries over consumption and \((n-1)\)-period AA choice problems that begin the next period.

Finally, let \( X^* := \times_{n=1}^{\infty} X_n \) and endow it with the product topology. The Tychonoff metric induces this topology and is given as follows: For \( x = (x_1, x_2, \ldots), y = (y_1, y_2, \ldots) \in X^* \),
\[
d(x, y) := \sum_{n} \frac{d^{(n)}_H(x_n, y_n)}{2^n}
\]
It is easy to see that for all \( x, y \in X^* \), \( d(x, y) \leq 1 \). Moreover, and this is easy to verify (because it holds for \( d^{(n)}_H \) for each \( n \)), \( d\left(\frac{1}{2}x + \frac{1}{2}y, \right) = \frac{1}{2}d(x, y) \).

The space of racps \( X \) is all members of \( X^* \) that are consistent. Intuitively, \( x = (x_1, x_2, \ldots) \) is consistent if deleting the last period in the \( n \)-period problem \( x_n \) results in the \((n-1)\)-period problem \( x_{n-1} \). The space of racps, \( X \), is our domain for choice, and it follows from standard arguments that \( X \) is (linearly) homeomorphic to \( \mathcal{H}(\mathcal{T}(\Delta(C \times X))) \). We denote this homeomorphism by \( X \simeq \mathcal{H}(\mathcal{T}(\Delta(C \times X))) \). In what follows, we shall abuse notation and use \( d \) as a metric both on \( X \) as well as \( \mathcal{H}(\mathcal{T}(\Delta(C \times X))) \). It will be clear from the context precisely which space we are interested in, so there should be no cause for confusion.

There is a natural notion of inclusion in the space of racps: For \( x, y \in X \), \( y \subset x \) if \( y_n \subset x_n \) for all \( n \geq 1 \).

### A.3. Recursive Information Constraints

Recall that \( \mathcal{P} \) is the space of all partitions of \( S \), where a typical partition is \( P \). The partition \( P \) is finer than the partition \( Q \) if every cell in \( Q \) is the union of cells in \( P \). For a partition \( P \), define its entropy \( H(P) := -\sum_{J \in P} \mu(J) \log \mu(J) \). Then, we can define a metric \( d \) on \( \mathcal{P} \) as \( d(P, Q) := 2H(P \land Q) - H(P) - H(Q) \). In Section H of the Supplementary Appendix, we show that \( d \) is indeed a metric. Thus, \( (\mathcal{P}, d) \) is a metric space.

Let \( \Omega_1 := \mathcal{H}(\mathcal{P}) \), and define recursively for \( n > 1 \), \( \Omega_n := \mathcal{H}_0(\mathcal{P} \times \Omega_{n-1}^S) \) (see Section 2.4 for a definition of \( \mathcal{H}_0 \)). Then, we can set \( \Omega := \times_{n=1}^{\infty} \Omega_n \). A typical member of \( \Omega_n \) is \( \omega_n \), while \( \omega_n = (\omega_n, s) \in S \) denotes a typical member of \( \Omega_n^S \).

Let \( \psi_1 : \mathcal{P} \times \Omega_1^S \rightarrow \mathcal{P} \) be given by \( \psi_1(P, \omega_1) = P \), and define \( \psi_1 : \Omega_1 \rightarrow \Omega_1 \) as \( \psi_1(\omega_2) := \{\psi_1(P, \omega_1) : (P, \omega_1) \in \omega_2\} \). Now define recursively, for \( n > 1 \), \( \psi_n : \mathcal{P} \times \Omega_n^S \rightarrow \mathcal{P} \times \Omega_{n-1}^S \) as \( \psi_n(P, \omega_n) := (P, (\psi_{n-1}(\omega_n, s)))_S \), and \( \psi_n : \Omega_{n+1} \rightarrow \Omega_n \) by \( \psi_n(\omega_{n+1}) := \{\psi_n(P, \omega_n) : (P, \omega_n) \in \omega_{n+1}\} \).

(47) See also Gul and Pesendorfer (2004) for a more formal definition in a related setting.
An \( \omega \in \Omega' \) is consistent if \( \omega_{n-1} = \Psi_{n-1}(\omega_n) \) for all \( n > 1 \). A Recursive Information Constraint is a consistent element in \( \Omega \). The set of Recursive Information Constraints (\( \text{RICS} \)) is

\[
\Omega := \{ \omega \in \Omega' : \omega \text{ is consistent} \}
\]

that is, the set of \( \text{RICS} \) is the space of all consistent elements of \( \Omega' \).

Notice that \( \Omega_1 \) is a compact metric space when endowed with the Hausdorff metric. Then, inductively, \( \mathcal{P} \times \Omega_n^S \) with the product metric is a compact metric space, so that endowing \( \Omega_n \) with the Hausdorff metric in turn makes it a compact metric space. Thus, \( \Omega \) endowed with the product metric is a compact metric space. (Moreover, \( \Omega \) is isomorphic to the Cantor set, ie, it is separable and completely disconnected.)

Therefore, for \( \omega, \omega' \in \Omega \), where \( \omega := (\omega_n)_{n=1}^\infty \) and \( \omega' := (\omega'_n)_{n=1}^\infty \), \( \omega \neq \omega' \) if, and only if, there is a smallest \( N \geq 1 \) such that for all \( n < N \), \( \omega_n = \omega'_n \), but \( \omega_N \neq \omega'_N \).

**Theorem 3.** The set \( \Omega \) is homeomorphic to \( \mathcal{K}_0(\mathcal{P} \times \Omega^S) \).

We write the homeomorphism as \( \Omega \simeq \mathcal{K}_0(\mathcal{P} \times \Omega^S) \). The theorem is not proved, though it can be in a straightforward way, by adapting the arguments in Mariotti, Meier, and Piccione (2005).

### A.4. Representation

We now prove Proposition 2.2 for the case of canonical \( \text{MCIS} \), ie, \( \text{RICS} \). The extension to the case of general \( \text{MCIS} \) is straightforward. In what follows, let \( \mathcal{C}(X \times \Omega \times (S \cup \{s_0\})) \) be the space of continuous functions over \( X \times \Omega \times (S \cup \{s_0\}) \) endowed with the supremum norm.

**Proposition A.1.** There is a unique value function \( V \in \mathcal{C}(X \times \Omega \times (S \cup \{s_0\})) \) satisfying [Val] that represents \( \text{DM} \)'s preference over \( \text{RACPS} \). Moreover, there is an optimal dynamic information plan.

**Proof.** Define the operator \( T : \mathcal{C}(X \times \Omega \times (S \cup \{s_0\})) \rightarrow \mathcal{C}(X \times \Omega \times (S \cup \{s_0\})) \) as follows:

\[
TW(x, \omega, s') = \max_{(P, \omega') \in \omega} \sum_{f \in F} \max_{c \in \mathcal{C}} \sum_{s \in S} E^{f(s)} [u_s(c) + \delta W(y, \omega'_s, s)] \pi'_s(s | I) \pi'_s(I)
\]

Recall that \( x \) is compact. It follows from the Theorem of the Maximum (using standard arguments) that \( T \) is well defined. It is also easy to see that \( T \) is monotone (ie, \( W \leq W' \) implies \( TW \leq TW' \)) and satisfies discounting (ie, \( T(W + c) \leq TW + \delta \omega \)), so \( T \) is a contraction mapping with modulus \( \delta \in (0, 1) \). Therefore, for each \( \text{DM} \) who is characterized by \( (\omega_s)_{s \in S}, \Pi, \delta, \omega \), there exists a unique \( V \in \mathcal{C}(X \times \Omega \times (S \cup \{s_0\})) \) that satisfies the functional equation [Val].

The optimal dynamic information plan is merely the mapping \( (x, \omega, s') \mapsto (P, \omega') \in \omega \). Because the set of such \( (P, \omega') \) is finite, it follows that there is a conserving plan. But given that \( C \) is bounded and because of discounting, this implies that the conserving plan is actually optimal — see Orkin (1974) or Proposition A6.8 of Kreps (2012).

---

(48) A plan (respectively, an action) at some date and state is conserving if it achieves the supremum in Bellman’s equation. See, for instance, Kreps (2012).
A.5. Recursive Blackwell Order

Let $\hat{\omega} \in \Omega$ denote the ríc that delivers the coarsest partition in each period in every state. Define $\hat{\Omega}_0 := \mathcal{K}_0(\mathcal{P} \times \{\hat{\omega}\})$, and inductively define $\hat{\Omega}_{n+1} := \mathcal{K}_n(\mathcal{P} \times \hat{\Omega}_n)$ for all $n \geq 0$. Notice that for all $n \geq 0$, $\hat{\Omega}_n \subset \hat{\Omega}_{n+1}$. We now define an order $\succsim_0$ on $\hat{\Omega}_0$ as follows: $\omega_0 \succsim_0 \omega_0'$ if for all $(P', \hat{\omega}) \in \omega_0'$, there exists $(P, \hat{\omega}) \in \omega_0$ such that $P$ is finer than $P'$. This allows us to define inductively, for all $n \geq 1$, an order $\succsim_n$ on $\hat{\Omega}_n$. For all $\omega_n, \omega_n' \in \hat{\Omega}_n$, $\omega_n \succsim_n \omega_n'$ if for all $(P', \omega_{n-1}') \in \omega_n'$, there exists $(P, \omega_n-1) \in \omega_n$ such that (i) $P$ is finer than $P'$, and (ii) $\omega_n \succsim_{n-1} \omega_{n-1,s}$ for all $s \in S$.

It is easy to see that $\succsim_n$ is reflexive and transitive for all $n$. There is a natural sense in which $\omega_n$ extends $\omega_{n-1}$, as we show next.

**Lemma A.2.** For all $n \geq 0$, $\succsim_{n+1}$ extends $\succsim_n$, ie, $\omega_n \succsim_{n+1} \omega_n$.

**Proof.** As observed above, $\hat{\Omega}_n \subset \hat{\Omega}_{n+1}$ for all $n$. First consider the case of $n = 0$ and recall that by construction $\hat{\omega} \in \hat{\Omega}_0$. Let $\omega_0 \succsim_0 \omega_0'$. Then, for $(P', \hat{\omega}) \in \omega_0'$, there exists $(P, \hat{\omega}) \in \omega_0$ such that $P$ is finer than $P'$. Moreover, because $\succsim_0$ is reflexive, $\omega \succsim_0 \omega$. But this implies $\omega_0 \succsim_1 \omega_0'$. Conversely, let $\omega_0 \succsim_1 \omega_0'$. Then, for all $(P', \hat{\omega}) \in \omega_0'$, there exists $(P, \hat{\omega}) \in \omega_0$ such that (i) $P$ is finer than $P'$, and (ii) $\omega \succsim_0 \omega$ for all $s \in S$. This allows us to define inductively, for all $n \geq 1$, $\omega_n \succsim_{n+1} \omega_n'$.

As our inductive hypothesis, we suppose that $\omega_n \succsim_{n-1} \omega_n'$. Then, for all $(P', \omega_{n-1}') \in \omega_n'$, there exists $(P, \omega_{n-1}) \in \omega_n$ such that (i) $P$ is finer than $P'$, and (ii) $\omega_{n-1,s} \succsim_{n-1,s} \omega_{n-1,s}'$ for all $s \in S$. By the induction hypothesis, this is equivalent to $\omega_{n-1,s} \succsim_{n-1,s} \omega_{n-1,s}'$ for all $s \in S$, which implies that $\omega_n \succsim_{n+1} \omega_n'$.

Let $\hat{\Omega} := \bigcup_{n \geq 0} \hat{\Omega}_n$. Let $\succsim$ be a partial order defined on $\hat{\Omega}$ as follows: $\omega \succsim \omega'$ if there is $n \geq 1$ such that $\omega, \omega' \in \Omega_n$ and $\omega \succsim_n \omega'$.

By definition of $\hat{\Omega}$, there is some $n$ such that $\omega, \omega' \in \hat{\Omega}_n$, and by Lemma A.2, the precise choice of this $n$ is irrelevant. This implies $\succsim$ is well defined. We now show that $\succsim$ has a recursive definition as well.

**Proposition A.3.** For any $\omega, \omega' \in \hat{\Omega}$, the following are equivalent.

(a) $\omega \succsim \omega'$.
(b) for all $(P', \omega') \in \omega'$, there exists $(P, \omega) \in \omega$ such that (i) $P$ is finer than $P'$, and (ii) $\omega_s \succsim \omega'_s$ for all $s \in S$.

Therefore, $\succsim$ is the unique partial order on $\hat{\Omega}$ defined as $\omega \succsim \omega'$ if (b) holds.

**Proof.** (a) implies (b). Suppose $\omega \succsim \omega'$. Then, by definition, there exist $n$ such that $\omega, \omega' \in \hat{\Omega}_n$, $\omega \succsim_n \omega'$. This implies that for all $(P', \omega_{n-1}') \in \omega_n'$, there exists $(P, \omega_{n-1}) \in \omega_n$ such that (i) $P$ is finer than $P'$, and (ii) $\omega_{n-1,s} \succsim_{n-1,s} \omega_{n-1,s}'$ for all $s \in S$. But the latter property implies $\omega_s \succsim \omega'_s$ for all $s \in S$, which established (b). The proof that (b) implies (a) is similar and is therefore omitted.

The uniqueness of $\succsim$ on $\hat{\Omega}$ follows immediately from the uniqueness of $\succsim_n$ for all $n \geq 0$. 

32
We can now prove the existence of a recursive order on $\Omega$. (Notice that $\mathrm{cl}(\hat{\Omega}) = \Omega$.) In particular, for all $\omega, \omega' \in \Omega$, we say that $\omega$ recursively Blackwell dominates $\omega'$ if for all $(P', \tilde{\omega}') \in \omega'$, there exists $(P, \hat{\omega}) \in \omega$ such that (i) $P$ is finer than $P'$, and (ii) $\tilde{\omega}_s$ recursively Blackwell dominates $\hat{\omega}_s$ for all $s \in S$. The following proposition characterizes a natural recursive Blackwell order.

**Proposition A.4.** The order $\succeq$ on $\hat{\Omega}$ has a unique continuous extension to $\Omega$, also denoted by $\succeq$. Moreover, on $\Omega$, $\succeq$ is the unique non-trivial and continuous recursive Blackwell order.

*Proof.* Because $\Omega = \mathrm{cl}(\hat{\Omega})$, we simply extend $\succeq$ to $\Omega$ by re-defining it to be $\mathrm{cl}(\hat{\omega})$. It is easy to see that $\succeq$ so defined is continuous and non-trivial. That $\succeq$ is a unique recursive Blackwell order follows immediately from the facts that $\hat{\Omega}$ is dense in $\Omega$, the continuity of $\succeq$, and Proposition A.3. $\square$

Let $\text{proj}_n : \Omega \rightarrow \hat{\Omega}_n$ be the natural map associating with each $\omega$, the ‘truncated and concatenated’ version $\omega_n$ which offers the same choices of partition as $\omega$ for $n$ stages, but then offers $\hat{\omega}$, the coarsest partition, forever. It is easy to see that given $\omega \in \Omega$, the sequence $(\omega_n)$ is Cauchy, and converges to $\omega$. The next corollary gives us an easy way to establish dominance.

**Corollary A.5.** For $\omega, \omega' \in \Omega$, $\omega \succeq \omega'$ if, and only if, for all $n \in \mathbb{N}$, $\omega_n \succeq \omega'_n$.

*Proof.* The ‘only if’ part is straightforward. The ‘if’ part follows from the continuity of $\succeq$. $\square$

Notice that if $m \geq n$, then $\omega_n = \text{proj}_n \omega = \text{proj}_n \omega_m$. This observation implies the following corollary.

**Corollary A.6.** For all $\omega, \omega' \in \Omega$ and $m \geq n \geq 1$, $\omega_m \succeq \omega'_m$ implies $\omega_n \succeq \omega'_n$ for all $1 \leq n \leq m$.

*Proof.* Notice that $\omega_m, \omega'_m \in \Omega$. Therefore, by Corollary A.5, it follows that for all $n \geq 1$, $\text{proj}_n \omega_m \succeq \text{proj}_n \omega'_m$.

A.6. Isomorphisms of Mics

*Proof of Proposition 2.4.* We first show that (a) implies (b). Towards this end, let $\mathcal{M} = (\Theta, \theta_0, \mathcal{P}, \Gamma, \tau)$ be a mic. Recall the definition of the space $\Omega_n$ from Appendix A.3 and define the maps $\Phi_n : \Theta \rightarrow \Omega_n$ as follows. Let

- $\Phi_1(\theta) := \Gamma(\theta)$,
- $\Phi_2(\theta) := \{(P, (\Phi_1(\tau(P, \theta, s)))_{s \in S}) : P \in \Gamma(\theta)\}$,
- $\vdots$
- $\Phi_{n+1}(\theta) := \{(P, (\Phi_n(\tau(P, \theta, s)))_{s \in S}) : P \in \Gamma(\theta)\}$.

It is easy to see that for each $\theta \in \Theta$, $\Phi_n(\theta) \in \Omega_n$, ie, $\Phi_n$ is well defined.

Now, given $\theta_0$, set $\Phi_n(\theta_0) := (\omega_n) \in \Omega_n$. It is easy to see that the sequence

$$(\omega_1, \omega_2, \ldots, \omega_n, \ldots) \in \prod_{n \in \mathbb{N}} \Omega_n$$

is consistent in the sense described in Appendix A.3. Therefore, there exists $\omega \in \Omega$ such that $\omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots)$, ie, the MDP $\mathcal{M}$ corresponds to an ric $\omega$. 

33
To see that (b) implies (a), let \( \omega \in \Omega \). A partition \( P \) is supported by \( \omega \) if there exists \( \omega' \in \Omega^S \) such that \((P, \omega') \in \omega \). Now set \( \Theta = \Omega, \theta_0 = \omega, \Gamma^*(\theta) = \{ P : P \) is supported by \( \theta \} \), and \( \tau^*(P, \omega, s) = \omega'_s \) where \( \omega' \in \Omega^S \) is the unique collection of mics such that \((P, \omega') \in \omega \). This results in the mic \( \omega_M = (\Theta, \Gamma^*, \tau^*, \theta_0 = \omega) \) that is uniquely determined by \( \mathcal{M} \). \( \square \)

Thus, \( \Omega \) is the space of canonical mics in that every mic can be embedded in \( \Omega \). Let \( \mathcal{M} = (\Theta, \Gamma, \tau, \theta_0) \) and \( \mathcal{M}' = (\Theta', \Gamma', \tau', \theta'_0) \) be two mics in \( \mathcal{M} \). Define the function \( D : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \) as follows:

\[
D(\mathcal{M}(\theta_0), \mathcal{M}'(\theta'_0)) := \\
\max \left[ d_H(\Gamma(\theta_0), \Gamma'(\theta'_0)) \wedge 1, \frac{1}{2} \max_{P \in \Gamma(\theta_0), s \in S} D(\mathcal{M}(\tau(\theta_0, P, s)), \mathcal{M}'(\tau'(\theta'_0, P, s))) \right]
\]

where \( \mathcal{M}(\theta) \) denotes the mic \( \mathcal{M} \) with initial state \( \theta \). The function \( D \) captures the discrepancy between the mics \( \mathcal{M} \) and \( \mathcal{M}' \). In what follows, let \( B(\mathcal{M} \times \mathcal{M}) \) denote the space of real-valued bounded functions defined on \( \mathcal{M} \times \mathcal{M} \) with the supremum norm.

**Lemma A.7.** There is a unique function \( D \in B(\mathcal{M} \times \mathcal{M}) \) that satisfies equation [A.1].

**Proof.** Consider the operator \( T : B(\mathcal{M} \times \mathcal{M}) \to B(\mathcal{M} \times \mathcal{M}) \) defined as

\[
TD' (\mathcal{M}(\theta_0), \mathcal{M}'(\theta'_0)) := \\
\max \left[ d_H(\Gamma(\theta_0), \Gamma'(\theta'_0)) \wedge 1, \frac{1}{2} \max_{P \in \Gamma(\theta_0), s \in S} D(\mathcal{M}(\tau(\theta_0, P, s)), \mathcal{M}'(\tau'(\theta'_0, P, s))) \right]
\]

for all \( D' \in B(\mathcal{M} \times \mathcal{M}) \). It is easy to see that \( T \) is monotone in the sense that \( D_1 \leq D_2 \) implies \( TD_1 \leq TD_2 \). It also satisfies discounting, i.e., \( T(D + a) \leq TD + \frac{1}{2}a \) for all \( a \geq 0 \). This implies that \( T \) has a unique fixed point in \( B(\mathcal{M} \times \mathcal{M}) \), and this fixed point \( D \) satisfies [A.1]. \( \square \)

We can now define an isomorphism between mics. Two mics \( \mathcal{M} \) and \( \mathcal{M}' \) are indistinguishable if \( D(\mathcal{M}(\theta_0), \mathcal{M}'(\theta'_0)) = 0 \). Intuitively, indistinguishable mics have the same set of choices of partitions after any history of choice, and so offer the same set of plans. We now have an easy, recursive characterization of indistinguishability.

**Lemma A.8.** Let \( \mathcal{M}, \mathcal{M}' \in \mathcal{M} \). Then, \( \mathcal{M} \) is indistinguishable from \( \mathcal{M}' \) if, and only if, (i) \( \Gamma(\theta_0) = \Gamma'(\theta'_0) \), and (ii) for all \( P \in \Gamma(\theta_0) \cap \Gamma'(\theta'_0) \) and \( s \in S \), the mic \( (\Theta, \Gamma, \tau, \tau(\theta_0, P, s)) \) is indistinguishable from the mic \( (\Theta', \Gamma', \tau', \tau'(\theta'_0, P, s)) \).

The proof follows immediately from the definition of the discrepancy \( D \) and so is omitted. We now regard \( \Omega \) as the canonical space of mics and each \( \omega \) as a canonical mic. In other words, every \( \omega \) is the canonical mic \( (\Omega, \Gamma^*, \tau^*, \omega) \).

**Corollary A.9.** Let \( \omega, \omega' \in \Omega \). Then, \( \omega \neq \omega' \) implies \( D(\omega, \omega') > 0 \).

**Proof.** It is easy to see that if \( D(\omega, \omega') = 0 \), then \( \omega_n = \omega'_n \) for all \( n \geq 1 \), which implies \( \omega = \omega' \), as required. \( \square \)
Corollary A.10. Let \( \omega, \omega' \in \Omega \) be such that \( \text{proj}_n(\omega) \not\succeq \text{proj}_n(\omega') \) for some \( n \geq 1 \), but for all \( m < n, \text{proj}_m(\omega) \succeq \text{proj}_m(\omega') \). Then, there exists finite sequences \( (P_k)_{k=1}^{n-1} \) and \( (s_k)_{k=1}^{n-1} \) which induce RICs \( \omega^i_{(n-k)} := \tau^*(\omega^i_{(n-k+1)}), P_k, s_k) \in \hat{\Omega}_{n-k} \) where \( P_k \in \Gamma^*(\omega^i_{(n-k+1)}) \), such that \( \Gamma^*(\omega^1) \) does not setwise-Blackwell dominate \( \Gamma^*(\omega^2) \).

**Proof.** If not, we would have \( \omega^1 \succeq \omega^2 \), a contradiction. \( \square \)

Let \( \triangleright \) be a recursive order on \( M \) defined as follows: For mics \( M = (\Theta, \Gamma, \tau, \theta_0) \) and \( M' = (\Theta', \Gamma', \tau', \theta'_0) \), 
\[
M \triangleright M' \text{ if for every } P' \in \Gamma'('0'), \text{ there exists } P \in \Gamma(\theta_0) \text{ such that (i) } P \text{ is finer than } P', \text{ and (ii) } (\Theta, \Gamma, \tau(\theta_0, P, s)) \triangleright (\Theta', \Gamma', \tau'(\theta'_0, P', s)) \text{ for all } s \in S.
\]

It is easy to see that such a recursive order exists. Indeed, for any mci \( M \), let \( \omega_M \) denote the canonical mci that is indistinguishable from it. (By Corollary A.9 there is a unique such \( \omega_M \).) The recursive Blackwell order is induced on \( M \) as follows: \( M \) recursively Blackwell dominates \( M' \) if (and only if) \( \omega_M \succeq \omega_{M'} \). The recursive Blackwell order on \( M \) clearly satisfies the condition [\( \star \)]. We now demonstrate that it is the *largest* order that satisfies [\( \star \)].

**Proposition A.11.** Let \( \triangleright \) be a recursive order on \( M \) that satisfies [\( \star \)]. If \( M \triangleright M' \), then \( M \) recursively Blackwell dominates \( M' \).

**Proof.** We will prove the contrapositive. If \( M \) does not recursively Blackwell dominate \( M' \), then \( \omega_M \not\succeq \omega_{M'} \). Corollaries A.5 and A.6 imply that there is a smallest \( n \) such that \( \omega_{M,n} \not\succeq \omega_{M',n} \) but that for all \( m < n, \omega_{M,m} \succeq \omega_{M',m} \) (where \( \omega_{M,n} = \text{proj}_n(\omega_M) \) as defined in Appendix A.5). From Corollary A.10 it follows that there exists a finite sequence of partitions \( (P_k) \) and states \( (s_k) \) such that \( \Gamma^*(\tau^{(n)}(\theta_0, (P_k), (s_k))) \) does not setwise Blackwell dominate \( \Gamma^*(\tau^{(n)}(\theta'_0, (P_k), (s_k))) \), where \( \tau^{(n)}(\theta_0, (P_k), (s_k)) \) represents the \( n \)-stage transition following the sequence of choices \( (P_k) \) and states \( (s_k) \). Now recall that \( M \) is indistinguishable from \( \omega_M \), and so is \( M' \) from \( \omega_{M'} \). This implies \( \Gamma'(\tau^{(n)}(\theta_0, (P_k), (s_k))) \) does not setwise Blackwell dominate \( \Gamma'(\tau^{(n)}(\theta'_0, (P_k), (s_k))) \). Thus, it must necessarily be that \( M \not\triangleright M' \). \( \square \)

### A.7. Consumption Streams and the raa Representation

The space \( L \) is defined as \( L \simeq \mathcal{F}(\Delta(C \times L)) \) and is a closed subspace of \( X \) (with the natural embedding).

Let \( u_s \in C(C) \) for all \( s \in S, \delta \in (0, 1), \Pi \) represent the transition operator for a fully connected Markov process on \( S \), and \( \pi_0 \) be the unique invariant distribution of \( \Pi \). A preference on \( L \) has a Recursive Anscombe-Aumann (raa) representation \((u_s)_{s \in S}, \Pi, \delta \) if \( W_0(\cdot) := \sum_s W(\cdot, s)\pi_0(s) \) represents it, where \( W(\cdot, s) \) is defined recursively as
\[
W(\ell; s) = \sum_{s' \in S} \Pi(s, s')[u_s(\ell_1(s')) + \delta W(\ell_2(s'); s')]
\]
and where \( u_s \) non-trivial for some \( s \in S \). Then, \( W_0 \) can also be written as
\[
W_0(\ell) = \sum_{s' \in S} \pi_0(s)[u_s(\ell_1(s)) + \delta W(\ell_2(s); s)]
\]
because $\pi_0$ is the unique invariant distribution of $\Pi$ and therefore satisfies $\pi_0(s) = \sum_s \pi_0(s') \Pi(s', s)$. The preference on $L$ has a standard RAA representation $((u_s)_{s \in S}, \Pi, \delta)$ if we also have $u_s(c^+_s) = 0$ for all $s \in S$ for some fixed $c^+_s \in C$.

We show in Section I of the Supplementary Appendix that $\succsim |_L$ has an RAA representation as described above. We cannot directly appeal to Corollary 5 from Krishna and Sadowski (2014) because they only consider finitely many prizes. Nonetheless, judicious and repeated applications of Corollary 5 of KS allows us to reach the same conclusion for a compact set of prizes.

It is clear that $L$ is compact, so the continuity of $\succsim$ implies that there exist $\succsim$-maximal and $\succsim$-minimal elements of $L$. These we call $\ell^*$ and $\ell_*$. Moreover, given that $\succsim |_L$ has an RAA representation as described above, for each $s \in S$, we let $c^+_s := \arg \max_{c \in C} u_s(c)$ and $c^-_s := \arg \min_{c \in C} u_s(c)$. Because each $u_s$ is continuous, such a $c^+_s$ and $c^-_s$ must exist. Now, define $f^+(s) := c^+_s$ — the Dirac measure concentrated at $c^+_s$ — for all $s \in S$, and similarly, define $f^-(s) := c^-_s$ for all $s \in S$. Then, $\ell^*$ is the (unique) consumption stream that delivers $f^+$ at each date and $\ell_*$ is the (unique) consumption stream that delivers $f^-$ at each date. Observe that the best and worst consumption streams are deterministic, and that for all $\alpha \in \Delta(C)$, $u_s(c^-_s) \leq u_s(\alpha) \leq u_s(c^+_s)$. An immediate consequence of this is that for any $c \in C$, $\ell \in L$ and $s \in S$, $(c, \ell^*) \succsim (c, \ell) \succsim (c, \ell_*)$. Lipschitz Continuity (Axiom 1(c)) implies that $\ell^* \succsim \ell_*$ (see Corollary E.3 in the Supplementary Appendix), so $(c, \ell^*) \succsim (c, \ell_*)$.

### A.8. Self-Generating Representations and Dynamic Plans

Recall that $C(X)$ is the space of all real-valued continuous functions on $X$. Let $\ell^+ \in L$ be the consumption stream that delivers $c^+_s$ in state $s$ at every date.

Suppose $(u_s, \emptyset, (v_s(\cdot, P)), \pi)$ is a tuple where

- $u_s \in C(C)$ for all $s \in S$,
- $\emptyset \subset \mathcal{P}(S)$,
- $v_s(\cdot, P) \in BL(X)$ for all $s \in S$ and $P \in \emptyset$,
- $\pi \in \Delta(S)$,
- $u_s(c^+_s) = v_s(\ell^+, P) = 0$ for all $s \in S$ and $P \in \emptyset$,
- $v_s(\cdot, P)$ is independent of $P$ on $L$, and
- $v_s(\cdot, P)$ is non-trivial on $L$, and hence on $X$, for all $s \in S$ and $P \in \emptyset$,

and $v \in \mathbb{R}^X$ is such that

$$v(x) = \max_{P \in \emptyset} \sum_{E \in P} \pi(E) \max_{f \in F} \sum_{s \in S} \pi(s \mid E) \left[ u_s(f_1(s)) + v_s(f_2(s), P) \right]$$

In that case, we say that the tuple $((u_s, \emptyset, (v_s(\cdot, P)), \pi)$ is a separable and partitional implementation of $v$, or in short, an implementation of $v$. (By definition, the implementation takes value 0 on $\ell^+(s)$ for all $s \in S$ and is linear on $L$. In what follows, we will not explicitly state these properties.)

---

(49) The space $BL(X)$ consists of all bounded Lipschitz functions on $X$; see Appendix A.2.
More generally, for any subset \( \Phi \subset C(X) \), define the operator \( A : 2^{C(X)} \to 2^{C(X)} \) as follows:

\[
A \Phi := \left\{ v \in C(X) : \exists \left( (u_s), (v_s(.), P) \right) \text{ that implements } v \right. \\
\text{and } v_s(. , P) \in \Phi \text{ for all } s \in S \text{ and } P \in \mathcal{Q} \left\} 
\]

**Proposition A.12.** The operator \( A \) is well defined and has a largest fixed point \( \Phi^* \neq \{ \emptyset \} \). Moreover, \( \Phi^* \) is a cone.

**Proof.** It is easy to see that for all nonempty \( \Phi \subset C(X) \), \( A \Phi \) is nonempty. (Simply take any \( \emptyset \), any \( \emptyset \neq v_s(\cdot, P) \emptyset \in \Phi \) for all \( P \in \emptyset \), and any \( u_s \), so that \( A \Phi \neq \emptyset \.) The operator \( A \) is monotone in the sense that \( \Phi \subset \Phi' \) implies \( A \Phi \subset A \Phi' \). Thus, it is a monotone mapping from the lattice \( 2^{C(X)} \) to itself, where \( 2^{C(X)} \) is partially ordered by inclusion. The lattice \( 2^{C(X)} \) is complete because any collection of subsets of \( 2^{C(X)} \) has an obvious least upper bound: the union of this collection of subsets. Similarly, a greatest lower bound is the intersection of this collection of subsets (which may be empty). Therefore, by Tarski’s fixed point theorem, \( A \) has a largest fixed point \( \Phi^* \in 2^{C(X)} \).

To see that \( \Phi^* \neq \{ \emptyset \} \), ie, \( \Phi^* \) does not contain only the trivial function \( \emptyset \), Fix \( \emptyset = \{ \{ s \} : s \in S \} \) so that it contains only the finest partition of \( S \). For the value function \( V \) in \([\text{Val}]\), take any \( u_s \in C(C) \setminus \{ \emptyset \} \) with \( u_s(e^t_s) = 0 \) for all \( s \in S \), a discount factor \( \delta \in (0, 1) \), and \( \pi \) as the uniform distribution over \( S \). Then \( V \) is implemented by \( ((u_s), (\emptyset, (\delta V), \pi)) \), while \( \delta V \) is implemented by \( ((\delta u_s), (\emptyset, \delta^2 V), \pi) \), and so on. Therefore, the set \( \Phi_V := \{ \delta^n V : n \geq 0 \} \) is clearly a fixed point of \( A \). Because \( \Phi_V \subset \Phi^* \), it must be that \( \Phi^* \) is nonempty.

Finally, to see that \( \Phi^* \) is a cone, let \( v \in \Phi^* \) and suppose \( ((u_s), (\emptyset, (v_s(\cdot, P)), \pi)) \) implements \( v \). Then, for all \( \lambda \geq 0 \), \( (\lambda u_s), (\emptyset, (\lambda v_s(\cdot, P)), \pi) \) implements \( \lambda v \), ie, \( \lambda \Phi^* \) is also a fixed point of \( A \). Because \( \Phi^* \) is the largest fixed point, it must be a cone. \( \square \)

Notice that each \( v \in \Phi^* \) is implemented by a tuple \( ((u_s), (\emptyset, (v_s(\cdot, P)), \pi)) \) with the property that each \( v_s(\cdot, P) \in \Phi^* \). Adapting the terminology of Abreu, Pearce, and Stacchetti (1990), the set \( \Phi^* \) consists of self-generating preference functionals that have a separable and partitional implementation. (Notice that unlike Abreu, Pearce, and Stacchetti (1990), our self-generating set lives in an infinite dimensional space. Also, unlike Abreu, Pearce, and Stacchetti (1990), the non-emptiness of \( \Phi^* \) follows relatively easily, as noted in the proof of Proposition A.12.) In what follows, if \( \succsim \) is represented by \( V \in \Phi^* \), we shall say that \( V \) is a self-generating representation of \( \succsim \).

Given a \( V \in \Phi^* \) that is a self-generating representation of \( \succsim \), we would like to extract the underlying (subjective) informational constraints. We show next that this is possible.

**Proposition A.13.** There is a unique map \( \varphi^* : \Phi^* \to \Omega \) that satisfies for some implementation \( ((u_s), (\emptyset, (v_s(\cdot, P)), \pi)) \) of \( v \), that

\[
\varphi^*(v) := \left\{ (P, \varphi^*(v_s(\cdot, P))) : P \in \mathcal{Q} \right\} 
\]

and is independent of the implementation chosen.

**Proof.** Let \( v^{(1)} \in \Phi_1 \), and suppose \( ((u_s), (\emptyset, (v_s(\cdot, P)), \pi)) \) implements \( v^{(1)} \). In this implementation, \( \emptyset \) is unique. (The argument follows from our identification argument below in Appendix B. It is easy to see that \( (u_s), (v_s(\cdot, P)), \) and \( \pi \) will typically not be unique.) Then, define \( \varphi_1 : \Phi_1 \to \Omega_1 \) as

\[
\varphi_1(v^{(1)}) := \emptyset, \text{ where } ((u_s), (\emptyset, (v_s(\cdot, P)), \pi)) \text{ implements } v^{(1)}
\]
Proceeding iteratively, we define \( \varphi_n : \Phi_n \rightarrow \Omega_n \) as
\[
\varphi_n(v^{(n)}) := \{ (P, \varphi_{n-1}(v^{(n-1)}_s)(\cdot, P)) : \exists (u_s, \mathcal{Q}, (v^{(n-1)}_s(\cdot, P)), \pi) \text{ that implements } v^{(n)} \text{ and } P \in \mathcal{Q} \}.
\]
Notice that the same argument that established the uniqueness of \( \varphi_1 \) also applies here, to provide the uniqueness of \( \varphi_n \).

Now, suppose \( v \in \Phi^* \). This implies \( v \) has a partitional and separable implementation \((u_s, \mathcal{Q}, (v_s(.P)), \pi)\), and where each \( v_s(.) \) also has a partitional and separable implementation, and so on, ad infinitum. Then, we may define, for all \( n \geq 1 \), \( \omega^{(n)} := \varphi_n(v) \). Now consider the infinite sequence \( \omega_0 := (\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(n)}, \ldots) \in \Omega \).

In particular, this allows us to define the map \( \varphi^* : \Phi^* \rightarrow \Omega \) as \( \varphi^*(v) = (\varphi_1(v), \varphi_2(v), \ldots) \), which extracts the underlying \( \text{ric} \) from any function \( v \in \Phi^* \), independent of the other components of the implementation, as claimed.

To recapitulate, we can now extract an \( \text{ric} \) from a self-generating representation. In other words, the identification of the \( \text{ric} \) \( \omega_0 \) doesn’t depend on the recursivity of the value function. This stands in contrast to the identification of the other preference parameters, which relies on recursivity. For a self-generating representation, we can find a (not necessarily unique) probability measure \( \pi \) over \( S^\infty \). The formal details are straightforward and hence omitted.

A dynamic plan consists of two parts: the first entails picking a partition for the present period (and the corresponding continuation constraint), and the second entails picking an act from \( x \), whilst requiring that the choice of act, as a function of the state, be measurable with respect to the chosen partition. The first part is a dynamic information plan while the second is a dynamic consumption plan.

An \( n \)-period history is an (ordered) tuple
\[
h_n = ((x_0, \omega_0), \ldots, (P^{(n-1)}, f^{(n-1)}, s^{(n-1)}, x^{(n-1)}, \omega^{(n-1)}))
\]
Let \( \delta_n \) denote the collection of all \( n \)-period histories.

Formally, a dynamic information plan is a sequence \( \sigma_i = (\sigma_i^{(1)}, \sigma_i^{(2)}, \ldots) \) of mappings where \( \sigma_i^{(n)} : \delta_n \rightarrow \mathcal{P} \times \Omega S \). Similarly, a dynamic consumption plan is a sequence \( \sigma_c = (\sigma_c^{(1)}, \sigma_c^{(2)}, \ldots) \) of mappings where \( \sigma_c^{(n)} : \delta_n \rightarrow \mathcal{F}(\Delta(C \times X)) \). A dynamic plan \( \sigma \) is just a pair \( \sigma = (\sigma_i, \sigma_c) \).

A dynamic plan \( \sigma = (\sigma_i, \sigma_c) \) is feasible if (i) \( \sigma_i^{(n)}(h_n) \in \omega^{(n-1)} \), (ii) \( \sigma_c^{(n)}(h_n) \in x^{(n-1)} \), and (iii) given the information plan \( \sigma_i^{(n)}(h_n) = (P, \omega') \in \omega^{(n-1)} \), \( \sigma_c^{(n-1)}(h_n) \) is \( P \)-measurable, ie, for all \( I \in P \) and for all \( s, s' \in I \), \( \sigma_c^{(n)}(h_n)(s) = \sigma_c^{(n)}(h)(s') \).

Each dynamic plan along with initial states \( (x_0, \omega_0, \pi_0) \) induces a probability measure over \( (X \times \Omega \times S)^\infty \) or, put differently, an \( X \times \Omega \times S \) valued process. Let \( (x^{(n)}, \omega^{(n)}, s^{(n)}) \) be the \( X \times \Omega \times S \)-valued stochastic process of racps, rics, and objective states induced by a dynamic plan, where \( x^{(n)} \in X \) is the racp beginning at period \( n + 1 \), \( \omega^{(n)} \in \Omega \) is the ric beginning at period \( n + 1 \), and \( s^{(n)} \in S \) is the state beginning period \( n + 1 \). A dynamic plan is stationary if \( \sigma^{(n)}(h_n) \) only depends on \( (x^{(n-1)}, \omega^{(n-1)}, s^{(n-1)}) \).

(50) Of course, the choice of plan doesn’t affect the evolution of the objective states \( s^{(n)} \).
For a fixed $V \in \Phi^*$, let $v^{(n)}(\cdot, \omega^{(n)}, s^{(n)}, \sigma)$ denote the value function that corresponds to the $n$-th period implementation of $V$ when following the dynamic information plan $\sigma$, where $\omega^{(n)} = \varphi_n(V)$ as in Proposition A.13 and $s^{(n)}$ is the state in period $n$.

While we have shown that each $v \in \Phi^*$ can be written as the sum of some instantaneous utility and some continuation utility function that also lies in $\Phi^*$, we nonetheless need to verify that the value that $V$ obtains for any menu is indeed the infinite sum of consumption utilities. We verify this next.

Proposition A.14. Let $V \in \Phi^*$, and suppose $v^{(n)}(\cdot, \omega^{(n)}, s^{(n)}, \sigma)$ is defined as above. Then, for any feasible dynamic plan $\sigma = (\sigma_c, \sigma_i)$, we have

$$\lim_{n \to \infty} \left\| E^{\sigma, \pi} v^{(n)}(\cdot, \omega^{(n)}, s^{(n)}, \sigma) \right\|_\infty = 0$$

**Proof.** Consider $V \in \Phi^*$ with Lipschitz rank $\lambda$. Recall that for any $x \in X$, $\ell^\dagger \omega_n x \in X$ denotes the RACP that delivers $\ell^\dagger$ in every period until period $n - 1$ and then, in period $n$, in every state, delivers $x$. Recall further that $X$ is an infinite product space, and by the definition of the product metric (see Appendix A.2), it follows that for any $\varepsilon > 0$, there exists an $N > 0$ such that for all $x, y \in X$ and $n \geq N$, $d(\ell^\dagger \omega_n x, \ell^\dagger \omega_n y) < \varepsilon / \lambda$. Lipschitz continuity of $V$ then implies $|V(\ell^\dagger \omega_n x) - V(\ell^\dagger \omega_n y)| < \varepsilon$.

For a given $n$, $V(\ell^\dagger \omega_n x) = 0 + E^{\pi}[v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma)]$, which implies

$$\left| E^{\sigma, \pi} v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) - E^{\sigma, \pi} v^{(n)}(y, \omega^{(n)}, s^{(n)}, \sigma) \right| < \varepsilon$$

for all $n \geq N$. Recall that

$$\left\| E^{\sigma, \pi} v^{(n)}(\cdot, \omega^{(n)}, s^{(n)}, \sigma) \right\|_\infty = \sup_x \left| E^{\sigma, \pi} v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) \right|$$

Moreover, we have

$$\sup_x \left| E^{\sigma, \pi} v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) \right| = \sup_x \left| E^{\sigma, \pi} v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) - v^{(n)}(\ell^\dagger, \omega^{(n)}, s^{(n)}, \sigma) \right| < \varepsilon$$

which completes the proof. \hfill \square

Adapting the terminology of Dubins and Savage (1976), we shall say that a function $V \in \Phi^*$ is equalizing if [A.2] holds. (To be precise, if [A.2] holds, then every dynamic plan is equalizing in sense of Dubins and Savage (1976).)

Given an initial $(x, \omega) \in X \times \Omega$, each $\sigma$ induces a probability measure over $X_n \delta_n$, the space of all histories. It also induces a unique consumption stream $\ell_{\sigma(x, \omega)}$ that delivers consumption $\sigma_c(h_n)(s')$ after history $h_n$ in state $s'$ in period $n$. We show next that for any self-generating preference functional $V \in \Phi^*$, the utility from following the plan $\sigma$ given the RACP $x$ is the same as the utility from the consumption stream $\ell_{\sigma(x, \omega)}$. (Of course, given the consumption stream $\ell_{\sigma(x, \omega, s)}$, there are no consumption choices to be made.) Moreover, there is an optimal plan such that following this plan induces a consumption stream that produces the same utility as the RACP $x$.

Let $\Sigma$ denote the collection of all dynamic plans and let $L_{x, \omega} := \{ \ell_{\sigma(x, \omega)} : \sigma \in \Sigma \}$ be the collection of all consumption streams so induced by the RACP $x$ and the RIC $\omega$. In what follows, $V(x, \sigma)$ is the expected utility from following the dynamic plan $\sigma$ given the RACP $x$. 

39
**Lemma A.15.** Let $V \in \Phi^*$ be such that $\varphi^*(V) = \omega$. Then, for all $x \in X$, $V(x, \sigma) = V(\ell_{\sigma(x, \omega)})$ and $V(x) = \max_{\sigma \in \Sigma} V(x, \sigma) = \max_{\ell \in \mathcal{L}_{x, \omega}} V(\ell)$.

These are analogues of standard statements in dynamic programming, as the following proof demonstrates.

**Proof.** For $V \in \Phi^*$ and for any plan $\sigma'$, an agent with the utility function $V$ is indifferent between following $\sigma'$ and the consumption stream $\ell_{\sigma'(x, \omega)}$. This is essentially an adaptation of Theorem 9.2 in Stokey, Lucas, and Prescott (1989) where their equation 7 — which is also known as a no-Ponzi game condition, see Blanchard and Fischer (1989, p 49) — is replaced by the fact that $V$ is equalizing (condition [A.2] in Proposition A.14).

To see that there is an optimal plan, notice that $x$ is a compact set of acts, and because there are only finitely many partitions of $S$, it is possible to find a conserving action at each date after every history. This then gives us a conserving plan (see Footnote 48). We can now adapt Theorem 9.2 in Stokey, Lucas, and Prescott (1989) where, as above, their equation 7 is replaced by [A.2], to show that $\sigma$ is indeed an optimal plan. Loosely put, we have just shown that because the plan is conserving and because $V$ is equalizing, the plan must be optimal. This corresponds to the characterization of optimal plans in Theorem 2 of Karatzas and Sudderth (2010).

---

**B. Identification and other Proofs from Section 2.3**

Recall that $x$ is strongly aligned with $\omega$ if (i) $V(x, \omega, \pi_0) \geq V(x, \omega', \pi_0)$ for all $\omega' \in \Omega$, and (ii) $\omega'$ does not recursively Blackwell dominate $\omega$ implies $V(x, \omega, \pi_0) > V(x, \omega', \pi_0)$. We say that $P$ is supported by $\omega$ if there exists $\omega' \in \Omega^S$ such that $(P, \omega') \in \omega$. Given an $\omega$, we now consider an RACP with a number of agreeable properties.

**Lemma B.1.** Let $(P, \omega') \in \omega$. Then, there exists an RACP $x(P, \omega')$ recursively defined as

$$
\boxed{[\star]} \quad x(P, \omega') = \{f_J : J \in P\} \quad \text{with} \quad f_J := \begin{cases} 
\ell^*(s) & \text{if } s \in J \\
\ell_*(s) & \text{if } s \notin J
\end{cases}
$$

where Unif($\cdot$) is the uniform lottery over a finite set.

**Proof.** For a partition $P$ with generic cell $J$, define the act

$$
f_{1,J} := \begin{cases} 
\ell^*(s) & \text{if } s \in J \\
\ell_*(s) & \text{if } s \notin J
\end{cases}
$$

and for each $P$ that is supported by $\omega$, define $x_1(P) := \{f_{1,J} : J \in P\}$.

Now, proceed inductively, and for $n \geq 2$, suppose we have the menu $x_{n-1}(P, \omega')$ for each $(P, \omega') \in \omega$, and define, for each cell $J \in P$, the act

$$
f_{n,J} := \begin{cases} 
\ell^*(s) & \text{if } s \in J \\
\ell_*(s) & \text{if } s \notin J
\end{cases}
$$

Note that Stokey, Lucas, and Prescott (1989) directly work with the optimal plan, but the essential idea is the same — continuation utilities arbitrarily far in the future must contribute arbitrarily little.

---

Note that Stokey, Lucas, and Prescott (1989) directly work with the optimal plan, but the essential idea is the same — continuation utilities arbitrarily far in the future must contribute arbitrarily little.
Then, given \((P, \omega') \in \omega\), we have the menu \(x_n(P, \omega') := \{f_n, J : J \in P\}\).

It is easy to see that for a fixed \((P, \omega') \in \omega\), the sequence of \(\text{racps} \ (x_n(P, \omega'))\) is a Cauchy sequence. Because \(X\) is complete, this sequence must converge to some \(x(P, \omega') \in X\). Moreover, this means that the sequence of sets \(\{x_n(Q, \tilde{\omega}) : (Q, \tilde{\omega}) \in \omega'_s\}\) also converges to \(\{x(Q, \tilde{\omega}) : (Q, \tilde{\omega}) \in \omega'_s\}\). This allows us to denote the uniform lottery over this finite set of points in \(X\) by \(\text{Unif}(\omega'_s)\).

Thus, \(x(P, \omega')\) consists of the acts \(\{f_J : J \in P\}\) where for each \(J \in P\)

\[
f_J := \begin{cases} 
(c_0, \text{Unif}(\{x(Q, \tilde{\omega}) : (Q, \tilde{\omega}) \in \omega'_s\})) & \text{if } s \in J \\
\ell^*_s(s) & \text{if } s \notin J
\end{cases}
\]
as claimed.

It is straightforward to verify that

\[
V(\ell^*_s, \omega, \pi_0) = V(x(P, \omega'), \omega, \pi_0) \geq V(x(P, \omega'), \tilde{\omega}, \pi_0)
\]
for all \(\tilde{\omega} \in \Omega\). Indeed, \(V(x(P, \omega'), \omega, \pi_0) = V(x(P, \omega'), (P, \omega'), \pi_0)\). We now prove that \(\omega\) and \(x(P, \omega')\) are strongly aligned. But first an intermediate lemma.

**Lemma B.2.** Let \(P, Q \in \mathcal{P}\) and suppose \(Q\) is not finer than \(P\). Then, for any \(\omega \in \Omega^s\), the menu \(x(P, \omega)\) defined in [★] is such that for all \(\omega' \in \Omega^s\), \(V(x, (P, \omega), \pi_0) > V(x, (Q, \omega'), \pi_0)\).

**Proof.** Fix \((P, \omega) \in \Omega\) and consider the menu \(x(P, \omega)\) defined in [★]. As noted above, for all \(\omega'\), we have \(V(x(P, \omega), (P, \omega), \pi_0) = V(x(P, \omega'), (P, \omega'), \pi_0)\). Moreover, it must be that for all \((Q, \omega')\) (even for \(Q = P\), we have \(V(x(P, \omega), (P, \omega), \pi_0) \geq V(x(P, \omega'), (Q, \omega'), \pi_0)\) and in the case where \(Q\) is not finer than \(P\) and \(Q \neq P\), \(V(x(P, \omega'), (P, \omega'), \pi_0) > V(x(P, \omega'), (Q, \omega'), \pi_0)\) by construction of the menu \(x(P, \omega')\). (This is straightforward to verify and is a version of Blackwell’s theorem on comparison of experiments; see Blackwell (1953) or Theorem 1 on p59 of Laffont (1989).) \(\Box\)

**Lemma B.3.** Suppose \(\omega'\) does not recursively Blackwell dominate \(\omega\). Then, for some \((P, \tilde{\omega}) \in \omega\), \(x(P, \tilde{\omega})\) defined in [★] is such that \(V(x(P, \tilde{\omega}), \omega, \pi_0) = V(x(P, \tilde{\omega}), (P, \tilde{\omega}), \pi_0) > V(x(P, \tilde{\omega}), \omega', \pi_0)\).

**Proof.** Suppose \(\omega'\) does not recursively Blackwell dominate \(\omega\). Then, there exists a smallest \(n \geq 1\) such that for all \(m < n\), proj\(_m\)(\(\omega'\)) recursively Blackwell dominates \(\text{proj}\_m(\omega)\), while \(\text{proj}\_n(\omega')\) does not recursively Blackwell dominate \(\text{proj}\_n(\omega)\).

From Corollary A.10 it follows that there exist finite sequences of partitions \((P_k)\) and \((P'_k)\), and states \((s_k)\) such that \(\Gamma^*(\tau^*(\omega', (P'_k), (s_k)))\) does not setwise Blackwell dominate the set \(\Gamma^*(\tau^*(\omega, (P_k), (s_k)))\), where \(\tau^*(\theta_0, (P_k), (s_k))\) represents the \(n\)-stage transition following the sequence of choices \((P_k)\) and states \((s_k)\), \(\omega_{n-k}^j = \tau^*(\omega_{n-k+1}^j, P_k, s_k)\) where \(P_k \in \Gamma^*(\omega_{n-k+1}^j)\), and \(\Gamma^*(\omega_{n-k+1}^j)\).

Let \((P_1, \tilde{\omega}) \in \omega\) be the unique first period choice under \(\omega\) that makes the sequence \((P_k)\) feasible. Then \(x(P_1, \tilde{\omega})\) defined in [★] is aligned with \((P_1, \tilde{\omega})\). That is, after \(n\) stages of choice and a certain path of states we can appeal to Lemma B.2, which completes the proof. \(\Box\)
Proof of Theorem 1. It follows from a straightforward extension of the arguments in Krishna and Sadowski (2014) (to the case of a compact prize space) that the collection \((u_s, \Pi, \delta)\) is unique in the sense of the Theorem. Now, define \(F_\omega := \{x(P, \tilde{\omega}) : (P, \tilde{\omega}) \in \omega\}\). It follows immediately from Lemma B.3 that \(F_\omega\) is uniformly strongly aligned with \(\omega\).

This brings us to the behavioral meaning of the recursive Blackwell order.

Corollary B.4. Let \(\omega, \omega' \in \Omega\). Then, the following are equivalent.

(a) \(\omega\) recursively Blackwell dominates \(\omega'\).

(b) For any \((u_s, \Pi, \delta)\) that induces \(\omega \mapsto V(\cdot, \omega, \cdot)\), we have \(V(x, \omega, \cdot) \geq V(x, \omega', \cdot)\) for all \(x \in X\).

Proof. That (a) implies (b) is easy to see. That (b) implies (a) is merely the contrapositive to Lemma B.3.

We are now in a position to prove Proposition 2.3.

Proof of Proposition 2.3. We first show the ‘only if’ part. On \(L\), we have \(\ell \succ^\dagger \ell'\) implies \(\ell \succ \ell'\). This implies, by Lemma 34 of Krishna and Sadowski (2014), that \(\succ^\dagger |_L = \succ |_L\). This, and the uniqueness of the \(\text{RAA}\) representation (Proposition 1.5) together imply that \((u_s, \delta, \Pi) = (u_s^\dagger, \delta^\dagger, \Pi^\dagger)\) after a suitable (and behaviorally irrelevant) normalization of the state-dependent utilities. Thus, part (b) of Corollary B.4 holds, which establishes the claim.

The ‘if’ part j8 follows immediately from Corollary B.4.

C. Existence

As always, \(C(C \times X)\) is the space of all uniformly continuous functions on \(C \times X\) and for \(\alpha \in \Delta(C \times X)\) and \(u \in C(C \times X)\), \(u(\alpha) := \int_{C \times X} u(c, x) \, d\alpha(c, x) =: \langle \alpha, u \rangle\). For each \(s \in S\), fix \(\ell^\dagger(s) \in \Delta(C \times X)\), and define \(\Omega_{s, \ell^\dagger(s)} := \{u_s \in C(C \times X) : u_s(\ell^\dagger(s)) = 0, \|u_s\|_\infty = 1\}\). Finally, define \(\Omega := \{(p_1, \ldots, p_n) : u_s \in \Omega_{s, \ell^\dagger(s)}, \ p_i \geq 0, \sum_i p_i = 1\}\). The space \(\Omega\) will serve as our subjective state space below. It is useful to reconsider \(\Omega\) as \(\Omega := \{(p, u) : p := (p_1, \ldots, p_n) \in \Delta(S)\}, u := (u_1, \ldots, u_n) \in \bigtimes_{s \in S} \Omega_{s, \ell^\dagger(s)}\)....

Throughout this section we assume that \(\succ\) is a binary relation on \(X\) and has a static representation \(V : X \to \mathbb{R}\) as follows:

\[ V(x) := \max_{\mu \in \mathcal{M}} \left[ \int_{\Omega} \max_{f \in \mathcal{F}} \sum_s p_s u_s(f(s)) \, d\mu(p, u) \right] \]

where the set \(\mathcal{M} \subset ba_n(\Omega)\) is weak* compact\(^{52}\) and \(\int_\Omega \max_{\alpha \in \mathcal{F}} \sum_i p_i \alpha_i \, d\mu(p, u)\) is independent of \(\mu\) for all \(\ell \in L\). Theorem 4 of Section E in the Supplementary Appendix shows that \(\succ\) satisfies Basic Properties (Axiom 1) and L-Independence (Axiom 2 (a)) if, and only if, it has a representation of the form \([\dagger]\).

For each \(\mu \in \mathcal{M}\), let \(V(x, \mu) := \int_{\Omega} \max_{f \in \mathcal{F}} \sum_s p_s u_s(f(s)) \, d\mu(p, u)\) be the utility from choosing the measure \(\mu\). Let \(Y : X \Rightarrow \mathcal{M}\) be the mapping selecting the maximizing \(\mu\) for each \(x\);\(^{52}\) Here, \(ba_n(\Omega)\) is the space of bounded additive (or finitely additive) measures (ie, charges) on \(\Omega\) that are also normal (ie, inner and outer regular).
that is, \( \Upsilon(x) := \arg \max_{\mu \in \mathcal{M}} V(x, \mu) \). It is easy to see that \( V(x, \mu) \) is continuous in \( \mu \), so it follows that \( \Upsilon \) is a correspondence that is closed valued. Notice that by definition, \( V \) is (i) convex, (ii) Lipschitz continuous, and (iii) \( L \)-affine in the sense that for all \( x \in X, \ell \in \mathcal{L} \) and \( t \in [0, 1] \), \( V((1-t)x + t\ell) = (1-t)V(x) + tV(\ell) \). We shall use these properties in the sequel.

Each of the following subsections will introduce a new axiom which will, in turn, impose further restrictions on the set \( \mathcal{M} \), eventually leading us to the desired representation theorem.

### C.1. Partitional Representation

In this section, we consider the representation in [\( \star \)] of \( \succeq \) and impose Indifference to Incentivized Contingent Commitment (henceforth IICC, Axiom 4). We begin with a definition.

**Definition C.1.** A menu \( x \) is nice if (i) it is finite, (ii) there is no \( f \in x \) and \( s \in S \) such that \( f(s) \in \Delta(C \times \mathcal{L}) \), and (iii) there is a unique \( \xi \in \mathcal{E}_x \) such that \( \mathcal{F}(\xi) \sim x \).

Let \( x \) be a nice menu, \( \xi \in \mathcal{E}_x \), and \( \mathcal{F}(\xi) := \{g_1, \ldots, g_m\} \), where \( \mathcal{E}_x \) is defined in Section 3.3. Each such \( \xi \) induces a partition \( J_1, \ldots, J_m \) of \( S \) wherein \( g_k(s) \notin \mathcal{L} \) if, and only if, \( s \in J_k \). In this case, we shall say that \( g_k \) is active in state \( s \in J_k \), so that \( J_k \) denotes all the states where \( g_k \) is active. By IICC (Axiom 4), there exists a \( \xi \in \mathcal{E}_x \) such that \( \mathcal{F}(\xi) \sim x \). It follows from the definition of \( \mathcal{E}_x \) that there exist \( f_1, \ldots, f_m \in x \) such that for each \( i = 1, \ldots, m \), \( f_i = \xi(s) \) and \( f_i \oplus_{(1,J_k)} \ell_\star = g_i \) for all \( s \in J_i \). The collection \( \{f_1, \ldots, f_m\} \) denotes a set of generators of the set \( x \). We shall also say that \( \{f_1, \ldots, f_m\} \) generates \( x \).

We show in Section J of the Supplementary Appendix that the space of nice menus is dense in \( X \), the space of all menus, and we clarify in Lemma J.1 the sense in which \( \{f_1, \ldots, f_m\} \) generates \( x \). That nice menus are dense in all menus only relies on Axioms 1, (parts of) 2, and 4.

The main consequence of assuming IICC (Axiom 4) is that instead of considering arbitrary finitely additive measures \( \mu \in \mathcal{M} \) over \( \mathcal{L} \) in the representation [\( \star \)], we can replace each \( \mu \) by a pair \((P, u)\) along with a prior belief \( \pi_0 \) over \( S \), where \( P \) is a partition of \( S \) and \( u \in \mathcal{C}(C \times \mathcal{L}) \).

**Proposition C.2.** Consider a preference relation \( \succeq \) on \( X \), and suppose \( V : X \rightarrow \mathbb{R} \) represents \( \succeq \) and has the form in [\( \star \)]. Then, the following are equivalent:

(a) \( \succeq \) satisfies IICC (Axiom 4).
(b) The function \( V \) has the form

\[
\text{[C.1]} \quad V(x) = \max_{(P, u) \in \mathcal{M}_p} \left[ \sum_{J \in P} \left( \max_{f \in x} \sum_{s \in J} \pi_0(s \mid J) u_s(f(s)) \right) \pi_0(J) \right]
\]

where \( \mathcal{M}_p \) is a collection of pairs \((P, u)\) where \( P \) is a partition and \( u = (u_s)_{s \in S} \) is a collection of state dependent (vN-M) utility functions on \( C \times \mathcal{L} \) with the property that for all \( s \in S, u_s(\alpha) = u'_s(\alpha) \) for all \((P, u), (P', u') \in \mathcal{M}_p \) and \( \alpha \in \Delta(C \times L) \).

Notice that each partition \( P \) along with a prior \( \pi_0 \) is equivalent to a posterior belief over \( S \), while \( u \) corresponds to a Dirac measure over \( \mathcal{L} \), both of which are countably additive. Thus, an essential part of the proof of Proposition C.2 is to show that IICC (Axiom 4) allows us to replace each
It is easy to see that for all \( \mu \in \mathcal{M} \) by a countably additive measure without affecting the representation. The proof is lengthy precisely due to the complications that arise from dealing with \( \mu \) in \([\dagger]\) that are finitely additive. If we knew beforehand that each \( \mu \) was countably additive, the proof would simply formalize the intuition behind IICC (Axiom 4) and be considerably shorter. The complete proof can be found in Section K of the Supplementary Appendix.

**C.2. Separable Representation**

We now investigate the implication of imposing State-Contingent Indifference to Correlation (henceforth SCIC, Axiom 3). Suppose \( V : X \to \mathbb{R} \) represents \( \succeq \) and takes the form \([C.1]\). For each \((P, u)\), define

\[
V(x, (P, u)) := \sum_{J \in P} \left( \max_{f \in x} \sum_{s \in J} \pi_0(s \mid J) u_s(f(s)) \right) \pi_0(J)
\]

to be the expected utility when the pair \((P, u)\) is chosen from \( \mathcal{M}_p \).

For each \( \alpha \in \Delta(C \times X) \), define the equivalence class \([\alpha] := \{ \alpha' \in \Delta(C \times X) : \alpha_1 = \alpha'_1, \alpha_2 = \alpha'_2 \}\) of lotteries with identical marginals over \( C \) and \( X \). Consider now the collection

\[
\mathcal{M}_p' := \left\{ (P, u') : (P, u) \in \mathcal{M}_p, u'_s(\alpha) = \min_{\alpha' \in [\alpha]} u_s(\alpha'), \text{ and } \alpha \in \Delta(C \times X) \right\}
\]

and observe that \( u'_s : \Delta(C \times X) \to \mathbb{R} \) is continuous and linear\(^{53} \) so that \( u'_s \in \mathcal{C}(C \times X) \). Moreover, for all \((P, u'), \tilde{P}, \tilde{u}') \in \mathcal{M}_p', u'_s|_{C \times L} = \tilde{u}'_s|_{C \times L} \). This implies that \( V(\tilde{P}, (P, u')) \) is independent of \((P, u') \in \mathcal{M}_p'\).

Now define \( V' : X \to \mathbb{R} \) as

\[
[C.2] V'(x) := \max_{(P, u') \in \mathcal{M}_p'} V(x, (P, u'))
\]

Observe that \( V' \) is monotone, ie, \( x \subset x' \) implies \( V'(x) \leq V'(x') \). This follows immediately from the form of \( V' \) in \([C.2]\). We claim that \( V' \) also represents \( \succeq \).

**Lemma C.3.** Let \( V \) and \( V' \) be defined as in \([C.1]\) and \([C.2]\) respectively. Then, for all \( x \in X \), \( V(x) = V'(x) \).

**Proof.** Because \( V \) is Lipschitz, it suffices to show that \( V(x) = V'(x) \) for all finite \( x \). Notice first that for all \( x \in X \), \( V'(x) \leq V(x) \). To see this, fix \( x \) and let \((P, u')\) be a maximizing pair for \( V' \). That is, \( V'(x) = V(x, (P, u')) \). But \( V(x, (P, u')) \leq V(x, (P, u)) \leq V(x) \), where the first inequality follows from the definition of \( u'_s \), which entails that for each \( \alpha \in \Delta(C \times X) \), \( u'_s(\alpha) \leq u_s(\alpha) \).

We shall now show that for all finite \( x \in X \), \( V(x) \leq V'(x) \). Note first that for each \( x \) and for any \((P, u)\) that is optimal for \( x \) with \( P = \{J_1, \ldots, J_m\} \), for \( i = 1, \ldots, m \) we can define the acts

\[
f_i := \arg \max_{f \in x} \sum_{s} \pi_0(s \mid J_i) u_s(f(s))
\]

(53) It is easy to see that for all \( \alpha' \in [\alpha] \) and \( \beta' \in [\beta] \), \( (\frac{1}{2} \alpha' + \frac{1}{2} \beta')_i = \frac{1}{2} \alpha_i + \frac{1}{2} \beta_i \) for \( i = 1, 2 \). This, the continuity of \( u'_s(\cdot; P) \), and the fact that \( u_s(\alpha'; P) \) is linear in \( \alpha' \), immediately imply that \( u'_s(\cdot; P) \) is linear.
Then, we see that $V(x) = V\{f_1, \ldots, f_m\}$, i.e., $\{f_1, \ldots, f_m\}$ is the generator set of $x$.

Now define the act $\hat{f}_i$ so that for each $s \in S$,

$$\hat{f}_i(s) = \arg\min_{\alpha \in [f_i(s)]} u_s(\alpha)$$

With this definition, we make the following observations.

(a) $V\{f_1, \ldots, f_m\} = V\{\hat{f}_1, \ldots, \hat{f}_m\}$ by repeated application of SCIC (Axiom 3).
(b) $V\{\hat{f}_1, \ldots, \hat{f}_m\}, (P, u) = V\{\hat{f}_1, \ldots, \hat{f}_m\}, (P, u')$ for all pairs $(P, u)$ and $(P, u')$. This follows from the definitions of $u'_i$ and $\hat{f}_i$, which imply that in any state $s$, $u_s(\hat{f}_i(s)) = u'_s(\hat{f}_i(s))$.
(c) $V\{\hat{f}_1, \ldots, \hat{f}_m\} = V\{\hat{f}_1, \ldots, \hat{f}_m\}, (\hat{P}, \hat{u})$ where $\hat{P}$ is a maximizing pair in $M_p$ for $\{\hat{f}_1, \ldots, \hat{f}_m\}$ under $V$.
(d) $V\{\hat{f}_1, \ldots, \hat{f}_m\}, (\hat{P}, \hat{u}') = V\{\hat{f}_1, \ldots, \hat{f}_m\}, (\hat{P}, \hat{u}')$. This follows from the definitions of $\hat{u}'$ and $\hat{f}_i$, which imply that in any state $s$, $\hat{u}'_s(\hat{f}_i(s)) = \hat{u}'_s(\hat{f}_i(s))$.

We can now use these equalities to form the following chain.

$$V(x) = V\{f_1, \ldots, f_m\}$$
$$= V\{\hat{f}_1, \ldots, \hat{f}_m\}$$
$$= V\{\hat{f}_1, \ldots, \hat{f}_m\}, (\hat{P}, \hat{u})$$
$$= V\{\hat{f}_1, \ldots, \hat{f}_m\}, (\hat{P}, \hat{u}')$$
$$\leq V'\{f_1, \ldots, f_m\}$$
$$\leq V'(x)$$

which completes the proof.

We can now state the main result of this section.

**Proposition C.4.** Let $V$ be as in [C.1] and suppose $V$ represents $\succeq$. Then, the following are equivalent.

(a) $\succeq$ satisfies SCIC (Axiom 3).
(b) There exist functions $u_s \in C(C)$ and a set $M'_p$ consisting of pairs of $(P, v_s)$ where $P$ is a partition and $v_s \in C(X)$ for each $s$ such that $(P, v_s), (P', v'_s) \in M'_p$ implies $v_s|_L = v'_s|_L$ for all $s \in S$, and $V$ can be written as

$$[C.3] \quad V(x) = \max_{(P, v_s) \in M'_p} \sum_{J \in P} \pi_0(J) \max_{J \in X} \sum_{s} \pi_0(s | J) \left[ u_s(f_1(s)) + v_s(f_2(s)) \right]$$

**Proof.** It is easy to see that (b) implies (a). We now show that (a) implies (b).

Lemma C.3 implies we can replace $V$ in [C.1] by $V'$ in [C.2]. Moreover, from the definition of $V$ in [C.1], $u_s(\alpha) = u'_s(\alpha)$ for all $(P, u), (P', u') \in M'_p$ and for all $\alpha \in \Delta(C \times L)$.

For any $\alpha \in \Delta(C \times X)$ with marginals $\alpha_1$ and $\alpha_2$, let $\alpha_1 \otimes \alpha_2 \in \Delta(C) \times \Delta(X)$ denote the product lottery with the same marginals. Recall that $\ell^\dagger \in L$ is such that $u_s(\ell^\dagger(s)) = 0$ for all $s$. Given $(P, u)$, now define

- $u_s(\alpha_1) := u_s(\alpha_1 \otimes \ell^\dagger_2(s))$ (and notice $u_s(\alpha) = u'_s(\alpha)$ for all $(P, u), (P', u') \in M_p$ and for all $\alpha \in \Delta(C \times L)$ because $\alpha_1 \otimes \ell^\dagger_2(s) \in \Delta(C \times L)$; and
With these definitions, \( u_s \in \mathcal{C}(C) \) while \( v_s(\cdot) \in \mathcal{C}(X) \). Notice that the lotteries \( \frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^s(s) \) and \( \frac{1}{2}(\alpha_1 \otimes \ell^s_2(s)) + \frac{1}{2}(\ell^s_1(s) \otimes \alpha_2) \) have identical marginals, which implies that for every \((P, u)\),

\[
    u_s \left( \frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^s(s) \right) = u_s \left( \frac{1}{2}(\alpha_1 \otimes \ell^s_2(s)) + \frac{1}{2}(\ell^s_1(s) \otimes \alpha_2) \right)
\]

This means we can write

\[
    \frac{1}{2}u_s(\alpha_1 \otimes \alpha_2) + \frac{1}{2}u_s(\ell^s(s)) = u_s \left( \frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^s(s) \right) = u_s \left( \frac{1}{2}(\alpha_1 \otimes \ell^s_2(s)) + \frac{1}{2}(\ell^s_1(s) \otimes \alpha_2) \right)
\]

\[
    = \frac{1}{2}u_s(\alpha_1 \otimes \ell^s_2(s)) + \frac{1}{2}u_s(\ell^s_1(s) \otimes \alpha_2) = \frac{1}{2}u_s(\alpha_1) + \frac{1}{2}v_s(\alpha_2)
\]

where the second equality holds because \( u_s(\cdot) \) is constant on the equivalence class of lotteries with identical marginals. The first and third equalities from the linearity of \( u_s(\cdot) \), while the last equality follows from the definitions of \( u_s \) and \( v_s(\cdot) \).

But we have already stipulated that \( u_s(\ell^s(s)) = 0 \), which implies that for all \( s \), we have

\[
    u_s(\alpha_1 \otimes \alpha_2) = u_s(\alpha_1) + v_s(\alpha_2)
\]

Substituting in [C.2] and invoking Lemma C.3 gives us [C.3], as desired.

As always, for each \((P, (v_s)) \in \mathcal{M}_p^\prime\prime\), define \( V(x, (P, (v_s))) \) as

\[
    V(x, (P, (v_s))) = \sum_{J \in P} \pi_0(J) \max_{f \in X} \sum_s \pi_0(s \mid J) \left[ u_s(f_1(s)) + v_s(f_2(s)) \right]
\]

### C.3. Representation with Deterministic Continuation Utilities

Thus far, we have seen that \( \succsim \) has a representation as in [C.3]. We now impose Concordant Independence (Axiom 5) and show that \( \succsim \) then has a representation of the form

\[
    [C.4] \quad V(x) = \max_{P \in \mathcal{M}_p^\prime} \sum_{f \in P} \left[ \max_{f \in X} \sum_s \pi_0(s \mid J) \left[ u_s(f_1(s)) + v_s(f_2(s), P) \right] \pi_0(J) \right]
\]

where \( \mathcal{M}_p^\prime \) is a finite collection of partitions \( P \) of \( S, u_s \in \mathcal{C}(C) \), and \( v_s(\cdot, P) \in \mathcal{C}(X) \) for each \( s \in S \) and \( P \in \mathcal{M}_p^\prime \), with the property that for all \( P, P' \in \mathcal{M}_p^\prime, s \in S, v_s(\cdot, P) \mid_L = v_s(\cdot, P') \mid_L \). As always, for any partition \( P \in \mathcal{M}_p^\prime \), we define

\[
    V(x, P) = \sum_{f \in P} \max_{f \in X} \left[ \sum_s \pi_0(s \mid J) \left[ u_s(f_1(s)) + v_s(f_2(s), P) \right] \pi_0(J) \right]
\]

For a fixed \( P \), let \( X'_P \) and \( \hat{X}_P \) be defined as follows:

\[
    X'_P := \{ x : V(x) = V(x, (P, (v_s))) \text{ for some } (P, (v_s)) \in \mathcal{M}_p^\prime \text{ and } V(x) > V(x, (Q, (v'_s))) \text{ for all } (Q, (v'_s)) \in \mathcal{M}_p^\prime \text{ such that } P \neq Q \}\]

\[
    \hat{X}_P := \{ x : V(x) = V(x, (P, (v_s))) \text{ for some } (P, (v_s)) \in \mathcal{M}_p^\prime \text{ and } V(x) \geq V(x, (Q, (v'_s))) \text{ for all } (Q, (v'_s)) \in \mathcal{M}_p^\prime \text{ such that } P \neq Q \}\]

46
Recall that \( x_1(P) := x(P, \hat{\omega}) \) as in \([\bullet]\) in Section 2.5. That is, for any partition \( P \), \( x_1(P) \in X \) is a one-period problem where the choice of \( P \) is optimal.

**Lemma C.5.** Let \( x \in X'_p \). Then, for all \( \varepsilon \in (0, 1), (1 - \varepsilon)x + \varepsilon x_1(P) \in X'_p \). Moreover, \( V((1 - \varepsilon)x + \varepsilon x_1(P)) = V((1 - \varepsilon)x + \varepsilon \ell^*) > V((1 - \varepsilon)x + \varepsilon x_1(Q)) \) if \( P \) is not finer than \( Q \).

**Proof.** We begin by establishing three claims.

(i) In the representation \([C.3]\), \( v_s(\ell^*) > v_s(\ell_*) \) for all \( s \in S \).

(ii) \( V(x_1(Q)) \leq V(\ell^*) \) for all \( Q \in \mathcal{P} \).

(iii) \( V(x_1(Q), (P, (v_s))) = V(\ell^*) \) if, and only if, \( P \) is finer than \( Q \).

To see (i), observe that by the \( \text{raa} \) representation of Appendix A.7, \([\ell^* \oplus (1, \mathcal{S} \setminus \ell_*) \ell_*] = \ell_* \) for all \( s \in S \). Because we have \( v_s(\ell) = v'_s(\ell) \) for all \( \ell \in L \) and \((P, (v_s)), (P', (v'_s)) \in \mathcal{M}'_p \) in \([C.3]\), \( v_s(\ell^*) = v_s(\ell_*) \) for all \( s \in S \).

Given claim (i), and that \( u_s(\ell^+) \geq u_s(\ell^-) \) for all \( s \) by construction, claim (ii) follows by evaluating \( V \) in \([C.3]\) at \( x_1(Q) \).

To establish claim (iii), consider first \( P \) finer than \( Q \), then

\[
V(x_1(Q), (P, (v_s))) = \sum_{J \in P} \pi_0(J) \max_{J' \in x_1(Q)} \sum_s \pi_0(s | J) [u_s(f_1(s)) + v_s(f_2(s))]
\]

\[
= \sum_{J \in P} \pi_0(J) \sum_s \pi_0(s | J) [u_s(\ell^+) + v_s(\ell^*)]
\]

\[
= V(\ell^*, (P, (v_s))) = V(\ell^*)
\]

Now suppose instead that \( P \) is not finer than \( Q \). Then there must be \( J \in P \) with \( s \in J \) such that

\[
\left[ \arg \max_{J' \in x_1(Q)} \left( \sum_s \pi_0(s' | J) [u_{s'}(f_1(s')) + v_{s'}(f_2(s'))] \right) \right](s) = \ell_*(s)
\]

Then, by claim (i) and because \( u_s(\ell^+) \geq u_s(\ell^-) \) for all \( s \) by construction, we find that \( V(\ell^*) > V(x_1(Q), (P, (v_s))) \).

With the claims in hand, observe that \( V((1 - \varepsilon)x + \varepsilon x_1(P)) \geq V((1 - \varepsilon)x + \varepsilon x_1(P), (P', \cdot)) \) for all \((P', \cdot) \in \mathcal{M}'_p \). Let \( (v_s) \) be such that \((P, (v_s)) \in \mathcal{M}'_p \) and \( V(x) = V(x, (P, (v_s))) \). Then

\[
V((1 - \varepsilon)x + \varepsilon x_1(P), (P, (v_s))) = (1 - \varepsilon) V(x) + \varepsilon V(x_1(P), (P, (v_s)))
\]

\[
= (1 - \varepsilon) V(x) + \varepsilon V(\ell^*) = V((1 - \varepsilon)x + \varepsilon \ell^*)
\]

by claims (ii) and (iii). Moreover, for any other \((Q, (v'_s)) \in \mathcal{M}'_p \),

\[
V((1 - \varepsilon)x + \varepsilon x_1(P), (Q, (v'_s))) = (1 - \varepsilon) V(x, (Q, (v'_s))) + \varepsilon V(x_1(P), (Q, (v'_s)))
\]

\[
< (1 - \varepsilon) V(x) + \varepsilon V(\ell^*)
\]

where the strict inequality is because \( V(x, (Q, (v'_s))) < V(x) = V(x, (P, \cdot)) \) (recall that \( x \in X'_p \)) and \( V(x_1(P), (Q, (v'_s))) \leq V(\ell^*) \) (claim (ii) above). This implies \((1 - \varepsilon)x + \varepsilon x_1(P) \in X'_p \). Moreover, it now follows immediately that \( V((1 - \varepsilon)x + \varepsilon x_1(P)) = V((1 - \varepsilon)x + \varepsilon \ell^*) \).

Finally, suppose \( P \) is not finer than \( Q \). Consider the menu \((1 - \varepsilon)x + \varepsilon x_1(Q) \) and suppose \((P', \cdot) \in \mathcal{M}'_p \) is optimal for this menu. Notice that if \( P' \neq P \), then \( V(x, (P', \cdot)) < V(x, (P, \cdot)) = V(x) \)
by virtue of \( x \in X'_p \), and that if \( P = P' \), then \( V(x_1(Q), (P, \cdot)) < V(\ell^*) \) by case (iii) because \( P \) is not finer than \( Q \). Thus,

\[
V((1-\varepsilon)x + \varepsilon x_1(Q)) = V((1-\varepsilon)x + \varepsilon x_1(Q), (P', \cdot))
\]

\[
= (1-\varepsilon)V(x, (P', \cdot)) + \varepsilon V(x_1(Q), (P', \cdot))
\]

\[
< (1-\varepsilon)V(x, (P, \cdot)) + \varepsilon V(\ell^*) = V((1-\varepsilon)x + \varepsilon \ell^*)
\]

which completes the proof. \( \square \)

**Lemma C.6.** \( X'_p \) is convex and consists of concordant rACP.

**Proof.** Let \( x, y \in X'_p \) such that \( x \succeq y \). It follows from IICC (Axiom 4) that \( x \succeq y \succeq \ell_* \). By Continuity of \( \succeq \) (Axiom 1 (b)), there exists a \( t \in [0, 1] \) such that \((1-t)x + t\ell_* \simeq y \). Because \( V \) is \( L \)-affine, any \((P, (v_{i}))\) that is optimal for \( x \) is also optimal for \((1-t)x + t\ell_* \). Moreover, \((1-t)x + t\ell_* \in X'_p \) because every optimizer for \( x \) is also an optimizer for \((1-t)x + t\ell_* \) and vice versa (this follows immediately from the \( L \)-affinity of \( V \)). Thus, it is without loss of generality to consider \( x, y \in X'_p \) such that \( x \sim y \).

By Lemma C.5, \((P, (\tilde{v}_i)) \in \mathcal{M}''_p \) remains optimal for both \((1-\varepsilon)x + \varepsilon x_1(P) \) and \((1-\varepsilon)y + \varepsilon x_1(P) \), for all \( \varepsilon \in (0, 1) \). It follows that \( x \) and \( y \) are \( \varepsilon \)-concordant (Definition 3.1), and by Concordant Independence (Axiom 5), so are \( x \) and \( \frac{1}{2}x + \frac{1}{2}y \). It follows that \( x, y, \) and \( \frac{1}{2}x + \frac{1}{2}y \) are concordant.

Now suppose \((Q, (v'_i)) \in \mathcal{M}''_p \) is optimal for \( \frac{1}{2}x + \frac{1}{2}y \). Then, \( V(x, (Q, (v'_i))) \leq V(x) \) and \( V(y, (Q, (v'_i))) \leq V(y) \). Because \( x, y, \) and \( \frac{1}{2}x + \frac{1}{2}y \) are \( \varepsilon \)-concordant, \( V(x) = V(y) = V(\frac{1}{2}x + \frac{1}{2}y) \), i.e., \((Q, (v'_i)) \) is optimal at \( x \) and \( y \). But \( x, y \in X'_p \), which implies that \( Q = P \). That is, \( \frac{1}{2}x + \frac{1}{2}y \in X'_p \).

Standard arguments now imply that every \( z \in [x, y] \) is concordant with \( x \) and \( y \) and the argument above establishes \((Q, (v'_i)) \) as a maximizer at each such \( z \), i.e., \( X'_p \) is convex. \( \square \)

**Lemma C.7.** For each \( x \in X \), there exists \((P, (v_i)) \in \mathcal{M}''_p \) such that \( x \in \text{cl}(X'_p) \).

**Proof.** Let \( x \in \tilde{X}_{P_1} \cap \cdots \cap \tilde{X}_{P_n} \) and suppose \( n \geq 2 \) (because if \( n = 1 \), then \( x \in X'_p \subset \text{cl}(X'_p) \)). Without loss of generality, suppose that none of \( P_2, \ldots, P_n \) are finer than \( P_1 \). In analogy to the arguments in the proof of Lemma C.5, we find that \( V((1-\varepsilon)x + \varepsilon x_1(P_1), (P_1, (v^1_i))) = V((1-\varepsilon)x + \varepsilon \ell^*) > V((1-\varepsilon)x + \varepsilon x_1(P_1), (P_1, (v^1_i))) \) for some \( v^1_i \) with \((P, (v^1_i)) \in \mathcal{M}''_p \) and all \((P_i, (v^1_i)) \in \mathcal{M}''_p \) for \( i = 2, \ldots, n \). That is, \((1-\varepsilon)x + \varepsilon x_1(P_1) \in X'_p \) for all \( \varepsilon \in (0, 1) \), which implies \( x \in \text{cl}(X'_p) \) as claimed. \( \square \)

**Lemma C.8.** Let \( x \in X'_p \) and let \( Y_x \) denote the set of rACP that (i) are concordant with \( x \), and (ii) have a unique optimal partition. Then, \( Y_x = X'_p \).

**Proof.** By hypothesis, \( P \) is uniquely optimal for \( x \). Let \( Q \neq P \) be optimal for \( y \in Y_x \). Because \( V \) is \( L \)-affine, we may assume without loss of generality, that \( x \sim y \). (This is made clear in the proof of Lemma C.6.) If \( P \) is not finer than \( Q \), by Lemma C.5, \((1-\varepsilon)y + \varepsilon x_1(Q) > (1-\varepsilon)x + \varepsilon x_1(Q) \), which contradicts our assumption that \( x \) and \( y \) are concordant. Conversely, if \( Q \) is not finer than \( P \), then an analogous argument establishes that \((1-\varepsilon)x + \varepsilon x_1(P) > (1-\varepsilon)y + \varepsilon x_1(P) \), which also contradicts our assumption that \( x \) and \( y \) are concordant. Therefore, \( P \) must be the unique optimal partition for any \( y \in Y_x \). Thus, \( Y_x \subset X'_p \). That \( X'_p \subset Y_x \) is an immediate consequence of Lemma C.6. \( \square \)
Notice that replacing $M_p'$ with its weak* closure (in the event that it is not weak* compact) in [C.3] does not affect the representation. Therefore, we shall now assume that $M_p'$ is weak*-compact.

**Lemma C.9.** Let $x \in \text{cl}(X'_p)$. Then, there exists $(v_x)$ such that $(P, (v_x)) \in M_p'$ is optimal for all $y \in \text{cl}(X'_p)$.

**Proof.** By Lemma C.8, $Y_x \subset X'_p$, which, by Lemma C.6, is convex. By Concordant Independence, $\succeq |X'_p$ satisfies Independence. That is, $V|X'_p$ is linear. It follows from Lemma F.5 in the Supplementary Appendix that there exists $(v_x)$ such that $(P, (v_x))$ is optimal for all $x \in X'_p$. Continuity now implies that $(P, (v_x))$ is optimal for all $x \in \text{cl}(X'_p)$. 

It follows that we can replace the set $M_p'$ by a finite collection $\{ (P_1, (v_1^1)), \ldots, (P_n, (v_n^n)) \} = M_p^\#$ as in [C.4]. Thus, we have shown that (a) implies (b) in the following proposition. That (b) implies (a) is clear.

**Corollary C.10.** Let $V$ be as in [C.3] and suppose $V$ represents $\succeq$. Then, the following are equivalent.
(a) $\succeq$ satisfies Concordant Independence (Axiom 5).
(b) $V$ can be written as in [C.4].

---

**C.4. Self-Generating Representation**

Recall that a representation $V : X \rightarrow \mathbb{R}$ of $\succsim$ is a self-generating representation if $V \in \Phi^*$ (see section A.8 for the definition of $\Phi^*$). In this section, we show that imposing Self-Generation (Axiom 6) on $\succeq$ implies it has a self-generating representation.

**Proposition C.11.** Let $\succeq$ be a binary relation on $X$. Then, the following are equivalent.
(a) $\succeq$ satisfies Axioms 1–6.
(b) $\succeq$ has a self-generating representation, that is, there exists a function $V \in \Phi^*$ that represents $\succsim$.

The proof is in Appendix C.4.2. We first show that $\succeq_{(x,y)}$ from Definition 3.2 is well defined. We begin with a preliminary lemma.

**Lemma C.12.** Let $x = \{ f_1, \ldots, f_m \}$, and $x' = \{ f'_1, f_2, \ldots, f_m \}$. Suppose $d(f_1, f'_1) < \varepsilon$. Then, $d(x, x') < \varepsilon$.

**Proof.** Notice that

$$d(f, x') = \begin{cases} 0 & \text{if } f \in \{ f_2, \ldots, f_m \} \\ \min\{d(f_1, f'_1), \min_i d(f_1, f_i)\} & \text{if } f = f_1 \end{cases}$$

Therefore, $d(f_1, x') \leq d(f_1, f'_1)$. Similarly, $d(f'_1, x) \leq d(f_1, f'_1)$. This implies that $d(x, x') = \max \left[ d(f_1, x'), d(f'_1, x) \right] \leq d(f_1, f'_1) < \varepsilon$, as claimed. 

Notice that $M_p^\#$ in [C.4] is finite and can be taken to be minimal (in the sense that if $N_p^\#$ is another set that represents $V$ as in [C.4], then $M_p^\# \subset N_p^\#$) without affecting the representation.
Lemma C.13. Let $\succsim$ have a representation as in [C.4]. For all $P \in M^p$, there exists a nice $x \in X'_p$. Moreover, for any nice $x$, there is $\xi \in \Xi_x$ such that $f = \xi(s)$ is the unique optimal choice of act from $x$ when the agent learns the cell that contains $s$.

Proof. The finiteness and minimality of $M^p$ in [C.4] implies that for any $P \in M^p$, there exists an open set $O \subset X'_p$. Because the space $X_0$ of nice menus is dense in $X$ (see Proposition I.5 in the Supplementary Appendix), there exists $x \in X_0 \cap O$.

Let $x$ be nice and $V(x) = V(x, P)$. Suppose, instead, that both $f$ and $f'$ are optimal, given the cell $J \in P$. Then, for all $s' \in J$, we may take $\xi(s')$ to be either $f$ or $f'$ while ensuring $x \succsim J(\xi)$, which contradicts the assumption that $x$ is nice.

Lemma C.14. In the representation [C.4], for all $P \in M^p$, $v_x(y, P) \geq v_x(\ell_\ast, P)$.

Proof. Suppose instead that $v_x(y, P) < v_x(\ell_\ast, P)$. Consider $x' \in X'_p$ nice, which exists by Lemma C.13. Again by Lemma C.13, we can consider $x = \xi(s)$ for the unique $\xi \in \Xi_x$ with $x' \sim J(\xi)$. Then, for $\epsilon > 0$ small enough such that $P$ and (the perturbation of) $x'$ remain optimal, $[x \succsim (f, \epsilon, s)] > [x \succsim (f, \epsilon, s)]$, contradicting part (a) of IICC (Axiom 4), which by Continuity of $\succsim$ implies that $[x \succsim (f, \epsilon, s)] > [x \succsim (f, \epsilon, s) \ast]$ for all $y \in X$.

Lemma C.15. Let $\succsim$ have a representation as in [C.4]. For a nice $x \in X'_p$ for some $P \in M^p$ and $s \in S$, if $x \succsim (x, s)$ $y'$, then this ranking is independent of the choice of $f \in x \in \Xi_x$ and $\epsilon \in (0, 1)$ for which Definition 3.2 applies. In particular, $\succsim (x, s)$ is represented by $v_x(\cdot, P)$ on any $X' \subset X$ where it is complete. Finally, if $x$ is nice, has a unique optimal partition, and is concordant with $x$, and if $\succsim (x, s)$ and $\succsim (x', s)$ both rank $y, y' \in X$, they must rank them the same.

Proof. Let $x \in X'_p$ be nice, so that $V(x) = V(x, P)$. Fix $s \in S$. Given the continuity of $V$ in [C.4], there is $\epsilon > 0$ such that for all $x' \in B(x; \epsilon)$, $P$ is the unique optimal partition, and hence all $x', x'' \in B(x; \epsilon)$ are concordant with each other. By Lemma C.12, $[x \succsim (f, \epsilon, s)]$, $[x \succsim (f, \epsilon, s)]$ $\in B(x; \epsilon)$. Then, $[x \succsim (f, \epsilon, s)]$, $[x \succsim (f, \epsilon, s)]$ are possible if, and only if, there is an act $f$ such that (i) $V(x \succsim (f, \epsilon, s), y) > V(x)$ (which implies $f \sim (\epsilon, s) y$ is optimal in $s$ from $x \succsim (f, \epsilon, s), y$, and (ii) $V(x \succsim (f, \epsilon, s), y) \geq V(x \succsim (f, \epsilon, s), y')$. Property (ii), in turn, holds if, and only if, $\epsilon = \binom{u_x(f_1(s)) + v_x(f_2(s), P)}{1 - \epsilon} [u_x(f_1(s)) + v_x(f_2(s), P)] + \epsilon [u_x(f_1(s)) + v_x(f_2(s), P)]$.

Lemma C.16. Let $\succsim$ have a representation as in [C.4]. Then for any $P \in M^p$, there is a nice sequence $(x_n)$ in $X'_p$ and a sequence $(Y_n)$ in $X$, such that (i) $\succsim (x_n, s)$ is complete on $Y_n$ and if $v_x(y, P) > v_x(y', P)$, then there is $N$ with $y \succsim (x_n, s) y'$ for all $n > N$, (ii) $\succsim (x_n, s) = \succsim (x_m, s)$ on $Y_n \cap Y_m$ for all $n, m > 0$, and (iii) $\bigcup_n Y_n$ is dense in $X$. 

50
We now relate preferences on $\succsim$ in Claim. (ii) from the product metric on $x$, and note that for all such $x$ the beginning the second period. Now, inductively define $x_{n+1} \succ x_n$ by $x_{n+1} = x_n + \epsilon_n$, where $\epsilon_n$ is a positive number such that $x_{n+1} = x_n + \epsilon_n + \epsilon_{n+1}$ for all $n$. By Lemma C.16, the binary relation $\succsim$ is dense in $x$, proving (iii). That $\succsim(x_n, s)$ is represented by $\succsim(x_{n+1}, s)$ on $x_{n+1}$, which completes the proof.

The claim implies that there is a sequence of sets $Y_n$ such that $\succsim(x_n, s)$ is complete on $Y_n$ and $Y := \bigcup_n Y_n$ is dense in $X$. For all $n > N$, there are $n > N$ such $x_n \notin \bigcup_{n \leq N} Y_n$. This implies that $x_n \notin \bigcup_{n \leq N} Y_n$ for all $n > N$. By the arguments in the proof of Lemma C.15, this completes the proof.

**Lemma C.17.** The binary relation $\succsim_{P,s}$ on $X$ which is represented by $\succsim_{P,s}(\cdot)_P$ satisfies Axioms 1–6.

**Proof.** By Lemma C.16, the binary relation $\succsim_{P,s}(\cdot)_P := \bigcup_n \succsim(x_n, s)$ is complete on $X$ and is represented by $\succsim_{P,s}(\cdot)_P$. By Self-Generation (Axiom 6) and because $\succsim(x_n, s) \succsim(x_{n+1}, s)$ on $Y_n \cap Y_{n+1}$ for all $n, m > 0$, $\succsim_{P,s}$ satisfies Axioms 1–6 on $X$. Because $Y_n$ is dense in $X$, there is a unique continuous extension of $\succsim_{P,s}$ to $X$, which satisfies Axioms 1–6 on $X$, which we denote by $\succsim_{P,s}$, and which is represented by $\succsim_{P,s}(\cdot)_P$ because $\succsim_{P,s}(\cdot)_P$ is itself uniformly continuous on $X$.

Before we prove Proposition C.11, an interlude.

### C.4.1. Some Properties of Consumption Streams

We now relate preferences on $L$ to those on $X$.

Let $\bar{X}_1 := \mathcal{K}(\mathcal{F}(\Delta(C \times \{\ell_*\})))$ be the space of one-period problems that always give $\ell_*$ from the beginning the second period. Now, inductively define $\bar{X}_{n+1} := \mathcal{K}(\mathcal{F}(\Delta(C \times \bar{X}_n)))$ for all $n \geq 1$, and note that for all such $n$, $\bar{X}_n \subset X$. Finally, let $\bar{X} := \bigcup_n \bar{X}_n$.

**Lemma C.18.** The set $\bar{X} \subset X$ is dense in $X$.

**Proof.** Recall that $X$ is the space of all consistent sequences in $X^* := \times_{n=1}^\infty X_n$, where $X_1 := \mathcal{K}(\mathcal{F}(\Delta(C)))$ and $X_{n+1} := \mathcal{K}(\mathcal{F}(\Delta(C \times X_n)))$. Clearly, every $x \in X$ is a sequence of the form $x = (x_1, x_2, \ldots, x_n, \ldots)$, and the metric on $X$ is the product metric.

For any $x = (x_1, x_2, \ldots) \in X$ and $n \geq 1$, let $\bar{X}_n \subset X$ be the set of all $x_n$ concatenated with $\ell_*$. It follows from the product metric on $X$ — see Appendix A.2 — that for any $\epsilon > 0$, there exists $n \geq 1$ such that $d(x, \bar{X}_n) < \epsilon$, as claimed.
Lemma C.19. Let $\succeq$ satisfy Axioms 1–5. Then, for any $s \in S$ and $P \in M^\#_P$, $\ell^* \succeq_{(P,s)} \ell \succeq_{(P,s)} \ell_*$ for all $\ell \in L$.

Proof. The preference $\succeq$ has a separable and partitional representation as in [C.4]. Therefore, $\succeq_s$ on $L$ is represented by $u_s(\cdot) + v_s(\cdot, Q)$ for all $Q$. Moreover, $\succeq_{\mid L}$ has an $\text{AAA}$ representation. As observed in Section A.7, $\succeq_s$ on $L$ is separable and has the property that for all $c \in C$, $\ell \in L$ and $s \in S$, $(c, \ell^*) \succeq_s (c, \ell)$ $\succeq_s (c, \ell_*)$. This implies that for all $\ell \in L$, $v_s(\ell^*, Q) \geq v_s(\ell, Q) \geq v_s(\ell_*, Q)$ for all partitions $Q \in M^\#_P$. By Lemma C.15, we may take $\ell^* \succeq_{(P,s)} \ell \succeq_{(P,s)} \ell_*$ for all $\ell \in L$, $s \in S$, and nice $x$. \hfill $\square$

Proposition C.20. Let $\succeq$ satisfy Axioms 1–6. Then, for all $x \in X$, $\ell^* \succeq x$.

Proof. By the continuity of $\succeq$ and by Lemma C.18, it suffices to show that for all $\tilde{x} \in \tilde{X}_n$. Suppose $\tilde{x} \in \tilde{X}_n$. We first consider the case $n = 1$. It follows immediately from the representation in [C.4] that $V(\tilde{x}_1) \leq V(\ell^*)$ for all $\tilde{x}_1 \in \tilde{X}_1$. Notice that this representation in [C.4] is equivalent to $\succeq$ satisfying Axioms 1–5. But $\succeq$ satisfies Axiom 6, so that $\succeq_{(P,s)}$ also satisfies Axioms 1–6 for any $P \in M^\#_P$, which implies that there exists $\ell^*_{(P,s)}$ such that $v_s(\ell^*_{(P,s)}) \geq v_s(\tilde{x}_1, 1)$ for all $\tilde{x}_1 \in \tilde{X}_1$. By Lemma C.19, we may take $\ell^* = \ell^*$, so that $v_s(\ell^*, P) \geq v_s(\tilde{x}_1, 1)$ for all $\tilde{x}_1 \in \tilde{X}_1$.

Now consider the induction hypothesis: If $\succeq$ satisfies Axioms 1–6, then for all $\tilde{x}_n \in \tilde{X}_n$, $\ell^* \succeq \tilde{x}_n$. Suppose the induction hypothesis is true for some $n \geq 1$. We shall now show that it is also true for $n + 1$.

Because $\succeq_{(P,s)}$ also satisfies Axioms 1–6 on $X$, we must also have $v_s(\ell^*, P) \geq v_s(\tilde{x}_n, P)$ for all $\tilde{x}_n \in \tilde{X}_n$ (where we have appealed to Lemma C.19 to establish that $\ell^*$ is the $v_s(\cdot, P)$-best consumption stream). In particular, this implies that for any lottery $\alpha_2 \in \Delta(\tilde{X}_n)$, $v_s(\ell^*, P) \geq v_s(\alpha_2, P)$.

Now consider any $\tilde{x}_{n+1} \in \tilde{X}_{n+1}$. We have, for any choice of $P$,

$$V(\tilde{x}_{n+1}, P) = \max_{f \in \tilde{x}_{n+1}} \sum_{J \in P} \pi_0(s \mid J)[u_s(f_1(s)) + v_s(f_2, P)]$$

$$\leq \sum_{J \in P} \pi_0(s \mid J)[u_s(c^+_s) + v_s(\ell^*, P)]$$

$$= V(\ell^*, P) = V(\ell^*)$$

where we have used the facts that $f_1(s) \in \Delta(C)$ and $f_2(s) \in \Delta(\tilde{X}_n)$, and that $u_s(c^+_s)$ and $v_s(\ell^*, P)$ respectively dominate all such lotteries, as established above. Thus, for all $\tilde{x}_{n+1} \in \tilde{X}_{n+1}$, $\ell^* \succeq \tilde{x}_{n+1}$, which completes the proof. \hfill $\square$

C.4.2. Proof of Proposition C.11

Proof. To see that (b) implies (a), suppose $\succeq$ has the representation [C.4]. By Proposition C.10 $\succeq$ satisfies Axioms 1–5. All that remains to establish is that $\succeq$ also satisfies Axiom 6.

Given a representation as in [C.4] that is self-generating, let $x \in X$ be finite and $P \in M^\#_P$ be an optimal partition for $x$. Observe first that if $x \in X'_p$ is nice, and if $\succeq_{(x,s)}$ is complete on all of $X'$, then by Lemma C.15 it is represented by $v_s(\cdot, P)$. Because the representation is self-generating, $\succeq_{(x,s)}$ must satisfy Axioms 1–5 on $X'$.
In general, $P$ may not be uniquely optimal for $x$, and $x$ may not be nice. By definition, $\succsim_{(x,s)}$ is complete on $X' \subset X$ only if for all $y, y' \in X'$ there is $\varepsilon > 0$ such that $[x \succ_{(f,\varepsilon,s)} y], [x \succ_{(f,\varepsilon,s)} y']$, and $x$ are pairwise concordant, and for at most one $y \in X'$ it is not the case that $[x \succ_{(f,\varepsilon,s)} y] \succ x$.

As in the proof of Lemma C.8 we assume, without loss of generality, that $[x \succ_{(f,\varepsilon,s)} y] \sim [x \succ_{(f,\varepsilon,s)} y']$. To build intuition, suppose the only optimal partitions for $x$ are $P, Q \in \mathcal{M}_p^\#$ with $P \neq Q$. Suppose, further, that $P$ is optimal for $x \succ_{(f,\varepsilon,s)} y$ and $Q$ is not, while $Q$ is optimal for $x \succ_{(f,\varepsilon,s)} y'$ and $P$ is not. Again without loss of generality, suppose that $Q$ is not finer than $P$. In that case

$$[(1 - t) x \succ_{(f,\varepsilon,s)} y + tx_1 (P)] \succ [(1 - t) x \succ_{(f,\varepsilon,s)} y' + tx_1 (P)]$$

violating concordance of $x \succ_{(f,\varepsilon,s)} y$ and $x \succ_{(f,\varepsilon,s)} y'$. Hence, it must be that either $P$ or $Q$ is optimal for both. The same argument applies if more than two partitions are optimal in $x$. Thus, if $\succsim_{(x,s)}$ is complete on $X'$, then there is $P \in \mathcal{M}_p^\#$ such that for every $y \in X'$ there is $\varepsilon > 0$ with $P$ optimal in $x \succ_{(f,\varepsilon,s)} y$. Therefore, if $\succsim_{(x,s)}$ is complete on $X'$, then it is represented on $X'$ by $v_x(\cdot, P)$ for some $P \in \mathcal{M}_p^\#$. Because the representation is self generating, $\succsim_{(x,s)}$ must satisfy Axioms 1–5 on $X'$. Because $V \in \Phi^*$, the same argument applies to preferences induced by $\succsim_{(x,s)}$, and so on, ad infinitum, which establishes Self-Generation (Axiom 6).

To see that (a) implies (b), note that Lemma C.17 has two implications. First, $\succsim_{(P,s)}$ has a separable and partitional representation $v'_x(\cdot, P)$ as in [C.4]. Because $v_x(\cdot, P)$ also represents $\succsim_{(P,s)}$, it follows that $v_x(\cdot, P)$ and $v'_x(\cdot, P)$ are identical up to a monotone transformation. But, by L-Indifference to Timing (Axiom 2(d)), it must be that $v_x(\cdot, P)$ and $v'_x(\cdot, P)$ are unique up to a positive affine transformation on $L$. Let us re-normalize $v'_x(\cdot, P)$ so that $v_x(\cdot, P) = v'_x(\cdot, P)$ on $L$.

Second, because $\succsim_{(P,s)}$ satisfies Axioms 1–6, it satisfies the hypotheses of Proposition C.20. Together with Lemma C.19 and IICC (Axiom 4), this implies that $\ell^* \succsim_{(P,s)} y \succsim_{(P,s)} \ell^*$ for all $y \in X$. Because $v_x(\cdot, P)$ and $v'_x(\cdot, P)$ both represent $\succsim_{(P,s)}$, they must agree on $X$ because they agree on $L$. It follows that $v_x(\cdot, P)$ also has a representation as in [C.4], that is, it can be written as

$$v_x(x, P) = \max_{P \in \mathcal{M}_p^\#(P)} \sum_{f \in Q} \pi_0(J) f \max_{f \in \mathcal{X}} \sum_s \pi_0(s | J) [u'_x(f_1(s)) + v'_x(f_2(s); P')]$$

Then, because $\succsim_{(x,s)}$ satisfies Axioms 1–6, it follows from the reasoning above that each $v'_x(\cdot, P')$ in the above representation of $v_x(\cdot, P)$ also has a representation as in [C.4], and so on, ad infinitum, which demonstrates that $V \in \Phi^*$.

\[\square\]

C.5. Recursive Representation

We now establish a recursive representation for $\succsim$, thereby proving Theorem 2.

Recall from Appendix A.7 that $\succsim_{|L}$ has a standard raa representation $((u_x, \delta, \Pi))$. That is, there exist functions $V^*_L(\ell, s) : L \to \mathbb{R}$ such that $V^*_L(\ell, \pi_0) := \sum_s \pi_0(s) V^*_L(\ell, s)$ represents $\succsim_{|L}$, and

$$V^*_L(\ell, s) := \sum_{s'} \Pi(s, s') [u_x(\ell_1(s')) + \delta V^*_L(\ell_2(s'), s')]$$

53
where \( u_s(c_s^\dagger) = 0 \) for all \( s \in S \). This implies \( V_L^*(\ell^\dagger, s) = 0 \) for all \( s \), so that \( V_L^*(\ell^\dagger, \pi_0) = 0 \). The function \( V_L^* \) (recall that \( V_L^* \) also denotes the linear extension of \( V_L^* \) to \( \Delta(L) \)) is uniquely determined by the tuple \( ((u_s)_{s \in S}, \delta, \Pi) \).

By Proposition C.11 \( \succeq \) has a self-generating representation \( V \in \Phi^* \) that satisfies \( V(\ell^\dagger) = 0 \). Now, \( V|_L \) and \( V_L^*(\cdot, \pi_0) \) both represent \( \succeq |_L \) on \( L \). Because \( \succeq |_L \) is continuous and satisfies Independence on \( L \), it follows from the Mixture Space Theorem — see Herstein and Milnor (1953) — that \( V|_L \) and \( V_L^*(\cdot, \pi_0) \) are identical up to a positive affine transformation. Given that \( V(\ell^\dagger) = V_L^*(\ell^\dagger, \pi_0) = 0 \), the Mixture Space Theorem implies \( V|_L \) and \( V_L^*(\cdot, \pi_0) \) only differ by a scaling. Therefore, rescale the collection \( (u_s)_{s \in S} \) by a common factor so as to ensure \( V|_L = V_L^*(\cdot, \pi_0) \) on \( L \).

Fix \( \omega_0 \) and observe that by Proposition A.1, the tuple \( ((u_s)_{s \in S}, \Pi, \delta, \omega_0) \) induces a unique value function that satisfies [Val]. Notice also that this value function agrees with \( V_L^*(\cdot, \pi_0) \) on \( L \). Therefore, we shall denote this value function, defined on \( X \times \Omega \times S \), by \( V^*(\cdot, \omega_0, \pi_0) \).

The next result proves Theorem 2.

**Proposition C.21.** Let \( V \) be a self-generating representation of \( \succeq \) such that \( \varphi^*(V) = \omega_0 \), and suppose \( V(\cdot) = V^*(\cdot, \omega_0, \pi_0) \) on \( L \). Then, \( V(\cdot) = V^*(\cdot, \omega_0, \pi_0) \) on \( X \).\(^{54}\)

**Proof.** For any \( x \), let \( \sigma(x, \omega_0) \) denote the optimal plan for the utility \( V \) and let \( \sigma_*(x, \omega_0) \) denote the optimal plan for \( V^* \). By Lemma A.15, there exist \( \ell_{\sigma(x, \omega_0)}, \ell_{\sigma_*(x, \omega_0)} \in L_{x, \omega_0} \) such that

\[
V(x) = V(\ell_{\sigma(x, \omega_0)}) \geq V(\ell_{\sigma_*(x, \omega_0)}) = V^*(\ell_{\sigma_*(x, \omega_0)}) = V^*(x, \omega_0, \pi_0)
\]

Reversing the roles of \( V \) and \( V^* \), we obtain once again from Lemma A.15 that

\[
V^*(x, \omega_0, \pi_0) = V^*(\ell_{\sigma_*(x, \omega_0)}) \geq V^*(\ell_{\sigma(x, \omega_0)}) = V(\ell_{\sigma(x, \omega_0)}) = V(x)
\]

In both displays, the second equality obtains because \( V \) and \( V^* \) agree on \( L \). Combining the two inequalities yields the desired result. \( \square \)

Suppose \( V \) represents \( \succeq \) and \( V \in \Phi^* \). Then, there exists an implementation of \( V \), given by \( ((u_s), \hat{Q}, (v_s^{(1)}(\cdot, P))), \pi) \). For ease of exposition, we shall say that the collection \( (v_s^{(n)}(\cdot, P)) \) implements \( V \). Then, for all \( n \geq 1 \), there exists \( (v_s^{(n)}) \in \Phi^* \) that implements \( v_s^{(n-1)} \) and so on. Notice that each \( v_s^{(n)} \) depends on all the past choices of partitions. However, our recursive representation \( V^* \) is only indexed by the current state of the rich, and so is entirely forward looking.

### D. Proof of Proposition 4.3

Let \( V(F, P) \) be the consumption value \( F \) generates under partition \( P \). Let \( \mathcal{P}_\delta \) be the set of partitions accessible under \( \mathcal{M} \). Since \( S \) is finite, so is \( \mathcal{P}_\delta \). Therefore, there is \( \hat{T} < \infty \) by which any \( P \in \mathcal{P}_\delta \) can be reached via some information plan. By part (a) of Definition 4.2, for any such \( P \) there

---

\(^{54}\) It follows immediately from Proposition C.21 that in considering dynamic plans, we may restrict attention to stationary plans. This is because we have a recursive formulation with discounting where all our payoffs are bounded, which obviates the need for non-stationary plans — see, for instance, Proposition 4.4 of Bertsekas and Shreve (2000) or Theorem 1 of Orkin (1974).
is an information plan that features learning $P$ every period after $\hat{T}$. For $H \in \mathcal{K} (\mathcal{F} (\Delta (C)))$ let $P_H \in \arg \max_{P \in \mathcal{P}_H} V (H, P)$. Consider now $T^* > \hat{T}$; for $\ell \circ_{T^*} F_{\infty}$, it must then be optimal to follow an information plan that features learning $P_F$ in each period after $T^*$. If $F_T G >_{T^*} F_{\infty}$, then there must be an information plan under which $G$ generates a strictly higher value in some periods than does $F$ under $P_F$. Hence, $V (G, P_G) > V (F, P_F)$, and therefore $G_{\infty} \succeq_{T^*} G_T F$. This establishes part (a) of Definition 4.1.

For partition $P$ let $F^P := \{ f_J : J \in P \} \in \mathcal{K} (\mathcal{F} (\Delta (C)))$ where $f_J (s) = c^+_J$ if $s \in J$ and is $c^-_J$ otherwise. Consider $P$ and $Q$ from part (b) of Definition 4.2. Without loss of generality, there exists $\alpha \in (0, 1]$ such that $V (F^P, P) = V (F^Q_\alpha, Q)$, where $F^Q_\alpha := \alpha F^Q + (1 - \alpha) f_{\emptyset}$. Because the singleton $f_{\emptyset}$ requires no choice, there is no risk of confusion in assuming that $V (F^P, P) = V (F^Q, Q)$.

Because $P$ is maximal in $\mathcal{P}_H$, only an information plan that features $P$ in every period after $T^*$ is optimal for $\ell \circ_{T^*} F^P_{\infty}$. Similarly, only an information plan that features $Q$ in every period after $T^*$ is optimal for $\ell \circ_{T^*} F^Q_{\infty}$. Further, because $Q \notin \Gamma (r (P, \theta, s))$ for any $\theta \in \Theta$ and $s \in S$, $F^P_{\infty} >_{T^*} F^P_{T^*} F^Q$. Analogously, because $P \notin \Gamma (r (Q, \theta, s)), (F^Q_{\infty} >_{T^*} F^Q_{T^*} F^P)$. This establishes part (b) of Definition 4.1.

References


Supplementary Appendix  
For Online Publication Only

All references to definitions and results in this Supplement refer to Dillenberger, Krishna, and Sadowski (2016a) unless otherwise specified.

E. Abstract Static Representation

Let $Y$ be a compact metric space. Then, $\Delta(Y)$ is the space of probability measures on $Y$. For compact metric spaces $Y_1, \ldots, Y_n$, we will consider the product space $Z := \Delta(Y_1) \times \cdots \times \Delta(Y_n)$. We are interested in the space of closed subsets of $Z$, $\mathcal{K}(Z)$ (endowed with the Hausdorff metric), and also in the space of closed and convex subsets $\mathcal{K}_c(Z)$. It is well known that $\mathcal{K}_c(Z)$ is a closed subset of $\mathcal{K}(Z)$.

The convex hull of a set $A$ (in the relevant ambient vector space) is denoted by $\text{ch} A$. If the ambient vector space has a topology, then $\text{cch} A$ denotes the closed convex hull of $A$.

Recall that $C(Y_i)$ is the space of all uniformly continuous functions on $Y_i$ and for $\alpha_i \in \Delta(Y_i)$ and $u_i \in C(Y_i)$, $u_i(\alpha_i) := \int_{Y_i} u_i(y_i) \, d\alpha_i(y_i) =: \langle \alpha_i, u_i \rangle$; endowed with the supremum norm, $C(Y_i)$ is a Banach space. For each $s \in S$, let $L_s \subseteq \Delta(Y_s)$ be a closed subset, and define $L := \times_{s \in S} L_s$. Fix $\ell^s_s \in L_s$, and define $\mathcal{U}_{Y_s, \ell^s_s} := \{u_s \in C(Y_s) : u_s(\ell^s_s) = 0, \|u\|_\infty = 1\}$. Finally, define $\mathcal{U} := \{(p_1 u_1, \ldots, p_n u_n) : u_s \in \mathcal{U}_{Y_s, \ell^s_s}, p_s \geq 0, \sum_s p_s = 1\}$. The space $\mathcal{U}$ will serve as our subjective state space below. It is useful to reconsider $\mathcal{U}$ as $\mathcal{U} := \{(p, u) : p := (p_1, \ldots, p_n) \in \Delta(S), u := (u_1, \ldots, u_n) \in \times_{s \in S} \mathcal{U}_{Y_s, \ell^s_s}\}$.

Specifically, if we consider the domain $X$, then each $Y_s := C \times X$, which then results in a corresponding definition of $\mathcal{U}$.

Theorem 4. Let $\succsim$ be a binary relation on $X$. Then, the following are equivalent:

(a) $\succsim$ satisfies Basic Properties (Axiom 1) and L-Independence (Axiom 2 (a)).

(b) There exists a metric space of continuous functions $\mathcal{U}$ (as defined above) and a minimal set $\mathcal{M}$ of finite, normal, and positive charges\(^{(55)}\) on $\mathcal{U}$ that is weak* compact such that

[i] For all $\ell \in L$ and $s \in S$, $\int_\mathcal{U} p_s u_s(\ell_s) \, d\mu(p, u)$ is independent of $\mu \in \mathcal{M}$, and

[ii] The function $V : X \rightarrow \mathbb{R}$ given by

\[ V(x) := \max_{\mu \in \mathcal{M}} \left[ \int_\mathcal{U} \max_{\alpha \in x} \sum_s p_s u_s(\alpha_s) \, d\mu(p, u) \right] \]

represents $\succsim$.

The proof of Theorem 4 follows immediately from Propositions E.10, E.11, and E.12 below.

E.1. Algebraic Representation

Recall that our domain is $X \simeq \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$. We shall first show that under our assumptions, every closed subset is indifferent to its closed convex hull.

\(^{(55)}\) A charge is a finitely additive measure.
Lemma E.1. If $\succeq$ satisfies Axiom 1, then for each $x \in \mathcal{H}(Z)$, $x \sim \text{cch}(x)$.

Proof. First consider $x \in X$ that is finite and follow Ergin and Sarver (2010a, Lemma 2). Notice that $\text{cch}(x) \succeq x$ by Monotonicity (Axiom 1(d)). Let $x^0 := x$, and for each $k \geq 1$, define $x^k := \frac{1}{2}x^{k-1} + \frac{1}{2}x^{k-1}$. Then, by Aversion to Randomization (Axiom 1(e)), $x^{k-1} \succeq x^k$. In other words, by Order (Axiom 1), $x \succeq x^k$ for all $k \geq 1$. But notice that $d(x^k, \text{cch}(x)) \to 0$ as $k \to \infty$. Therefore, by Continuity (Axiom 1(b)), it follows that $x \succeq \text{cch}(x)$, which proves that $x \sim \text{cch}(x)$ for all finite subsets of $X$.

Now consider the general case, where $x \in X$ is arbitrary. Then, there exists a sequence of finite sets $(x_m)$ such that (i) $x_m \subset x$ for all $m$, and (ii) $d(x_m, x) \to 0$ (in the Hausdorff metric). But each $x_m \sim \text{cch}(x_m)$. It is also easy to see that $d(\text{cch}(x), \text{cch}(x_m)) \to 0$ as $m \to \infty$. Continuity (Axiom 1(b)) now implies that $x \sim \text{cch}(x)$, which proves the claim. □

In light of lemma E.1, in what follows, we may restrict attention to the space $\mathcal{H}_c(X)$.

Lemma E.2. If $\succeq$ satisfies Continuity (Axiom 1(b)) and L-Independence (Axiom 2(a)), then there exists a continuous and affine function $\zeta : L \to \mathbb{R}$ such that $\zeta$ represents $\succeq |_L$, ie, for all $\ell, \ell' \in L$, $\ell \succeq \ell'$ if, and only if, $\zeta(\ell) \geq \zeta(\ell')$.

Proof. Independence and Continuity hold on $L$, so by the Expected Utility Theorem, the claim follows. □

Corollary E.3. If $\succeq$ satisfies Axiom 1, there exist $\ell_\#, \ell_\in \in L$ such that $\ell_\# > \ell_\in$.

Proof. Consider $\ell_\#, \ell_\in \in L$ that exist by Lipschitz continuity (Axiom 1(c)). Set $x = y = \{\ell_\#\}$ and $\alpha = \frac{1}{2}$. Lipschitz continuity then implies $\ell_\# > \frac{1}{2}\ell_\# + \frac{1}{2}\ell_\in$. Similarly, let $x = y = \{\ell_\in\}$ and $\alpha = \frac{1}{2}$, so Lipschitz continuity implies $\frac{1}{2}\ell_\# + \frac{1}{2}\ell_\in > \ell_\in$. It follows immediately that $\ell_\# > \ell_\in$. □

Lemma E.4. Given the function $\zeta : L \to \mathbb{R}$ from lemma E.2 above, there exists $V : X \to \mathbb{R}$ such that

(a) $x \succeq y$ if, and only if, $V(x) \geq V(y)$ for all $x, y \in X$,
(b) for all $\ell \in L$, $V(\ell) = \zeta(\ell)$, and
(c) $V$ is continuous.

Proof. By Corollary E.3, $\ell^* \sim \ell_\ast$. First, consider the case where $x \in X$ is such that $\ell^* \succeq x \succeq \ell_\ast$. By Continuity (Axiom 1(b)), there exists $\alpha \in [0, 1]$ such that $x \sim \alpha \ell^* + (1 - \alpha)\ell_\ast$. Define $V(x) := \zeta(\alpha \ell^* + (1 - \alpha)\ell_\ast) = a\zeta(\ell) + (1 - a)\zeta(\ell_\ast)$. It is easy to see that for all $\ell \in L$, $V(\ell) = \zeta(\ell)$.

Next, consider the case where $x \succ \ell^*$. By Continuity, for any $\ell \in L$, there exists $\alpha \in [0, 1]$ such that $ax + (1 - \alpha)\ell_\ast \sim \ell$. Now, set $V(x) = [V(\ell) - (1 - a)V(\ell_\ast)]/a$.

To see that $V(x)$ is independent of the choice of $\ell$, suppose $\ell' \in L$ and $a' \in [0, 1]$ are such that $\ell \succeq \ell'$ and $a'x + (1 - a')\ell_\ast \sim \ell'$, so that $V(x) = [V(\ell') - (1 - a)V(\ell_\ast)]/a'$. Because $ax + (1 - a)\ell_\ast \sim \ell$, for all $b \in [0, 1]$, $b(ax + (1 - a)\ell_\ast) + (1 - b)\ell_\ast \sim b\ell + (1 - b)\ell_\ast$. Now, choose $b$ such that $b\ell + (1 - b)\ell_\ast \sim \ell'$. Then, $b(ax + (1 - a)\ell_\ast) + (1 - b)\ell_\ast \sim \ell'$, which implies $ba = a'$. Using the
fact that $V(\ell') = bV(\ell) + (1 - b)V(\ell_*)$, we see that

$$V(x) = \frac{V(\ell') - (1 - a')V(\ell_*)}{a'} = \frac{[bV(\ell) + (1 - b)V(\ell_*)] - (1 - ba)V(\ell_*)}{ba} = \frac{V(\ell) - (1 - a)V(\ell_*)}{a}$$

which is independent of the choice of $b$, or equivalently, the choice of $\ell'$.

We can deal with case where $\ell_* \succ x$ in a similar fashion. The continuity of $V$ follows immediately from the continuity of $\succ$ and from the continuity of $\zeta$, which completes the proof.

Lemma E.5. If $tx + (1 - t)\ell > ty + (1 - t)\ell$ then $x > y$.

Proof. Suppose not. Then, by L-Independence, there are $x, y, \ell, t$ such that $x \sim y$ and $tx + (1 - t)\ell > ty + (1 - t)\ell$. By Lipschitz Continuity (Axiom 1(c)), and because $d(x, x) = 0$, we have $t'x + (1 - t')\ell_\# > t'\ell_\# + (1 - t')\ell_\#$ for all $t' > 0$. Observe that by Negative Transitivity of the strict relation $\succ$, it must be that for all $t'$, either $t'x + (1 - t')\ell_\# > x$ or $x > t'x + (1 - t')\ell_\#$ holds, and the same for $y$. There are three cases to consider.

Case 1: For all $\varepsilon > 0$ there is $(1 - t') < \varepsilon$ with $x \succ t'x + (1 - t')\ell_\#$. Then, since $x \sim y$, L-Independence implies that $ty + (1 - t)\ell > t(t'x + (1 - t')\ell_\#) + (1 - t)\ell$ for all such $(1 - t') > 0$. At the same time, by continuity, we can pick $(1 - \overline{t}) > 0$ small enough, such that by replacing $x$ with $\overline{t}x + (1 - \overline{t})\ell_\#$, $t(\overline{t}x + (1 - \overline{t})\ell_\#) + (1 - t)\ell > ty + (1 - t)\ell$ still holds. Taking $\varepsilon \leq (1 - \overline{t})$ establishes a contradiction.

Case 2: For all $\varepsilon > 0$ there is $(1 - t') < \varepsilon$ with $t'y + (1 - t')\ell_\# > y$. This case is analogous to case 1.

Case 3: There is $\varepsilon > 0$ such that for all $(1 - t') < \varepsilon$ both, $t'x + (1 - t')\ell_\# \succeq x$ and $y \succ t'y + (1 - t')\ell_\#$. We claim that this case can never occur. To see this, first observe that by continuity, if $t'x + (1 - t')\ell_\# \succeq x$ for all $(1 - t') < \varepsilon$ then $\ell_\# \succeq x$; and if $y \succ t'y + (1 - t')\ell_\# \succeq x$ for all $(1 - t') < \varepsilon$ then $y \succ \ell_\#$. But then we have $y \succeq \ell_\# > \ell_\# \succeq x$, which contradicts the premise that $x \sim y$.

Corollary E.6. It follows immediately from L-Independence and Lemma E.5 that $tx + (1 - t)\ell > ty + (1 - t)\ell$ if, and only if, $x > y$.

Lemma E.7. $\ell \succ \ell'$ if, and only if, $tx + (1 - t)\ell > tx + (1 - t)\ell'$.

Proof. If $x > \ell_*$, by continuity there are $a \in (0, 1)$ and $\overline{t} \in L$ with $ax + (1 - a)\ell_* \sim \overline{t}$. Applying Corollary E.6 repeatedly yields that $\ell \succ \ell'$ if, and only if, $t'[ax + (1 - a)\ell_*] + (1 - t')\ell \sim t'\overline{t} + (1 - t')\ell' \sim t'[ax + (1 - a)\ell_*] + (1 - t')\ell'$ for all $t' \in (0, 1)$. Again by Corollary E.6, and for $t' = \frac{t}{a + t(1 - a)}$, this is equivalent to $tx + (1 - t)\ell \succ tx + (1 - t)\ell'$. The case where $\ell_* \succ x$ is similar and hence omitted.

Lemma E.8. The function $V$ defined in the proof of Lemma E.4 has the following properties:

a) $V$ is monotone, ie, $V(x \cup y) \geq V(x)$ for all $x, y \in X$;

b) $V$ is $L$-affine, ie, for all $x \in X$, $\ell \in L$ and $a \in [0, 1]$, $V(ax + (1 - a)\ell) = aV(x) + (1 - a)V(\ell)$;
(c) \( V \) is midpoint convex, i.e., \( V(\frac{1}{2}x_1 + \frac{1}{2}x_2) \leq \frac{1}{2}V(x_1) + \frac{1}{2}V(x_2) \);

(d) \( V \) is convex.

**Proof.** To ease notational burden, we shall assume only in this part of the proof, and without loss of generality, that \( V(\ell^*) = 1 \) while \( V(\ell_*) = 0 \). We prove the claims in turn.

(a) \( V \) represents \( \succeq \), so it is clear that it is monotone.

(b) Let \( x \in X \) and \( \ell \in L \). Consider first the case where \( \ell^* \succ x \succ \ell_* \). Then, there exists \( \ell_x \in L \) such that \( x \sim \ell_x \). Then, by \( L \)-Independence, for all \( \alpha \in (0, 1) \), \( ax + (1-a)\ell \sim a\ell_x + (1-a)\ell \). Therefore, \( V(ax + (1-a)\ell) = V(a\ell_x + (1-a)\ell) = aV(\ell_x) + (1-a)V(\ell) = aV(x) + (1-a)V(\ell), \) as required.

Now consider the case where \( x \succ \ell^* \), the case where \( \ell_* \succ x \) being analogous. Because \( \ell \succeq \ell_* \), Lemma E.7 yields \( t\ell_x + (1-t)\ell \succeq \ell_* \), and then by Corollary E.6 \( tx + (1-t)\ell \succ \ell_* \). By continuity, there are \( \alpha \in (0, 1) \) and \( \bar{\ell} \), such that \( \ell^* \succ \alpha (tx + (1-t)\ell) \). Further, let \( \beta \in [0, 1] \) be such that \( \ell \sim \beta \ell^* + (1-\beta)\ell_* \) (so that \( V(\ell) = \beta \)), and let \( \gamma \in (0, 1) \) be such that \( \ell \sim \gamma \ell^* + (1-\gamma)\ell_* \). First, from Corollary E.6 and the definition of \( V \) it is easy to verify that \( V(tx + (1-t)\ell) = \frac{\gamma}{\alpha} \) (independent of whether \( tx + (1-t)\ell \succeq \ell^* \) or not). Next, by Lemma E.7, \( tx + (1-t)\ell \sim tx + (1-t)(\beta \ell^* + (1-\beta)\ell_*) \). Then, by Corollary E.6,

\[
\alpha (tx + (1-t)(\beta \ell^* + (1-\beta)\ell_*)) + (1-\alpha)\ell_* \sim \gamma \ell^* + (1-\gamma)\ell_*
\]

or

\[
\alpha tx + \alpha (1-t)\beta \ell^* + [1-\alpha t - \alpha (1-t)\beta] \ell_* \sim \gamma \ell^* + (1-\gamma)\ell_*
\]

Because \( x \succ \ell^* \), Corollary E.6 and Lemma E.7 further imply that \( \alpha (1-t)(1-\beta) + (1-\alpha) > (1-\gamma) \) or \( \gamma - \alpha (1-t) \beta > \alpha t > 0 \). This implies that \( \gamma > \alpha (1-t) \beta \). Corollary E.6 then yields that

\[
\frac{\alpha t}{D_1} x + \frac{1 - \alpha t - \alpha (1-t)\beta}{D_1} \ell_* \sim \gamma \frac{1 - \alpha t - \alpha (1-t)\beta}{D_1} \ell_*
\]

where \( D_1 = \gamma - \alpha (1-t) \beta + (1-\gamma) = 1 - \alpha (1-t) \beta \).

It follows that \( 1 - \gamma < 1 - \alpha t - \alpha (1-t) \beta \), and hence, again by Corollary E.6,

\[
\frac{\alpha t}{D_2} x + \frac{1 - \alpha t - \alpha (1-t)\beta - (1-\gamma)}{D_2} \ell_* \sim \ell^*
\]

where \( D_2 = \alpha t + 1 - \alpha t - \alpha (1-t) \beta - (1-\gamma) = \gamma - \alpha (1-t) \beta \).

Hence, \( \frac{\alpha t}{\alpha t - \alpha (1-t) \beta} x + \left[ 1 - \frac{\alpha t}{\alpha t - \alpha (1-t) \beta} \right] \ell_* \sim \ell^* \), so that \( V(x) = \frac{\gamma - \alpha (1-t) \beta}{\alpha t} \). Putting everything together establishes the lemma, i.e.,

\[
tV(x) + (1-t)V(\ell) = \frac{\gamma}{\alpha} = V(tx + (1-t)\ell)
\]

(c) Suppose first that \( x_1 \sim x_2 \). Then, by Aversion to Randomization (Axiom 1 (e)), \( x_1 \succeq \frac{1}{2}x_1 + \frac{1}{2}x_2 \), from which it follows immediately that \( V(\frac{1}{2}x_1 + \frac{1}{2}x_2) \leq \frac{1}{2}V(x_1) + \frac{1}{2}V(x_2) \).

Let us now suppose that \( x_1 \succ x_2 \) and consider the case where \( \ell^* \succ x_1 \). By continuity, there exists \( \lambda \in (0, 1) \) such that \( y := \lambda x_2 + (1-\lambda) \ell^* \sim x_1 \). Notice that because \( V \) is \( L \)-affine, \( V(y) = \lambda V(x_2) + (1-\lambda) V(\ell^*) \). Let \( \bar{x} := \frac{\lambda}{1+\lambda} x_1 + \frac{1}{1+\lambda} y = \frac{\lambda}{1+\lambda} (\frac{1}{2}x_1 + \frac{1}{2}x_2) + (1+\lambda) \ell^* \), so that \( V(\bar{x}) = \frac{2\lambda}{1+\lambda} V(\frac{1}{2}x_1 + \frac{1}{2}x_2) + \frac{1-\lambda}{1+\lambda} V(\ell^*) \), where we have used the \( L \)-affinity of \( V \). But
notice also that $V(\tilde{x}) \leq \frac{\lambda}{1+\lambda} V(x_1) + \frac{1}{1+\lambda} V(y)$ by Aversion to Randomization (Axiom 1 (e)) because $x_1 \sim y$. We also have $\frac{\lambda}{1+\lambda} V(x_1) + \frac{1}{1+\lambda} V(y) = \frac{\lambda}{1+\lambda} (V(x_1) + V(x_2)) + \frac{1-\lambda}{1+\lambda} V(\ell^*)$. Substituting in the value of $V(\tilde{x})$ obtained above, we see that $V(\frac{1}{2}x_1 + \frac{1}{2}x_2) \leq \frac{1}{2} V(x_1) + \frac{1}{2} V(x_2)$, as claimed.

Now consider the case where $x_1 \succ x_2$ but $x_1 \succ \ell^*$. Then, by continuity, there exists $a \in [0, 1]$ such that $y = ax_1 + (1-a)\ell^* \sim x_2$. Therefore, $V(y) = aV(x_1) + (1-a)V(\ell^*) = V(x_1)$. Set $\tilde{x} = \frac{a}{1+a} x_2 + \frac{1}{1+a} y = \frac{2a}{1+a} (\frac{1}{2}x_1 + \frac{1}{2}x_2) + \frac{1-a}{1+a} \ell^*$. Then, using the $L$-affinity of $V$, we obtain $V(\tilde{x}) = \frac{2a}{1+a} V(\frac{1}{2}x_1 + \frac{1}{2}x_2) + \frac{1-a}{1+a} V(\ell^*)$.

But notice that $x_2 \sim y$, so that by Aversion to Randomization (Axiom 1 (e)), $V(\tilde{x}) \leq \frac{a}{1+a} V(x_2) + \frac{1}{1+a} V(y)$. We also have $\frac{a}{1+a} V(x_1) + \frac{1}{1+a} V(y) = \frac{a}{1+a} (V(x_1) + V(x_2)) + \frac{1-a}{1+a} V(\ell^*)$. Substituting in the value of $V(\tilde{x})$ obtained above, we see that $V(\frac{1}{2}x_1 + \frac{1}{2}x_2) \leq \frac{1}{2} V(x_1) + \frac{1}{2} V(x_2)$, as claimed.

(d) As noted above, $V$ is continuous, and because it is midpoint convex, it is convex. □

Recall that $V$ is Lipschitz if there exists a constant $K > 0$ such that for all $x, y \in X$, $|V(x) - V(y)| \leq Kd(x, y)$, where $d(\cdot, \cdot)$ is the metric on $X$.

**Lemma E.9.** If $\succ$ satisfies Lipschitz continuity (Axiom 1(c)) and is represented by a continuous and $L$-affine $V$, then $V$ is Lipschitz. Conversely, if $V$ is Lipschitz, non-trivial, $L$-affine, and represents $\succ$, then it satisfies Lipschitz continuity.

**Proof.** Let $N > 0$ be as given in Lipschitz continuity. Fix $\beta \in (0, 1)$ such that $N\beta < 1$. First consider the case where $x, y \in X$ are such that $0 < d(x, y) \leq \beta$ and let $\alpha = Nd(x, y)$. Then, by Lipschitz Continuity, $(1-\alpha)x + \alpha \ell^* > (1-\alpha)y + \alpha \ell^*$. By the $L$-affinity of $V$, it follows that $V(y) - V(x) < \frac{\alpha}{L} [V(\ell^*) - V(\ell^*)]$. But notice that $\alpha/N \leq \beta$, so setting $K = N/(1-N\beta) \frac{[V(\ell^*) - V(\ell^*)]}{d(x, y)}$, we find that

$$V(y) - V(x) < \frac{\alpha}{1-\alpha} [V(\ell^*) - V(\ell^*)]$$

$$< \frac{N}{1-\alpha} [V(\ell^*) - V(\ell^*)] d(x, y)$$

$$< Kd(x, y)$$

We now follow Dekel et al. (2007) and remove the restriction on the $x$ and $y$. For arbitrary $x, y \in X$, let $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_{J+1} = 1$ such that $(\lambda_{j+1} - \lambda_j)d(x, y) \leq \beta$ for all $j = 0, \ldots, J + 1$. Define $x_j := \lambda_j x + (1-\lambda_j)y$, so $d(x_{j+1}, x_j) = (\lambda_{j+1} - \lambda_j)d(x, y) < \beta$. From the result established above, we see that $V(x_{j+1}) - V(x_j) \leq Kd(x_{j+1}, x_j) = K(\lambda_{j+1} - \lambda_j)d(x, y)$. Summing over $j$, we find $V(y) - V(x) \leq Kd(x, y)$. Interchanging the roles of $x$ and $y$, it follows that $|V(x) - V(y)| \leq Kd(x, y)$, as claimed. The converse is as in Dekel et al. (2007) and is omitted. □

In sum, we have proven that (a) implies (b) in the following representation result.

**Proposition E.10.** Let $\succ$ be a binary relation. Then, the following are equivalent:

(a) $\succ$ satisfies Basic Properties (Axiom 1) and $L$-Independence (Axiom 2(a)).

(b) There exists a function $V : X \to \mathbb{R}$ that represents $\succ$ and is $L$-affine, Lipschitz Continuous, and convex. Moreover, any such representation of $\succ$ is unique up to a positive affine transformation.

The proof that (b) implies (a) is standard and is omitted.
E.2. Abstract Convex and Monotone Representation

Every $\alpha \in \mathcal{F}(\Delta(C \times X))$ is a product lottery of the form $\alpha_1 \times \cdots \times \alpha_n$. A function $u \in \mathcal{U}$ acts on $\mathcal{F}(\Delta(C \times X))$ as follows: $u(\alpha) := \sum_i p_i u_i(\alpha_i)$. For any $x \in \mathcal{K}_c(\mathcal{F}(\Delta(C \times X)))$, define its support function $H_x : \mathcal{U} \to \mathbb{R}$ as $H_x(u) := \max_{\alpha \in \mathcal{X}} u(\alpha)$. The extended support function of $x \in \mathcal{K}_c(\mathcal{F}(\Delta(C \times X)))$ is the unique extension of the support function $H_x$ to span($\mathcal{U}$) by positive homogeneity. Theorem 5.102 and Corollary 6.27 of Aliprantis and Border (1999) imply that a function defined on span($\mathcal{U}$) is sublinear, norm continuous, and positively homogeneous if, and only if, it is the extended support function of some weak* closed, convex subset of $\mathcal{F}(\Delta(C \times X))$. Therefore, a function $H : \mathcal{U} \to \mathbb{R}$ is a support function if its unique extension to span($\mathcal{U}$) by positive homogeneity is sublinear and norm continuous.

Given a function $H : \mathcal{U} \to \mathbb{R}$ whose extension to span($\mathcal{U}$) by positive homogeneity is sublinear and norm continuous, we may define $x_H := \{ \alpha \in \text{aff}(Z) : u(\alpha) \leq H(u) \text{ for all } u \in \mathcal{U} \}$. Support functions enjoy the following duality: For any weak* compact, convex subset $x$ of aff($Z$), $x_{H_x} = x$, and for any function $H$ as defined above, $H_{x_H} = H$.

For weak* compact, convex subsets $x$ and $x'$ of $X$, support functions exhibit the following properties: (i) $x \subseteq x'$ if, and only if, $H_x \leq H_{x'}$, (ii) $H_{tx+(1-t)x'} = tH_x + (1-t)H_{x'}$ for all $t \in (0, 1)$, (iii) $H_{x \cap x'} = H_x \wedge H_{x'}$, and (iv) $H_{\text{ch}(x \cup x')} = H_x \vee H_{x'}$. (By Lemma 5.14 of Aliprantis and Border (1999), ch($x \cup x'$) is compact because $x$ and $x'$ are compact, which ensures that $H_{\text{ch}(x \cup x')}$ is well defined.) Finally, observe that for $\ell^\dagger := \ell^\dagger_1 \times \cdots \times \ell^\dagger_n$, $H_{\ell^\dagger} = 0$.

**Proposition E.11.** Let $V : X \to \mathbb{R}$ be Lipschitz, convex, and $L$-affine. Then, there exists a minimal set $\mathcal{M}$ of finite normal charges on $\mathcal{U}$ so that $V$ can be written as

\[
V(x) = \max_{\mu \in \mathcal{M}} \left[ \int_{\mathcal{U}} \max_{\alpha \in \mathcal{X}} \sum_i p_i u_i(\alpha_i) \, d\mu(p, u) \right]
\]

where the set $\mathcal{M} \subseteq \text{ba}_n(\mathcal{U})$ is weak* compact and $\int_{\mathcal{U}} \max_{\alpha \in \mathcal{X}} \sum_i p_i u_i(\alpha_i) \, d\mu(p, u)$ is independent of $\mu$ for all $x \in L$.

Moreover, for a dense set of points in $X$, there is a unique $\mu \in \mathcal{M}$ that achieves the maximum in [•].

In Proposition E.11 above, $\text{ba}_n(\mathcal{U})$ is the space of bounded additive (or finitely additive) measures (ie, charges) on $\mathcal{U}$ that are also normal (ie, inner and outer regular). The last part of the proposition reflects the fact that $V$ is linear on $L$. The set $\mathcal{M}$ is minimal in the sense that if $\mathcal{N} \subseteq \mathcal{M}$ is compact, then there exists $x \in X$ such that $V(x) > \max_{\mu \in \mathcal{M}} \left[ \int_{\mathcal{U}} \max_{\alpha \in \mathcal{X}} \sum_i p_i u_i(\alpha_i) \, d\mu(p, u) \right]$.

**Proof.** By Lemma E.1, for every $x \in \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$, $V(x) = V(\text{ch}(x))$. Therefore, we may restrict attention to convex menus.

Let $\Psi : \mathcal{K}_c(\mathcal{F}(\Delta(C \times X))) \to \mathcal{C}_b(\mathcal{U})$ be the map that associates each compact, convex subset $x$ of $\mathcal{F}(\Delta(C \times X))$ with its support function, $\Psi : x \mapsto H_x$. Note that $\Psi$ is invertible. Moreover, $\Psi$ is an isometry because $d(x, x') = \|H_x - H_{x'}\|_\infty$ for all $x, x' \in \mathcal{K}_c(\mathcal{F}(\Delta(C \times X)))$. Thus $\Psi$ is an affine isometric embedding of $\mathcal{K}_c(\mathcal{F}(\Delta(C \times X)))$ in $\mathcal{C}_b(\mathcal{U})$. Moreover, $\Psi(\ell^*) = 0$. In sum, $\Psi(\mathcal{K}_c(\mathcal{F}(\Delta(C \times X))))$ is a compact and convex subset of $\mathcal{C}_b(\mathcal{U})$ that contains the origin.

(56) Recall that $\text{ba}_n(\mathcal{U})$ is the space of finite normal charges on $\mathcal{U}$.
Let \( \hat{V} : \Psi(\mathcal{K}(\Delta(C \times X))) \to \mathbb{R} \) be defined as follows: \( \hat{V}(H) := V(x) \) where \( H = H_x \) for some \( x \). Because \( \Psi \) is injective, it follows that \( \hat{V} \) is well defined. Thus, \( \hat{V} \) is Lipschitz, convex, and \( \Psi(L) \)-affine. Recall that by definition, \( V(\ell^*) = 0 = \hat{V}(H(\ell^*)) \), and \( \Psi(\ell^*) = 0 \). Therefore, \( \hat{V} \) is positively homogeneous. Extending \( \hat{V} \) to \( \text{cone}(\Psi(\mathcal{K}(\Delta(C \times X)))) \) by positive homogeneity, it follows by Proposition F.4 below that \( \hat{V} \) (and hence \( V \)) has the desired representation.

**Proposition E.12.** Let \( V : \mathcal{K}(\Delta(C \times X)) \to \mathbb{R} \) be as in \([\star]\). Then, the following are equivalent.
(a) \( V \) is monotone, in the sense that \( x \geq x' \) implies \( V(x) \leq V(x') \).
(b) Every charge \( \mu \in \mathcal{M} \) is positive, ie, \( \mu(E) \geq 0 \) for all (Borel) measurable \( E \subset \Omega \).

**Proof.** That (b) implies (a) is easy to see. That (a) implies (b) follows from Theorem S.2 of Ergin and Sarver (2010a) after observing that \( \hat{V} \) (defined in the proof of E.12) is monotone. We note that a similar statement is contained in the proof of Lemma 3.5 of Gilboa and Schmeidler (1989).

The following corollary follows immediately from Lemma F.5.

**Corollary E.13.** Let \( V : \mathcal{K}(\Delta(C \times X)) \to \mathbb{R} \) have a representation as in \([\star]\). Suppose \( E \subset \mathcal{K}(\Delta(C \times X)) \) is convex and \( V|_E \) is linear. Then, there exists \( \mu \in \mathcal{M} \) such that \( V(x) = \int_{\Omega} \max_{\alpha \in \Xi} \sum_i p_i u_i(\alpha_i) \, d\mu(p, u) \) for all \( x \in E \).

### F. Convex Duality

We review some notions from convex analysis. Our review follows Ekeland and Turnbull (1983).

Let \( X \) be a Banach space, \( X^* \) its norm dual, \( C \subset X \), and \( f : C \to X \) a convex and Lipschitz function. The **subdifferential** of \( f \) at \( x \in C \) is \( \partial f(x) := \{ x^* \in X^* : (y - x, x^*) \leq f(y) - f(x) \} \) for all \( y \in C \). A necessary and sufficient condition for the existence of a subdifferential at \( x \in C \) is that there exists \( K \geq 0 \) such that for all \( y \in X \), \( f(x) - f(y) \leq K \| y - x \| \). To see this, recall that the set \( \text{epi}(f) := \{(x, t) \in X \times \mathbb{R} : t \geq f(x)\} \), the **epigraph** of the function \( f \), is a convex set (if, and only if, \( f \) is a convex function). For each \( x \in C \), we define \( A(x) := \{(y, t) \in X \times \mathbb{R} : f(x) - t \geq K \| y - x \| \} \). It is easy to see that the set \( A(x) \) is (i) nonempty, (ii) convex, and (iii) open. It is also easy to show that \( \text{epi}(f) \cap A(x) = \emptyset \), so there exists a non-vertical hyperplane that separates the two sets. Following the arguments in Gale (1967), we can conclude that \( \partial f(x) \neq \emptyset \), and moreover, there exists \( x^* \in \partial f(x) \) such that \( \| x^* \| \leq K \). This is the content of the Duality Theorem of Gale (1967). (Indeed, Gale (1967) also shows that local Lipschitzness is a necessary condition for \( \partial f(x) \) to be nonempty.) We will rely on the following result in the sequel.

**Proposition F.1** (Duality Theorem in Gale (1967)). Let \( C \subset X \) be convex and suppose \( f : C \to \mathbb{R} \) is convex and Lipschitz of rank \( K \). Then, there exists \( x^* \in \partial f(x) \) such that \( \| x^* \| \leq K \).

In what follows, we will denote by \( \partial_K f(x) := \{ x^* \in \partial f(x) : \| x^* \| \leq K \} \). For each \( x^* \in X^* \) and \( a \in \mathbb{R} \), we can define the continuous affine functional \( \varphi(\cdot, x^*) : X \to \mathbb{R} \) as \( \varphi(y; x^*) := (y, x^*) - a \). The function \( \varphi \leq f \) for all \( y \in C \) if, and only if, \( (y, x^*) - a \leq f(y) \), and is **exact** at \( x \in C \) if \( \varphi(x; x^*) = f(x) \). If \( \varphi \) is exact, the value of \( a \) which makes it so is given by \( -a(x^*) := f(x) - \langle x, x^* \rangle \). Therefore, \( x^* \in \partial f(x) \) if, and only if, the continuous affine functional \( \varphi(y; x^*) =
\( f(x) + \langle y - x, x^* \rangle \leq f(y) \) for all \( y \in C \) with \( \varphi(x; x^*) = f(x) \). In other words, \( x^* \in \partial f(x) \) if, and only if, \( \varphi(y; x^*) = f(x) + \langle y - x, x^* \rangle \) is a supporting hyperplane for the graph of \( f \) at \( x \).

Notice that for any intercept \( a \geq a(x^*), \langle x, x^* \rangle - a < \langle x, x^* \rangle - a(x^*), \) so \( a(x^*) = \inf \{a \in \mathbb{R} : f(x) \geq \langle x, x^* \rangle - a \} = \sup \{x \in C : \langle x, x^* \rangle - f(x) \} \). This smallest intercept is the Fenchel conjugate of \( f \), and is denoted by \( f^* : X^* \to \mathbb{R} \cup \{-\infty, +\infty\} \), and is given by

\[
 f^*(x^*) := \sup_{x \in C} \left[ \langle x, x^* \rangle - f(x) \right]
\]

Proposition 2 of Ekeland and Turnbull (1983) shows that \( x^* \in \partial f(x) \) if, and only if, \( f(x) + f^*(x^*) = \langle x, x^* \rangle \).

By Proposition F.1, it follows that for Lipschitz \( f \), the conjugate function is given by \( f^*(x^*) := \max_{x \in C} \left[ \langle x, x^* \rangle - f(x) \right] \). We now show that for positively homogeneous functions, the conjugate function \( f^* \) is identically 0.

**Proposition F.2.** Let \( C \subset X \) be a convex cone, and let \( f : C \to \mathbb{R} \) be convex and Lipschitz. Then, the following are equivalent:

(a) \( f \) is positively homogeneous, i.e., \( f(\lambda x) = \lambda f(x) \) for all \( \lambda > 0 \);  
(b) \( f^*(x^*) \in \mathbb{R} \) implies \( f^*(x^*) = 0 \).

**Proof.** Suppose \( f^* = 0 \). Fix \( x \in C \), and recall that because \( f \) is convex and Lipschitz, there exists \( x^* \in \partial f(x) \). This implies \( f(x) = \langle x, x^* \rangle \). It is easy to see that \( x^* \in \partial f(\lambda x) \) for all \( \lambda > 0 \), so that \( f(\lambda x) = \lambda f(x) \). That is, \( f \) is positively homogeneous.

Now suppose \( f \) is positively homogeneous. Fix \( x \in C \) and suppose \( x^* \in \partial f(x) \). We will first show that for any \( \lambda > 0, x^* \in \partial f(\lambda x) \). Then, by the definition of \( \partial f \), for any \( y \in C, \langle y - x, x^* \rangle \leq f(y) - f(x^*) \). Now let \( \lambda > 0 \) and let \( y \in C \) be arbitrary. Because \( C \) is a cone, there exists \( z \in C \) such that \( \lambda z = y \). This implies \( y - \lambda z = \lambda (z - x^*) \leq \lambda [f(z) - f(x)] = f(y) - f(\lambda x) \), which proves that \( x^* \in \partial f(x) \) implies \( x^* \in \partial f(\lambda x) \) for all \( \lambda > 0 \).

Now suppose \( x^* \) is such that \( f^*(x^*) \in \mathbb{R} \). Because \( f \) is positively homogeneous, we have \( f(0) = 0 \). (To see this, note that \( f(0) = f(2 \times 0) = 2f(0) \) which implies \( f(0) = 0 \).) Therefore, \( f^*(x^*) \geq \langle 0, x^* \rangle - f(0) = 0 \). Now suppose \( f^*(x^*) > 0 \). Then, for any \( \varepsilon \in (0, f^*(x^*)) \), there exists \( x \in C \) such that \( f^*(x^*) - \varepsilon = \langle x, x^* \rangle - f(x) > 0 \). But then we can choose \( \lambda > 0 \) such that \( \langle \lambda x, x^* \rangle - f(\lambda x) > f^*(x^*) \), which is a contradiction. Therefore, it must be that \( f^*(x^*) = 0 \).

This allows us to establish the following corollary.

**Corollary F.3.** Let \( C \subset X \) be a convex cone, and \( f \in \mathbb{R}^C \) be convex, Lipschitz, and positively homogeneous. Then, there exists a weak* compact set \( \mathcal{M} \subset X^* \) such that \( f(x) = \max \{\langle x, x^* \rangle : x^* \in \mathcal{M} \} \).

**Proof.** We have already established that for each \( x \in C \), there exists \( x^* \in \partial f(x) \) such that \( \|x^*\| \leq K \), where \( K \) is the Lipschitz constant of \( f \). We have also established that \( x^* \in \partial f(\lambda x) \) for all \( \lambda \geq 0 \). Therefore, \( f(y) \geq \langle y, x^* \rangle \) for all \( y \in C \). Letting \( \mathcal{M} = \text{cl}(\{x^* \in \partial f(x) : x \in C, \|x^*\| \leq K \}) \) (in the weak* topology) establishes the claim.
If $C$ is convex and $A \subseteq C$ is also convex, then $f : C \to \mathbb{R}$ is $A$-affine if for all $x \in C$, $a \in A$, and $t \in (0, 1)$, we have $f(tx + (1-t)a) = tf(x) + (1-t)f(a)$.

For a fixed $x \in C$, notice that $f$ is affine on the set $\text{ch}(\{x\} \cup A)$. Let $\mathcal{C}_x$ be the collection of all (convex) subsets of $C$ such that if $E \in \mathcal{C}_x$ then (i) $x \in E$ and (ii) $f|_E$ is affine. A simple application of Zorn’s lemma shows that for each $x \in C$, there is a largest set $E_x$ that contains $x$ and where $f|_{E_x}$ is affine.

Notice that there exist $x \in X$ such that this maximal set $E_x$ is not unique. Indeed, for any $a \in A$, and $x, y \in C$ such that $f$ is not affine on $[x, y]$ (the closed line segment joining $x$ and $y$), then $a \in E_x \cap E_y$, but $E_x \cup E_y$ (or it’s convex hull) is not a member of $\mathcal{C}_a$.

If $f$ is Lipschitz continuous (as we shall assume below), then it is easy to see that the set $E_x$ must be closed as well.

**Proposition F.4.** Let $C \subseteq X$ be a convex set, and $f \in \mathbb{R}^C$ be convex and Lipschitz of rank $K$. Let $A \subseteq C$ be convex and suppose that $0 \in A$, $f(0) = 0$, and that $f$ is $A$-affine. Then, for each $x$, there exists $x^* \in X^*$ such that $x^* \in \partial f_K(y)$ for all $y \in E_x$ where $E_x$ is defined above. Moreover, there exists a weak* compact set $\mathcal{M}_f \subseteq X^*$ such that $f(x) = \max\{\langle x, x^* \rangle : x^* \in \mathcal{M}_f \}$ and $\langle a, x^* \rangle$ is independent of $x^* \in \mathcal{M}_f$ for all $a \in A$.

**Proof.** Fix $x \in C$, let $y_1, \ldots, y_n \in E_x$, and define $y := \frac{1}{n} \sum_i n$. Then, by Proposition F.1, there exists $y^* \in \partial f(y)$. Recall the affine function $\varphi(\cdot, y^*) : X \to \mathbb{R}$ given by

$$\varphi(x; y^*) := \langle x - y, y^* \rangle + f(y)$$

The affine function $\varphi$ satisfies the following two properties:

- $f(x) \geq \varphi(x; y^*)$ for all $x \in C$, and
- $f(y) = \varphi(y; y^*)$.

The first requirement implies that $f(y_i) \geq \varphi(y_i; y^*)$ for all $i = 1, \ldots, n$. Summing up and dividing by $n$, we see that $\frac{1}{n} \sum_i f(y_i) \geq \frac{1}{n} \sum_i \varphi(y_i; y^*)$. However, $f$ restricted to $E_x$ is affine which implies $\frac{1}{n} \sum_i f(y_i) = f(y)$; similarly, $\varphi$ is affine, which implies $\frac{1}{n} \sum_i \varphi(y_i; y^*) = \varphi(y; y^*)$.

But we have noted above that $f(y) = \varphi(y; y^*)$, which is possible if, and only if, $f(y_i) = \varphi(y_i; y^*)$ for all $i = 1, \ldots, n$. But this is equivalent to saying that $y^* \in \partial f(y_i)$.

For any $y \in E_x$, $\partial f(y)$ is a (nonempty) closed (and hence compact) subset of $\{x^* \in X^* : \|x^*\| \leq K\}$. 57 Thus, $(\partial f(y))_{y \in E_x}$ is a collection of closed subsets of the compact set $\{x^* \in X^* : \|x^*\| \leq K\}$. But we have just established that for any $y_1, \ldots, y_n \in E_x$, $\bigcap_{i=1}^n \partial f(y_i) \neq \emptyset$. In other words, the collection of closed sets $(\partial f(y))_{y \in E_x}$ has the finite intersection property. The compactness of $\{x^* \in X^* : \|x^*\| \leq K\}$ then implies that $\bigcap_{y \in E_x} \partial f(y) \neq \emptyset$. Thus, there exists $\zeta_x \in \bigcap_{y \in E_x} \partial f(y)$ which proves the first part.

Fix this $\zeta_x$ and notice that $\varphi(y; \zeta_x) = f(y)$ for all $y \in E_x$. Because $0 \in A$, this implies $\varphi(0; \zeta_x) = 0$. In other words, $f^*(\zeta_x) = 0$. (In geometric terms, the supporting hyperplane determined by $\zeta_x$ passes through the origin.) Now, let $\mathcal{M}_f := \text{cl}\{\zeta_x \in X^* : x \in C\}$. It is immediate that $\mathcal{M}_f$ is closed. Because $f(a) = \langle a, \zeta_x \rangle$ for all $x \in C$, it follows that the same holds for all $x^* \in \mathcal{M}_f$, which completes the proof.

---

57 By the Banach-Alaoglu Theorem — see, for instance, Theorem 6.25 of Aliprantis and Border (1999) — the set $\{x^* \in X^* : \|x^*\| \leq K\}$ is a weak* compact subset of the dual $X^*$. 
We end with an easy observation.

**Lemma F.5.** Let \( C \subseteq X \) be a convex set, and \( f \in \mathbb{R}^C \), and \( M_f \) a weak* compact subset of \( X^* \) such that for all \( x \in C \), \( f(x) = \max \{ \langle x, x^* \rangle : x^* \in M_f \} \). (This implies \( f \) is convex and Lipschitz of rank \( K \) for some \( K \).) Let \( C_0 \subseteq C \) be convex. Then, the following are equivalent.

(a) The function \( f |_{C_0} \) is linear.
(b) There exists \( x_0^* \in M_f \) such that \( x_0^* \in \bigcap_{x \in C_0} \partial_K f(x) \) (which is equivalent to saying that \( f(x) = \langle x, x_0^* \rangle \) for all \( x \in C_0 \)).

**Proof.** It is easy to see that (b) implies (a). To prove that (a) implies (b), we shall prove the contrapositive. So, suppose \( \bigcap_{x \in C_0} \partial_K f(x) = \emptyset \). Then, there exist \( x_1, \ldots, x_n \in C_0 \) such that \( \bigcap_{i=1}^n \partial_K f(x_i) = \emptyset \). Let \( \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \).

Then, for all \( x^* \in M_f \) we have

- \( \langle x_i, x^* \rangle \leq \langle x_i, x_i^* \rangle = f(x_i) \) for all \( i = 1, \ldots, n \), and
- \( \langle x_i, x^* \rangle < \langle x_i, x_i^* \rangle = f(x_i) \) for some \( i \in \{1, \ldots, n\} \)

This implies \( \frac{1}{n} \sum_{i=1}^n \langle x_i, x^* \rangle = \langle \bar{x}, x^* \rangle < \frac{1}{n} \sum_{i=1}^n f(x_i) \). Since this is true for all \( x^* \in M_f \), and because \( M_f \) is compact, it follows that \( f(\bar{x}) = \max \{ \langle \bar{x}, x^* \rangle : x^* \in M_f \} < \frac{1}{n} \sum_{i=1}^n f(x_i) \), which proves that \( f \) is not linear on \( C_0 \), as claimed.

\[ \square \]

**G. Minimal RICs**

Let \( \hat{\Omega}_n \) be defined for all \( n \in \mathbb{N} \) as in Appendix A.5. Define inclusion for \( n = 0 \) as follows: for \( \omega_0, \omega'_0 \in \hat{\Omega}_0 \), \( \omega_0 \subseteq_0 \omega'_0 \) if \( (P, \hat{\omega}) \in \omega_0 \) implies \( (P, \hat{\omega}) \in \omega'_0 \).

Let us inductively define a partial order representing inclusion for all \( n \geq 0 \): for \( \omega_{n+1}, \omega'_{n+1} \in \hat{\Omega}_{n+1}, \omega_{n+1} \subseteq_{n+1} \omega'_{n+1} \) if \( (P, \omega_n) \in \omega_{n+1} \) implies there exists \( (P, \omega'_n) \in \omega'_{n+1} \) such that \( \omega_{n,s} \subseteq \omega'_{n,s} \) for all \( s \in S \).

In analogy with Lemma A.2, it can be shown that \( \subseteq_{n+1} \mid_{\hat{\Omega}_n} = \subseteq_n \). As before, then, for \( \omega, \omega' \in \hat{\Omega} \), \( \omega \subseteq^* \omega' \) if \( \omega_{n,s} \subseteq \omega'_{n,s} \) for some \( n \) with \( \omega, \omega' \in \hat{\Omega}_n \).

By definition of \( \hat{\Omega} \), there is some \( n \) such that \( \omega, \omega' \in \hat{\Omega}_n \), and because \( \subseteq_n \) extends faithfully, the precise choice of \( n \) is immaterial. Thus, \( \subseteq^* \) is a well defined partial order on \( \hat{\Omega} \). We now show that \( \subseteq^* \) has a recursive definition as well.

**Proposition G.1.** For any \( \omega, \omega' \in \hat{\Omega} \), the following are equivalent.

(a) \( \omega \subseteq^* \omega' \).
(b) for all \( (P, \hat{\omega}) \in \omega \), there exists \( (P, \hat{\omega}') \in \omega' \) such that \( \hat{\omega}_s \subseteq^* \hat{\omega}'_s \) for all \( s \in S \).

Therefore, \( \subseteq^* \) is the *unique* partial order for inclusion on \( \hat{\Omega} \) defined as \( \omega \subseteq^* \omega' \) if (b) holds.

The *proof* of Proposition G.1 is analogous to the proof of Proposition A.3, and so is omitted.

Finally, just as in Proposition A.4, \( \subseteq^* \) has a unique continuous extension to \( \Omega \). Thus, \( \subseteq^* \) is the unique partial order on \( \Omega \) that signifies inclusion. Moreover, for \( \omega, \omega' \in \Omega \), let \( \omega \cap^* \omega' \) represent the \( \subseteq^* \)-greatest lower bound of both \( \omega \) and \( \omega' \). Naturally, \( \cap^* \) then represents recursive set inclusion.

For \( \omega, \omega' \in \Omega \), let \( \omega_n := \text{proj}_n \omega \) and \( \omega'_n := \text{proj}_n \omega' \). The following is an easy corollary.

**Corollary G.2.** For \( \omega, \omega' \in \Omega, \omega := \omega \cap^* \omega' \) if, and only if, \( \omega_n := \text{proj}_n \omega = \omega_n \cap^* \omega'_n \) for all \( n \in \mathbb{N} \).
Proof. The ‘only if’ part is straightforward. The ‘if’ part follows from the continuity of $\mathcal{C}^*$. □

Let $\approx$ denote the symmetric part of $\geq$, the recursive Blackwell order, and note that $\approx$ is transitive. Then, $\omega \approx \omega'$ if, and only if, $\omega$ and $\omega'$ recursively Blackwell dominate each other. Similarly, for $\omega_n, \omega_n' \in \hat{\Omega}_n$, define $\tilde{\omega}_n := \omega_n \cap^* \omega_n'$, where $\cap^*$ is recursive set inclusion as described above.

**Lemma G.3.** Let $\omega_0, \omega_0' \in \hat{\Omega}_0$ such that $\omega_0 \approx \omega_0'$. Then, $\tilde{\omega}_0 \approx \omega_0$.

Proof. It is easy to see that $\tilde{\omega}_0 \subset^* \omega_0$, and so $\omega_0 \gtrsim \tilde{\omega}_0$ (and similarly for $\omega_0'$). All that remains is to show that $\tilde{\omega}_0 \gtrsim \omega_0$.

Towards this end, let $(P(0), \hat{\omega}) \in \omega_0$, and suppose $(P(0), \hat{\omega}) \notin \tilde{\omega}_0$. Then, because $\omega_0' \gtrsim \omega_0$, there exists $(P(1), \hat{\omega}) \in \omega_0'$ such that $P(1)$ is (strictly) finer than $P(0)$. But now because $\omega_0 \gtrsim \omega_0'$, either $(P(1), \hat{\omega}) \in \omega_0$ and hence $\tilde{\omega}_0$, in which case we are done, or there exists $(P(2), \hat{\omega}) \in \omega_0$ where $P(2)$ is strictly finer than $P(1)$. Continuing in this fashion, we get a sequence $(P(j))$ of strictly finer partitions, where the even members belong to $\omega_0$ (in the obvious sense) and the odd members belong to $\omega_0'$. But this sequence is finite, and so the final member must belong to both $\omega_0$ as well as $\omega_0'$, otherwise we would contradict the assumption that $\omega_0 \approx \omega_0'$. Let $P(n)$ be this final member of the sequence. Then, $(P(n), \hat{\omega}) \in \tilde{\omega}_0$, so that $\tilde{\omega}_0 \gtrsim \omega$, which proves the claim. □

A similar result holds for all $\hat{\Omega}_n$.

**Lemma G.4.** Let $\omega_n, \omega_n' \in \hat{\Omega}_n$. Then, for all $n \geq 0$, $\omega_n \approx \omega_n'$ implies $\tilde{\omega}_n \approx \omega_n$.

Proof. It is easy to see that $\tilde{\omega}_n \subset^* \omega_n$, and so $\omega_n \gtrsim \tilde{\omega}_n$ (and similarly for $\omega_n'$). All that remains is to show that $\tilde{\omega}_n \gtrsim \omega_n$. We shall establish the proof by induction. $\omega_n \approx \omega_n'$, and that the result is true for $n - 1$.

Let $(P(0), \omega_n^{(0)} ) \in \omega_n$, and suppose $(P(0), \cdot ) \notin \tilde{\omega}_n$. Then, there exists $(P(1), \omega_n^{(1)} ) \in \omega_n'$ such that $P(1)$ is finer than $P(0)$ and $\omega_{n-1,s}^{(1)} \gtrsim \omega_{n-1,s}^{(0)}$. Continuing just as we did in Lemma G.3, we note that there exists a sequence $(P(j), \omega_n^{(j)} )$ where $P(j)$ is strictly finer than $P(j-1)$ and $\omega_n^{(j)} \gtrsim \omega_n^{(j-1)}$ for all $s \in S$, and where the even members belong to $\omega_n$ (in the obvious sense) and the odd members belong to $\omega_n'$. But this sequence is finite, and there must be eventual members of this sequence where $(P(m), \omega_n^{(m)} ) \in \omega_n$ and $(P(m), \omega_n^{(m+1)} ) \in \omega_n'$, and $\omega_n^{(m)} \approx \omega_n^{(m+1)}$ for all $s \in S$, because by hypothesis, $\omega_n \approx \omega_n'$. Moreover, we must also have that $P(m)$ is strictly finer than $P(0)$ and $\omega_n^{(m)} \gtrsim \omega_n^{(0)}$.

Then, $(P(m), \omega_n^{(m)} ) \in \tilde{\omega}_n$, where $\omega_n^{(m)} := \omega_{n-1,s}^{(m)} \cap \omega_n^{(m+1)}$. But, by the induction hypothesis, $\omega_n^{(m)} \approx \omega_n^{(m+1)}$. This implies, $\tilde{\omega}_n \gtrsim \omega_n$, as claimed. □

We can now show that the recursive intersection of two recursively Blackwell equivalent classes is also in the same equivalence class.

**Proposition G.5.** For $\omega, \omega' \in \Omega$, $\omega \approx \omega'$ implies $\omega \cap^* \omega =: \tilde{\omega} \approx \omega$.

Proof. As in Appendix A.5, let $\omega_n := \text{proj}_n \omega$ and $\omega'_n := \text{proj}_n \omega'$. By Corollary A.5, $\omega_n \approx \omega'_n$ for all $n \geq 0$. Corollary G.2 implies that $\tilde{\omega}_n := \omega_n \cap^* \omega'_n$, and Lemma G.4 implies $\tilde{\omega}_n \approx \omega_n$ for all $n \geq 0$. Corollary A.5 now implies $\tilde{\omega} \approx \omega$, as claimed. □
Let \([\omega] := \{\omega' : \omega' \sim \omega\}\) denote the \(\sim\)-equivalence class of \(\omega\). Note that \([\omega]\) is a closed (and hence compact) subset of \(\Omega\) because \(\preceq\) is continuous. For each \(\omega' \in [\omega]\), define the set \(\mathcal{D}(\omega') := \{\tilde{\omega} : \tilde{\omega} \preceq^* \omega'\}\). We are now ready to prove the existence of \(\preceq^*\)-minimal rics.

**Proposition G.6.** Each \([\omega]\) has a unique \(\preceq^*\)-minimal element given by

\[
\bigcap_{\omega' \in [\omega]} \mathcal{D}(\omega')
\]

**Proof.** Recall that, by construction, \(\preceq^*\) is a continuous partial order. Therefore, the set \(\mathcal{D}(\omega') := \{\tilde{\omega} : \tilde{\omega} \preceq^* \omega'\}\) is closed for each \(\omega' \in [\omega]\). Moreover, for any finite collection \(\omega^1, \ldots, \omega^m \in [\omega]\), the intersection \(\bigcap_{i=1}^m \mathcal{D}(\omega^i)\) is non-empty by Proposition G.5. Thus, the collection of closed sets \((\mathcal{D}(\omega'))_{\omega' \in [\omega]}\) has the finite intersection property. Because \(\Omega\) is compact, the intersection

\[
\bigcap_{\omega' \in [\omega]} \mathcal{D}(\omega')
\]

is non-empty. By Proposition G.5, this intersection must have a unique element, which proves the claim. \(\square\)

**H. A Metric on the Space of Partitions**

In this section, we define a natural metric on the space of partitions that is related to the informational content of the partitions. The metric we introduce is fairly standard. However, we have been unable to find a formulation suitable for our purposes, so we prove that our proposed metric is indeed a metric. It is also worth noting that all the results in this section remain valid if the state space \(S\) is an arbitrary countable set, \(\mu\) is a countably additive measure on \(S\), and \(\mathcal{P}\) represents the space of all partitions of \(S\) with countably many (measurable) cells.

Let \(S\) be a finite set, and \(\mathcal{P}\) be the space of all partitions of \(S\). Let \(\mu\) be a probability measure on \(S\). Define the entropy of the partition \(P \in \mathcal{P}\) as

\[
H(P) := -\sum_{J \in P} \mu(J) \log \mu(J)
\]

Let \(\succeq\) be a partial order on \(\mathcal{P}\), wherein \(P \succeq Q\) if \(P\) is coarser than \(Q\) (or equivalently, \(Q\) is finer than \(P\)). We shall say that \(P > Q\) if \(P \succeq Q\) and \(P \neq Q\).

We may also define the coarsest refinement of \(P\) and \(Q\), denoted by \(P \wedge Q\). If \(P = (I_m)\) and \(Q = (J_n)\), then \(P \wedge Q = (I_m \cap J_n)_{m,n}\), so

\[
H(P \wedge Q) = -\sum_m \sum_n \mu(I_m \cap J_n) \log (\mu(I_m \cap J_n))
\]

Similarly, \(P \vee Q\) is the finest partition coarser than \(P\) and \(Q\). Then, \((\mathcal{P}, \succeq, \vee, \wedge)\) is a lattice, with greatest (coarsest) element \(\{S\}\), and least (finest) element \(\{\{s\} : s \in S\}\). Notice that \(H(\{S\}) = 0\), while \(H(P) > 0\) for all other partitions \(P\). Define the conditional entropy \(H(P \mid Q)\) as

\[
H(P \mid Q) := H(P \wedge Q) - H(Q)
\]
It is easy to see that

\[ H(P \mid Q) = -\sum_n \mu(J_n) \sum_m \frac{\mu(I_m \cap J_n)}{\mu(J_n)} \log \left( \frac{\mu(I_m \cap J_n)}{\mu(J_n)} \right) \]

We now come to the main result of this section.

**Proposition H.1.** The function

\[ d(P, Q) := 2H(P \wedge Q) - H(P) - H(Q) = H(P \mid Q) + H(Q \mid P) \]

is a metric on \( \mathcal{P} \).

We begin with some lemmata.

**Lemma H.2.** \( H \) is anti-monotone, ie, \( P \geq Q \) implies \( H(Q) \geq H(P) \). Moreover, \( H \) is strictly anti-monotone, ie, \( P > Q \) implies \( H(Q) > H(P) \).

The proof is trivial and is omitted.

**Lemma H.3.** The function \( H(P \mid Q) \) is anti-monotone in \( P \), and is monotone in \( Q \).

*Proof.* Notice that if \( P' \geq P \), then \( P' \wedge Q \geq P \wedge Q \), so the anti-monotonicity of \( H \) implies that \( H(P \mid Q) \) is anti-monotone in \( P \). We say that \( Q \) is an elementary refinement of \( Q' \) if \( Q' = \{J_1, \ldots, J_N\} \) and \( Q = \{\hat{J}_1, \ldots, \hat{J}_{n-1}, \hat{J}_N, \hat{J}_{N+1}\} \), where \( \hat{J}_n := J_n \) for all \( n = 1, \ldots, N - 1 \), while \( J_N = \hat{J}_N \cup \hat{J}_{N+1} \). In other words, \( Q \) and \( Q' \) are identical except that there exists a cell \( J_N \in Q' \) that is the union of exactly two cells in \( Q \).

Let \( Q' \geq Q \). Then, there exist \( Q_1, \ldots, Q_k \in \mathcal{P} \) such that \( Q' = Q_k \geq Q_{k-1} \geq \cdots \geq Q_1 = Q \), and where \( Q_i \) is an elementary refinement of \( Q_{i+1} \). Thus, in order to show that \( H(P \mid Q) \) is monotone in \( Q \), it suffices to consider \( Q \) and \( Q' \) where \( Q \) is an elementary refinement of \( Q' \).

Let \( P = \{I_1, \ldots, I_M\} \) and \( Q \) and \( Q' \) be as above. In what follows, we shall let \( \eta(x) = x \log x \) for all \( x > 0 \) and \( \eta(0) = 0 \). Then \( \eta \in \mathbb{R}^{\mathbb{R}^+} \) is strictly convex and continuous on its domain. Let

\[ \Lambda = -\sum_{n=1}^{N-1} \mu(J_n) \sum_m \eta \left( \frac{\mu(I_m \cap J_n)}{\mu(J_n)} \right) \]
This allows us to write

\[
H(P \mid Q) = - \sum_n \mu(j_n) \sum_m \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \log \left( \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \right)
\]

\[
= \Lambda - \sum_{n=N,N+1} \mu(j_n) \sum_m \eta \left( \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \right)
\]

\[
= \Lambda - \mu(J_N) \sum_m \sum_{n=N,N+1} \eta \left( \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \right)
\]

\[
\leq \Lambda - \mu(J_N) \sum_m \eta \left( \sum_{n=N,N+1} \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \right)
\]

\[
= \Lambda - \mu(J_N) \sum_m \eta \left( \frac{\mu(I_m \cap j_n)}{\mu(j_n)} \right)
\]

\[
= H(P \mid Q')
\]

where we have used the fact that \(-\eta\) is concave to establish the inequality.

\[\blacksquare\]

**Lemma H.4.** The function \(H\) is submodular, ie, \(H(P \land Q) + H(P \lor Q) \leq H(P) + H(Q)\).

**Proof.** Fix \(P\) and \(Q\), and let \(Q \leq Q'\). We shall use the fact that the function \(H(P \mid Q)\) is anti-monotone in \(P\) and monotone in \(Q\). Then, \(H(P \land Q) - H(Q) = H(P \mid Q) \leq H(P \mid Q') = H(P \land Q') - H(Q')\). Now set, \(Q' := P \lor Q\), so that \(P \land (P \lor Q) = P\), which implies \(H(P \land Q) - H(Q) \leq H(P) - H(P \lor Q)\). Therefore, \(H\) is submodular. \[\blacksquare\]

We now list some properties of the lattice \((\mathcal{P}, \geq, \lor, \land)\).

**Lemma H.5.** For \(P, Q, R \in \mathcal{P}\), the following hold:

(a) \(R \geq (P \land R) \lor (Q \land R)\).

(b) \((P \lor Q) \land R \geq (P \land R) \lor (Q \land R)\).

**Proof.** Note that \(R \geq P \land R\) and \(R \geq Q \land R\), so \(R \geq (P \land R) \lor (Q \land R)\), which establishes (a).

To see (b), note that \(P \geq P \land R\), while \(Q \geq Q \land R\). Therefore, \(P \lor Q \geq (P \land R) \lor (Q \land R)\). But we also have that \(R \geq (P \land R) \lor (Q \land R)\), from (a). The definition of \(\land\) then implies that \((P \lor Q) \land R \geq (P \land R) \lor (Q \land R)\), as required. \[\blacksquare\]

**Proof of Proposition H.1.** The proof relies on the fact that conditional entropy \(H(P \mid Q)\) is anti-monotone (Lemmas H.2 and H.3) and submodular (Lemma H.4). Because \(H\) is anti-monotone (Lemma H.2), \(d(P, Q) \geq 0\) for all \(P, Q\). We have already established that \(P < Q\) implies \(H(P) > H(Q)\). If \(P\) and \(Q\) are distinct, then \(P \land Q\) is distinct from either \(P\) or \(Q\), so that \(d(P, Q) > 0\).

It is easy to see that \(d(P, Q) \leq d(P, R) + d(R, Q)\) if, and only if,

\[\bigcirc\]

\[H(P \land Q) + H(R) \leq H(P \land R) + H(Q \land R)\]

By lemma H.5, we see that \(R \geq (P \land R) \lor (Q \land R)\) and \((P \lor Q) \land R \geq (P \land R) \lor (Q \land R)\). Set \(P' = P \land R\) and \(Q' = Q \land R\). The submodularity of \(H\) implies \(H(P' \lor Q) + H(P' \land Q') \leq \]
Lemma I.3. By the Mixture Space Theorem, applying L-Independence, we find Corollary E.3. We now show that for any for a formal definition.) By the definition of the product metric (see Appendix A.2), this means that all for length `n` is dense in 0 1. In other words, each L(n) is naturally embedded in L0.

**Proposition I.1.** The space L0 is dense in L.

**Proof.** Because probability measures on C with finite support are dense in Δ(C), it follows that for all n ≥ 1, L0(n) is dense in L(n). (The metrics defined on L(n) make this clear — see Appendix A.2 for a formal definition.) By the definition of the product metric (see Appendix A.2), this means that for any ℓ ∈ L and ε > 0, there exists an n and an ℓ(n) ∈ L(n) such that d(ℓ, ℓ(n) ◦ ℓ†) < ε, where ℓ(n) ◦ ℓ† is the concatenation of ℓ† to ℓ(n). This completes the proof.

It follows immediately from Lipschitz Continuity (Axiom 1(c)) that ≻ |L is non-trivial, see Corollary E.3. We now show that ≻ |s (as defined in Section 3.1) is also non-trivial for each s ∈ S.

**Lemma I.2.** Let ℓ0, ℓ1 ∈ L. Then, ℓ0(s) ≈ s ℓ1(s) for all s ∈ S implies ℓ0 ≈ |L ℓ1.

**Proof.** By definition of ≻ |s, ℓ0(s) ≈ s ℓ1(s) if, and only if, ℓ0 ⊕ (1,S\s) ℓ∗ ∝ |L ℓ1 ⊕ (1,S\s) ℓ∗. Repeatedly applying L-Independence, we find

\[
\frac{1}{n} \ell^0 + \frac{n - 1}{n} \ell_\ast = \frac{1}{n} \sum_{s \in S} \ell^0 \oplus (1,S\setminus s) \ell_\ast \sim |L \frac{1}{n} \sum_{s \in S} \ell^1 \oplus (1,S\setminus s) \ell_\ast = \frac{1}{n} \ell^1 + \frac{n - 1}{n} \ell_\ast
\]

By L-Independence, we find ℓ0 ∼ |L ℓ1. (More precisely, this follows immediately once we note that, by the Mixture Space Theorem, ≻ |L has an affine representation.)

**Lemma I.3.** There exists s ∈ S such that ℓ*(s) ≈ s ℓ*(s). For all s ∈ S, there exists s′ ∈ S such that (c, ℓ* ⊕ (1,S\s) ℓ*) ∝ s (c, ℓ*).

I. Consumption Streams and the RAA Representation

To see that L ∼ F(Δ(C × L)), note that we can define L(1) := F(Δ(C)) and then recursively define L(n) := F(Δ(C × L(n−1))) as the space of consumption streams of length n. Just as with the definition of the space of racps X in Appendix A.2, we say that L is the space of all consistent sequences in ∞

The support of a consumption stream ℓ ∈ L is a set F ⊂ C such that at any date and in any state, the realized consumption lies in F. A consumption stream has finite support if its support is finite. For any finite set F ⊂ C, we can define LF as the space of all consumption streams with prizes in F. Formally, LF ∼ F(Δ(C × LF)). Let L0 be the space of all consumption streams with finite support. That is, L0 := {L_F : F ⊂ C, F finite}.

Recall the consumption stream ℓ† ∈ L which delivers c†(s) in state s at every date. Clearly, the support of ℓ† is finite. Analogous to L0, we can define L0(n) as the space of consumption streams of length n with finite support. For any ℓ(n) ∈ L0(n), ℓ(n) ◦ ℓ† ∈ L0, where ℓ(n) ◦ ℓ† is the concatenation of ℓ† to ℓ(n). In other words, each L(n) is naturally embedded in L0.
Proof. Corollary E.3 says that $\ell^* > \mid_L \ell_*$. Therefore, by (the contrapositive to) Lemma I.2, there must exists an $s$ such that $\ell^*(s) \sim_s \ell_*(s)$. In particular, this means that $\ell^* \oplus (1, S \setminus s) \ell_* \sim \mid_L \ell_*$.

To see the second part, let us suppose by way of contradiction that for all $s' \in S$, $(c, \ell^* \oplus (1, S \setminus s') \ell_*) \sim_{s'} (c, \ell_*)$. Now, set $\ell^0, \ell^1$ such that $\ell^0(s') = (c, \ell^* \oplus (1, S \setminus s') \ell_*)$, while $\ell^1(s') = (c, \ell_*)$. It follows from Lemma I.2 that $(c, \ell^* \oplus (1, S \setminus s') \ell_*) \sim \mid_L (c, \ell_*)$.

Now, L-Stationarity (Axiom 2) and the fact that $\ell^* \oplus (1, S \setminus s') \ell_* \sim \mid_L \ell_*$ imply $(c, \ell^* \oplus (1, S \setminus s') \ell_*) \sim \mid_L (c, \ell_*)$, which yields the desired contradiction. \qed

**Proposition I.4.** For all $s \in S$, $\succsim_s$ is non-trivial.

**Proof.** Lemma I.3 and (the contrapositive to) L-History Independence (Axiom 2) imply $(c, \ell^* \oplus (1, S \setminus s') \ell_*) \sim_{s''} (c, \ell_*)$ for all $s'' \in S$, as claimed. \qed

**Proposition I.5.** The preference $\succsim \mid_L$ on $L$ has a standard raa representation. Moreover, $\Pi$ and $\delta$ are unique and the collection $(u_s)_{s \in S}$ is unique up to a common positive scaling.

As described in Section 3.1, for each $s \in S$, $\succsim_s$ is an induced preference over $\Delta(C \times L)$. Let $\succsim^C_s$ denote the induced preference over $\Delta(C)$ in state $s$. It is clear that $\succsim^C_s$ is well defined, continuous on $\Delta(C)$, and satisfies Independence. These properties imply there exist $\succsim^C_s$-maximal and -minimal lotteries that are degenerate; denote them by $c^*(s)$ and $c_*(s)$. Let $F_0$ be the finite set of consumption defined as

$$F_0 := \{c_*(s), c^<(s), c^>(s) : s \in S\}$$

Recall that for any finite set $B \subset C$, $L_B$ denotes the space of consumption streams where the consumption in all periods and states is entirely within $B$. Then, $L_B \simeq \mathcal{F}(\Delta(B \times L_B))$.

**Lemma I.6.** For any finite set $B \subset C$, the induced preference $\succsim \mid_{L_B}$ satisfies the Axioms stated in Corollary 5 of Krishna and Sadowski (2014, henceforth KS).

**Proof.** It follows from Proposition I.4 that each $\succsim_s$ is non-trivial. That is, $\succsim \mid_L$ is state-wise nontrivial. In addition, $\succsim \mid_L$ is continuous, satisfies Independence, and is separable in $\ell_1$ and $\ell_2$, thereby satisfying Axioms 2, 3, and 5 in KS. Axioms 6, 7, and 9 in KS correspond to properties (c), (d), and (b) of $L$-Properties (Axiom 2).

We now proceed to the proof of Proposition I.5.

**Proof of Proposition I.5.** Let $B \subset C$ be finite. By Lemma I.6, $\succsim \mid_{L_B}$ satisfies the Axioms in Corollary 5 of KS. This implies there exists a tuple $((u_s^B)_{s \in S}, \Pi^B, \delta^B)$ that is an raa representation of $\succsim \mid_{L_B}$. If $F_0 \subset B$, then we may assume, without loss of generality, that $u_s^B(c^>(s)) = 0$ for all $s \in S$. Then, Corollary 5 in KS says that the collection of utilities $(u_s^B)$ is uniquely identified up to a joint scaling, and that $\Pi^B$ and $\delta^B$ are also uniquely determined.

Now, consider any other finite set $D$ such that $F_0 \subset B \subset D$. By Lemma I.6, $\succsim \mid_{L_D}$ also has an raa representation $((u_s^D)_{s \in S}, \Pi^D, \delta^D)$. As before, if we set $u_s^D(c^>(s)) = 0$ for all $s \in S$, then the collection of utilities $(u_s^D)$ is identified up to a common scaling. Now, because $B \subset D$, we have $L_B \subset L_D$. Therefore, the raa representation $((u_s^D)_{s \in S}, \Pi^D, \delta^D)$ of $\succsim \mid_{L_D}$ when restricted to $L_B$, is also a representation of $\succsim \mid_{L_B}$. And this representation has the feature that $u_s^D(c^>(s)) = 0$ for all
$s \in S$. Once again, the uniqueness of the raa representation implies that a single joint scaling of the collection $(u^B_s)$ results in $u^B_s = u^D_s$ for all $s \in S$, $\Pi^B = \Pi^D$, and $\delta^B = \delta^D$.

Recall that $c^*(s) \succeq_C \alpha \succeq_C c_\alpha(s)$ for all $\alpha \in \Delta(C)$. Because $u^B_s$ and $u^D_s$ represent, respectively, $\succeq_C \Delta(B)$ and $\succeq_C \Delta(D)$, it must be that $\lambda^*(s) := u^j_s(c^*(s))$ and $\lambda_*(s) := u^j_s(c_*(s))$ for $j = B, D$. Since $B$ and $D$ are arbitrary, it follows that it holds for all finite $B$ that contains $F_0$. In other words, the Markov transition operator $\Pi$ has been identified uniquely, as has the discount factor $\delta \in (0, 1)$.

Let $u_s \in C(C)$ be a vN-M utility representation of $\succeq_C$ such that $u_s(c^*(s)) = 0$. Both $u_s|_{\Delta(B)}$ and $u^B_s$ are vN-M representations of $\succeq_C |_{\Delta(B)}$ and by the Mixture Space Theorem, differ at most by a positive affine transformation. Therefore, if we scale $u_s$ so that $u_s(c^*(s)) = \lambda^*(s)$, we must necessarily have $u_s(c_*(s)) = \lambda_*(s)$ for all $s \in S$.

Consider, now, the tuple ($(u_s, \Pi, \delta)$, and the functional $W_0 : L \to \mathbb{R}$ defined as $W_0(\ell) := \sum_s \pi_0(s)W(\ell, s)$, where

$$W(\ell, s) := \sum_{s'} \Pi(s, s')[u_s'(\ell_1(s')) + \delta W(\ell_2(s'), s')]$$

It is easy to see that the function $W_0 \overset{(58)}{\text{is uniquely determined by the tuple }} ((u_s)_{s \in S}, \delta, \Pi)$. As established above, $W_0$ represents $\succeq |_{L_B}$ for every finite $B$. In other words, $W_0$ represents $\succeq |_{L_0}$. Proposition I.1 says that $L_0$ is dense in $L$, and because $W_0$ is (uniformly) continuous, it also represents $\succeq$ on $L$. The uniqueness of the raa representation of $\succeq |_{L}$ (given our normalizations) follows immediately, which concludes the proof.

\section*{J. The Space of Nice Menus}

\textbf{Lemma J.1.} Let $\{f_1, \ldots, f_m\}$ be the generator set for some finite $x$. Then, $x \sim \{f_1, \ldots, f_m\}$.

\textit{Proof.} Notice that

$$x \succeq \{f_1, \ldots, f_m\} \quad \text{by Monotonicity (Axiom 1(d))}$$

$$\succeq \mathcal{F}(\xi_x) \quad \text{by IICC and Continuity}$$

$$\sim x \quad \text{by Axiom IICC}$$

for some $\xi_x \in \mathcal{E}_x$, which establishes the claim. \hfill $\square$

\textbf{Lemma J.2.} $y \sim y \cup \{\ell_*\}$ for all finite $y \in X$.

\textit{Proof.} By Monotonicity (Axiom 1(d)), $y \cup \{\ell_*\} \succeq y$. To see that $y \succeq y \cup \{\ell_*\}$, let $(f_n) \in \mathcal{F}^\infty$ be such that $f_n \to f$ but $f_n \not\in y$ for all $n \in \mathbb{N}$. Then, by IICC (Axiom 4), $y \cup \{f_n\} \succeq y \cup \{\ell_*\}$. By Continuity (Axiom 1(b)), $y = \lim_{n \to \infty} y \cup \{f_n\} \succeq y \cup \{\ell_*\}$, as claimed. \hfill $\square$

\textbf{Lemma J.3.} If $y > x$ then $y > \frac{1}{2}y + \frac{1}{2}x$.

\textit{Proof.} As always, $W_0$ also denotes the linear extension of $W_0$ to $\Delta(L)$.
Proof. By Aversion to Randomization (Axiom 1(e)), the lower contour sets are convex, which implies \( y \gtrsim \frac{1}{2} y + \frac{1}{2} x \). To show there cannot be indifference, we consider two cases.

Case 1: There exists \( \ell_{y} \in L \) such that \( \ell_{y} \sim y \). In this case, by L-Independence (Axiom 2(a)) and Continuity (Axiom 1(b)), there exists \( \ell_{x} \in L \) such that \( \ell_{x} \sim x \). Repeated use of L-Independence gives us \( y \sim \ell_{y} \sim \frac{1}{2} \ell_{y} + \frac{1}{2} y \sim \frac{1}{2} x \sim \frac{1}{2} \ell_{x} + \frac{1}{2} y \sim \frac{1}{2} \ell_{y} + \frac{1}{2} x \). Then, proving that \( \frac{1}{2} y + \frac{1}{2} \ell_{x} \gtrsim \frac{1}{2} y + \frac{1}{2} x \) will establish our claim.

To see that this is indeed the case, recall that \( V \) represents \( \gtrsim \) and \( V \) itself has the representation \([\bullet] \). Then, \( V \) is convex as well as \( L \)-affine, which implies that

\[
V(\frac{1}{2} y + \frac{1}{2} \ell_{x}) = \frac{1}{2} V(y) + \frac{1}{2} V(\ell_{x})
\]

\[
= \frac{1}{2} V(y) + \frac{1}{2} V(x)
\]

\[
\gtrsim V(\frac{1}{2} y + \frac{1}{2} x)
\]

which establishes \( y \gtrsim \frac{1}{2} y + \frac{1}{2} x \), as claimed.

Case 2: There does not exist \( \ell_{y} \in L \) such that \( \ell_{y} \sim y \). Then, by Lemma J.2 and Monotonicity (Axiom 1(d)), \( y \succ \ell_{*} \). Therefore, there exists \( a \in (0, 1) \) sufficiently small and \( \ell_{y_{a}} \in L \) such that \( \ell_{y_{a}} \sim y_{a} \), where \( y_{a} := ay + (1 - a) \ell_{*} \). Now let \( x_{a} := ax + (1 - a) \ell_{*} \), so that by L-Independence, \( y_{a} \succ x_{a} \). We are now back in Case 1 above, which establishes that \( y_{a} \gtrsim \frac{1}{2} y_{a} + \frac{1}{2} x_{a} \). We may rewrite this as

\[
ay + (1 - a) \ell_{*} \gtrsim \frac{1}{2} (ay + (1 - a) \ell_{*}) + \frac{1}{2} (ax + (1 - a) \ell_{*})
\]

\[
= a(\frac{1}{2} x + \frac{1}{2} y) + (1 - a) \ell_{*}
\]

Because \( V \) represents \( \gtrsim \) and \( V \) is \( L \)-affine, this can be rewritten as

\[
aV(y) + (1 - a)V(\ell_{*}) > aV(\frac{1}{2} x + \frac{1}{2} y) + (1 - a)V(\ell_{*})
\]

This implies \( V(y) > V(\frac{1}{2} x + \frac{1}{2} y) \), as desired. \( \square \)

A menu \( \{f_{1}, \ldots, f_{m}\} \) is minimal if \( \{f_{1}, \ldots, f_{m}\} \succ \{f_{1}, \ldots, f_{m}\} \setminus \{f_{i}\} \) for all \( i \in \{1, \ldots, m\} \).

Lemma J.4. Let \( x \in X \) be finite and suppose \( x \succ \ell_{*} \). Also suppose there exists a minimal menu \( \{f_{1}, \ldots, f_{m}\} \) such that \( x \sim \{f_{1}, \ldots, f_{m}\} \subset x \). Then, \( \{f_{1}, \ldots, f_{m}\} \succ \{(1 - \varepsilon) f_{1} + \varepsilon \ell_{*}, \ldots, f_{m}\} \) for all \( \varepsilon \in (0, 1] \).

Proof. Because \( \{f_{1}, \ldots, f_{m}\} \) is minimal, \( \{f_{1}, \ldots, f_{m}\} \succ \{f_{2}, \ldots, f_{m}\} \), and so by Lemma J.2, \( \{f_{1}, \ldots, f_{m}\} \succ \{\ell_{*}, f_{2}, \ldots, f_{m}\} \). (This proves the case where \( \varepsilon = 1 \).) By Lemma J.3, \( \{f_{1}, \ldots, f_{m}\} \succ (1 - \varepsilon) \{f_{1}, \ldots, f_{m}\} + \varepsilon \{\ell_{*}, f_{2}, \ldots, f_{m}\} \). Note that \( (1 - \varepsilon) \{f_{1}, \ldots, f_{m}\} \subset (1 - \varepsilon) \{f_{1}, \ldots, f_{m}\} + \varepsilon \{\ell_{*}, f_{2}, \ldots, f_{m}\} \), so Monotonicity (Axiom 1(d)) implies \( \{f_{1}, \ldots, f_{m}\} \succ \{(1 - \varepsilon) f_{1} + \varepsilon \ell_{*}, \ldots, f_{m}\} \), which concludes the proof. \( \square \)

Proposition J.5. The space \( X_{0} \) of nice menus is dense in \( X \).

Proof. It is easy to see that the space of finite menus that are strictly preferred to \( \ell_{*} \) is dense in \( X \). (This is because \( \ell^{*} \succ \ell_{*} \), and \( \supseteq \) satisfies L-Independence (Axiom 2(a))). Therefore, it will suffice to
show that $X_0$ is dense in the space of such finite menus. Towards that end, let $x$ be a finite menu and let $x_\varepsilon := (1 - \varepsilon)x + \varepsilon x_*$. Because $x > x_*$, it follows from Lemma J.3 that $x > x_\varepsilon$ for all $\varepsilon \in [0, 1)$.

Let $\xi_1, \ldots, \xi_m \in \mathcal{X}_x$ such that $\mathcal{F}(\xi_k) \sim x$ for all $k = 1, \ldots, m$. Suppose $\xi_1$ is such that $\mathcal{F}(\xi_1)$ is minimal. (It is easy to see that such a $\xi \in \mathcal{X}_x$ always exists.) Let $y := \mathcal{F}(\xi_1) \cup x_\varepsilon$. Then, $y \succeq x \sim \mathcal{F}(\xi_1)$ by Monotonicity (Axiom 1(d)).

We claim that $y \sim x$. To see this, let $\xi_y \in \mathcal{X}_y$ be such that $\mathcal{F}(\xi_y) \sim y$. Then, $\mathcal{F}(\xi_y) := \{f_1, \ldots, f_j, (1 - \varepsilon)f_{j+1} + \varepsilon x_*, \ldots, (1 - \varepsilon)f_m + \varepsilon x_*\}$, where $f_1, \ldots, f_j \in \mathcal{F}(\xi_1)$ and $(1 - \varepsilon)f_{j+1} + \varepsilon x_* \ldots, (1 - \varepsilon)f_m + \varepsilon x_* \in x_\varepsilon$.

Notice that $x \supset \{f_1, \ldots, f_m\}$, so that Monotonicity implies $x \sim \{f_1, \ldots, f_m\}$. If $m > j$, then by repeated application of Lemma J.4, we find that $x \sim \{f_1, \ldots, f_j, (1 - \varepsilon)f_{j+1} + \varepsilon x_* \ldots, (1 - \varepsilon)f_m + \varepsilon x_*\} = \mathcal{F}(\xi_y) \sim y$, which contradicts our assertion that $y \succeq x$.

Therefore, it must be that $m = j$, which implies that $\xi_y = \xi_1$, ie, $y \sim x$, as claimed. But $y$ is nice by construction, and $y \rightarrow x$ as $\varepsilon \rightarrow 0$. This proves that the space of nice menus is dense in the space of finite menus strictly better that $x_*$, which completes the proof.

\[ \square \]

K. Partitional Representation — Proof Details

In this section, prove Proposition C.2 in Appendix C.1 of Dillenberger, Krishna, and Sadowski (2016b).

A (static) \textit{strategy} for an agent at a menu $x$ given $\mu \in \mathcal{M}$ is a mapping $\xi^\mu_k : \mathcal{U} \rightarrow x$. The strategy $\xi^\mu_k$ is \textit{partitional} if there is a finite partition $(E_i)$ of $\mathcal{U}$ if for each $E_i$, there exists $f_i \in x$ such that $\xi^\mu_k(E_i) = f_i$. The value of this strategy is simply

$$V(x, \mu, \xi^\mu_k) = \sum_i \int_{E_i} \sum_s p_s u_s(f_i(s)) \, d\mu(p, u)$$

A strategy $\xi^\mu_k$ is \textit{optimal} at $x$ if there is no other strategy that gives a higher payoff. A partitional optimal strategy $\xi^\mu_k$ is an optimal strategy that is partitional, ie, one where

$$V(x, \mu, \xi^\mu_k) = \sum_i \int_{E_i} \langle (p, u), f_i \rangle \, d\mu(p, u)$$

$$\geq \max_{\mu \in \mathcal{M}} \left[ \int \max_{f \in x} \langle (p, u), f \rangle \, d\mu(p, u) \right]$$

where $\langle (p, u), f \rangle = \sum_s p_s u_s(f(s))$. Notice that if a partitional strategy $\xi^\mu_k$ is optimal at $x$ and if $f_i$ is the act chosen in the cell $E_i$, we must necessarily have for all $(p, u) \in E_i$, $\langle (p, u), f_i \rangle \geq \langle (p, u), f_j \rangle$ for all $j = 1, \ldots, m$.

In the sequel, $\xi^\mu_k$ denotes an optimal partitional strategy when one exists. It is easy to see that for a finite $x$, the optimal strategy is always partitional, though there may be many such strategies that are optimal. If $\xi^\mu_k$ induces the partition $(E_i)$, we refer to $(E_i)$ as an optimal partition for $\mu$ at $x$.

\textbf{Definition K.1.} Let $\{f_1, \ldots, f_m\}$ be a set of generators of $x$, and let $(E_i)_{i=1}^m$ be a partition of $\mathcal{U}$. Then, $(E_i)$ is a \textit{partition of $\mathcal{U}$ consistent with $\{f_1, \ldots, f_m\}$} if $(p, u) \in E_i$ implies $\langle (p, u), f_i \rangle \geq \langle (p, u), f_j \rangle$ for all $j = 1, \ldots, m$. 

19
Intuitively, a partition \((E_i)\) of \(\Omega\) is consistent with \(\{f_1, \ldots, f_m\}\) if there is some optimal \(\mu\) such that it is optimal to choose \(f_i\) when \((p, u) \in E_i\). The following lemma shows that finite menus always have consistent partitions.

**Lemma K.2.** Let \(x \in X\) be finite and suppose \(\{f_1, \ldots, f_m\}\) is a set of generators of \(x\). Then, \(\mu \in \Phi(\{f_1, \ldots, f_m\})\) implies \(\mu \in \mathcal{Y}(x)\).

**Proof.** Consider the following string of inequalities:

\[
V(x) = V(\{f_1, \ldots, f_m\}) \quad \text{because \(\{f_1, \ldots, f_m\}\) generates \(x\)} \\
= V(\{f_1, \ldots, f_m\}, \mu) \quad \text{definition of } \mu \\
\leq V(x, \mu) \quad \text{\(V(\cdot, \mu)\) is monotone} \\
\leq V(x) \quad \text{definition of } V
\]

which proves that \(\mu \in \mathcal{Y}(x)\), as claimed. \(\square\)

**Lemma K.3.** Let \(x\) be finite. For any \(\ell \in L\) and \(\varepsilon > 0\), (i) \(\mathcal{Y}(x) = \mathcal{Y}((1 - \varepsilon)x + \varepsilon \ell)\), (ii) if \(x\) is nice, then \((1 - \varepsilon)x + \varepsilon \ell\) is also nice, and (iii) if \(\mu \in \mathcal{Y}(x)\) and \((E_i)\) is an optimal partition for \(\mu\) at \(x\), then it is also an optimal partition for \(\mu\) at \((1 - \varepsilon)x + \varepsilon \ell\).

**Proof.** Let \(x\) be finite and \(\mu \in \mathcal{Y}(x)\). Then, \(V(x) = V(x, \mu) \geq V(x, \mu')\) for all \(\mu' \in \mathcal{M}\). We also have

\[
V((1 - \varepsilon)x + \varepsilon \ell, \mu) = (1 - \varepsilon)V(x, \mu) + \varepsilon V(\ell, \mu) \\
\geq (1 - \varepsilon)V(x, \mu') + \varepsilon V(\ell, \mu) \\
= (1 - \varepsilon)V(x, \mu') + \varepsilon V(\ell, \mu') \\
= V((1 - \varepsilon)x + \varepsilon \ell, \mu')
\]

where the inequality uses the fact that \(V(x, \mu) \geq V(x, \mu')\) and the second equality follows because \(V(\ell, \mu) = V(\ell, \mu')\) for all \(\mu, \mu' \in \mathcal{M}\) and \(\ell \in L\). This proves part (i). Part (ii) follows immediately from the definition.

To see part (iii), let \(\zeta^\mu_x\) be a partitional optimal strategy with optimal partition \((E_i)\). Then,

\[
V(x) = V(x, \mu, \zeta^\mu_x) = \sum_i \int_{E_i} \langle (p, u), f_i \rangle \, d\mu(p, u)
\]

For the menu \((1 - \varepsilon)x + \varepsilon \ell\), consider the strategy \(\zeta^{\mu}_{(1-\varepsilon)x+\varepsilon\ell} (E_i) = (1 - \varepsilon)f_i + \varepsilon \ell\). Then,

\[
V((1 - \varepsilon)x + \varepsilon \ell, \mu, \zeta^{\mu}_{(1-\varepsilon)x+\varepsilon\ell}) \\
= (1 - \varepsilon) \sum_i \int_{E_i} \langle (p, u), f_i \rangle \, d\mu(p, u) + \varepsilon \sum_i \int_{E_i} \langle (p, u), \ell \rangle \, d\mu(p, u) \\
= (1 - \varepsilon)V(x) + \varepsilon V(\ell) \\
\geq V((1 - \varepsilon)x + \varepsilon \ell, \mu')
\]

for all \(\mu' \in \mathcal{M}\) where the second equality follows from part (i). This proves that \(\zeta^{\mu}_{(1-\varepsilon)x+\varepsilon\ell}\) is a partitional optimal strategy at the menu \(x\) given the optimal \(\mu \in \mathcal{M}\) and completes the proof. \(\square\)
For a fixed partition \((E_i)\) of \(\Omega\), \(\mu \in \mathcal{M}\), and \(s \in S\), consider the map

\[(\mu, E_i, s) \mapsto \int_{E_i} p_s u_s(\cdot) \, d\mu(p, u)\]

Each tuple \((\mu, E_i, s)\) induces a continuous and linear preference functional \(\int_{E_i} p_s u_s(\cdot) \, d\mu(p, u)\) on \(\Delta(C \times X)\). By the Expected Utility Theorem, this linear functional has a vN-M utility representation which we denote by \(\tilde{p}_i(s) \bar{u}_{i,s}\). Thus, for all \(\alpha \in \Delta(C \times X)\), we have

\[\tilde{p}_i(s) \bar{u}_{i,s}(\alpha) = \int_{E_i} p(s) u_s(\alpha) \, d\mu(p, u)\]

Then, \(\tilde{p}_i(s) \bar{u}_{i,s}\) is a local EU representation of \(\mu\) on \(E_i\) for state \(s\).

**Definition K.4.** Let \(\mu \in \mathcal{M}\) and \((E_i)\) a partition of \(\Omega\). Then,

- A measure \(\mu\) is Type Ia on \(E_i\) in state \(s\) if \(\tilde{p}_i(s) \bar{u}_{i,s} = 0\), i.e., if \(\tilde{p}_i(s) \bar{u}_{i,s}\) is trivial.
- A measure \(\mu\) is Type Iib on \(E_i\) in state \(s\) if \(\tilde{p}_i(s) \bar{u}_{i,s}\) is non-trivial, \(\tilde{p}_i(s) \bar{u}_{s}\) is constant on \(\Delta(C \times L)\), and \(\ell_s\) maximizes \(\tilde{p}_i(s) \bar{u}_{i,s}\) on \(C \times X\).
- A measure \(\mu\) is Type IIa on \(E_i\) in state \(s\) if \(\tilde{p}_i(s) \bar{u}_{i,s}\) is non-trivial and not constant on \(\Delta(C \times L)\).
- A measure \(\mu\) is Type IIb on \(E_i\) in state \(s\) if \(\tilde{p}_i(s) \bar{u}_{i,s}\) is non-trivial, constant on \(\Delta(C \times L)\), and there exists \(\alpha \in \Delta(C \times X)\) such that \(\tilde{p}_i(s) \bar{u}_{i,s}(\alpha) > \tilde{p}_i(s) \bar{u}_{i,s}(\beta)\) for some (and hence all) \(\beta \in \Delta(C \times L)\).

It is easy to see that the above taxonomy of measures is both exclusive and exhaustive in the sense that any \(\mu \in \mathcal{M}\) can only be of one of the above types, but must be of some type. Analogous to the definition in Section 3.1 (and abusing notation), for any \(\alpha \in \Delta(C \times X)\), we define

\[(f \oplus_{\varepsilon,s} \alpha)(s') := \begin{cases} (1 - \varepsilon)f(s) + \varepsilon \alpha & \text{if } s' = s \\ f(s) & \text{otherwise} \end{cases}\]

**Lemma K.5.** Let \(x\) be a finite menu, \(\mu \in \mathcal{Y}(x)\), and suppose there is a partitional optimal strategy \(\zeta^U_x\) with optimal partition \((E_i)\), where \(\zeta^U_x(E_i) = f_i \in x\). Suppose \(\mu\) is Type II (a or b) on some \(E_i\) in state \(s \in S\) and there exists \(\alpha \in \Delta(C \times X)\) such that

\[\int_{E_i} p(s) u_s(\alpha - f_i(s)) \, d\mu(p, u) > 0\]

Then, the menu \(x_{i,\varepsilon} := x \setminus \{f_i\} \cup \{f_i \oplus_{\varepsilon,s} \alpha\}\) is such that \(V(x_{i,\varepsilon}) > V(x)\) for all \(\varepsilon > 0\).

**Proof.** Let \(\mu \in \mathcal{Y}(x)\) so that \(V(x) = V(x, \mu)\). If

\[\int_{E_i} p(s) u_s(\alpha - f_i(s)) \, d\mu(p, u) > 0\]

then it must necessarily be that \(\mu(E_i) > 0\). The measure \(\mu\) and the set \(E_i\) induce the functional

\[V_i(x, \mu, E_i) := \int_{E_i} \max_{f \in \mathcal{X}} \sum_s p(s) u_s(f(s)) \, d\mu(p, u)\]
on $X$. Let $V_i^0$ denote the restriction of $V_i$ to $\mathcal{F}(\Delta(C \times X))$. By construction,

$$V_i^0(f) = \int_{E_i} \sum_s p(s)u_s(f(s)) \, d\mu(p, u)$$

and because $\mu(E_i) > 0$, $V_i^0$ is non-trivial. In particular, by hypothesis, we have $V_i^0(f \oplus_{\varepsilon, s} \alpha) > V_i^0(f_i)$.

Consider, now, the menu $x_{i,\varepsilon}$ and the strategy which entails the choice of $f_j$ for $(p, u) \in E_j$ when $j \neq i$, and the choice of $f_i \oplus_{\varepsilon, s} \alpha$ when $(p, u) \in E_i$. This strategy delivers utility bounded above by $V(x_{i,\varepsilon}, \mu)$, ie,

$$V(x_{i,\varepsilon}, \mu) \geq \sum_{j \neq i} \left[ \int_{E_j} \sum_s p(s)u_s f_j(s) \, d\mu(p, u) \right] + \int_{E_i} \sum_s p(s)u_s(f_i(s)) \, d\mu(p, u)
+ \varepsilon \int_{E_i} p(s)u_s(\alpha - f_i(s)) \, d\mu(p, u)
= V(x) + \varepsilon \int_{E_i} p(s)u_s(\alpha - f_i(s)) \, d\mu(p, u)
> V(x)$$

because $\int_{E_i} p(s)u_s(\alpha - f_i(s)) \, d\mu(p, u) > 0$ by hypothesis. Noting that $V(x_{i,\varepsilon}) \geq V(x_{i,\varepsilon}, \mu)$ by the definition of $V$ completes the proof. \qed

Let $\mathcal{M}_0 := \{\gamma \in \mathcal{M} : \{f_1, \ldots, f_m\} \text{ generates } x \text{ for some } x \in X\}$. It follows from Lemma K.2 that for all finite $x$,

$$\max_{\mu \in \mathcal{M}_0} V(x, \mu) = \max_{\mu \in \mathcal{M}} V(x, \mu)$$

In what follows, we shall restrict attention to finite menus, and therefore, it suffices to consider the set $\mathcal{M}_0$. Let $\mathcal{Y}_0 : X_0 \Rightarrow \mathcal{M}_0$ be defined as $\mathcal{Y}_0(x) = \mathcal{Y}(x) \cap \mathcal{M}_0$.

**Lemma K.6.** Let $x_0 := \{f_1, \ldots, f_m\}$ be the generator set for some nice finite menu $x$, and suppose $\mu \in \mathcal{Y}(x_0)$. Let $J_i$ denote the states where $f_i$ is active, and also let $(E_i)$ represent an optimal partitional strategy (for $\mu$) at $x$ so that act $f_i$ is chosen in the set $E_i$. Then, $\mu$ is not Type II (a or b) at $E_i$ in state $s$ for all $i = 1, \ldots, m$ and $s \in J_i^c$.

**Proof.** Let $\mu \in \mathcal{Y}(x_0)$ so that $V(x) = V(x_0) = V(x_0, \mu)$ and suppose $\mu$ is Type II (a or b) at $E_i$ in state $s \in J_i^c$. Note also that because $x$ is nice, there is a unique $\xi_x \in \mathcal{S}_x$ such that $x \sim \mathcal{J}(\xi_x)$, and the generator of $x$ is unique.

**Case 1:** First consider the case where $f_1(s)$ is not a maximizer for $\bar{p}_i(s)\bar{u}_{i,s}(\cdot)$ on $\Delta(C \times X)$. Let $f^*_i$ be the act such that (i) $f^*_i(s') = f_i(s')$ for all $s' \neq s$, and (ii) while $f^*_i(s)$ maximizes $\bar{p}_i(s)\bar{u}_{i,s}(\cdot)$ on $\Delta(C \times X)$, so that $\bar{p}_i(s)\bar{u}_{i,s}(f^*_i(s)) > \bar{p}_i(s)\bar{u}_{i,s}(f_i(s))$. An act satisfying (ii) exists because $\mu$ is Type II at $E_i$ in state $s$.

Now, consider the menu $x_{i,\varepsilon} := \{f_1, \ldots, (1 - \varepsilon)f_i + \varepsilon f^*_i, \ldots, f_m\}$. By Lemma K.5, $V(x_{i,\varepsilon}) > V(x)$ for all $\varepsilon > 0$. Notice also that $x_{i,\varepsilon} \rightarrow x$ as $\varepsilon \rightarrow 0$. \qed
For any $\varepsilon > 0$, consider $\mathcal{E}_{x_1, e}$, and notice that the set-valued map $\varepsilon \mapsto \mathcal{E}_{x_1, e}$ is a continuous, closed, and compact valued correspondence. By Axiom IICC (Axiom 4), there exists $\xi \in \mathcal{E}_{x_1, e}$ such that $\mathcal{I}(\xi) \sim x_1, e$. Consider, now, the maximization problem (parametrized by $\varepsilon$)

\[
[P1] \quad W(\varepsilon) := \max \{ \mathcal{I}(\xi) \} \quad \text{s.t.} \quad \xi \in \mathcal{E}_{x_1, e}
\]

Notice that $W(0) = V(x)$ and that because $\mathcal{E}_{x_1, e}$ is finite, a solution to [P1] always exists. We claim that for any $\varepsilon > 0$, the value of problem [P1] is precisely the value of $x_1, e$, i.e., $W(\varepsilon) = V(x_1, e)$.

To see this, notice that from the proof of Lemma J.1, it follows that $V(x_1, e) \geq V(\mathcal{I}(\xi))$ for all $\xi \in \mathcal{E}_{x_1, e}$. By Axiom IICC (Axiom 4), there exists $\xi \in \mathcal{E}_{x_1, e}$ such that $V(\mathcal{I}(\xi)) = V(x_1, e)$. Therefore, $W(\varepsilon) \geq V(x_1, e)$. Combining the two inequalities establishes that $W(\varepsilon) = V(x_1, e)$ for all $\varepsilon > 0$.

By the Theorem of the Maximum — see for instance, Ok (2007, p306) — $W$ is continuous in $\varepsilon$. The Theorem of the Maximum also implies that the maximizer correspondence is upper hemi-continuous, and therefore for any $\xi^*_\varepsilon$ that is optimal for the problem [P1], the limit $\xi^*_0 := \lim_{\varepsilon \to 0} \xi^*_\varepsilon$ is also a maximizer. (The limit always exists because $\mathcal{E}_{x_1, e}$ is a continuous, closed, and compact valued correspondence.) The continuity of $W$ then implies that $W(0) = V(\mathcal{I}(\xi^*_0))$.

There are two possibilities now. The first is that for all $\varepsilon^0 > 0$, there exists $\varepsilon \in (0, \varepsilon^0)$ such that $\xi^*_\varepsilon(s) = (1 - \varepsilon)f_i + \varepsilon f_i^*$ is active in state $s$. Because $\xi^*_0 = \lim_{\varepsilon \to 0} \xi^*_\varepsilon$, it follows that $\xi^*_0(s) = f_i$, i.e., $f_i$ is active in state $s$. In other words, $\xi^*_0 \neq \xi_x$. But we have already established that $W(0) = V(x) = V(\mathcal{I}(\xi^*_0))$, which contradicts the assumption that $x$ is nice, which rules out this first possibility.

The only other possibility is that there exists an $\varepsilon_x > 0$ such that for all $\varepsilon < \varepsilon_x$, the act 

$(1 - \varepsilon)f_i + \varepsilon f_i^*$

is inactive in every such state $s \in J^*_\varepsilon$, i.e., $\xi^*_\varepsilon(s) \neq (1 - \varepsilon)f_i + \varepsilon f_i^*$. In this case, for all $\varepsilon < \varepsilon_x$, we have $\xi^*_0 = \xi^*_\varepsilon$. Because $x$ is nice, it must necessarily be that $\xi^*_0 = \xi_x$. This implies that for all such $\varepsilon$, $V(x_1, e) = W(\varepsilon) = W(0) = V(x)$. But this contradicts our earlier observation that $V(x_1, e) > V(x)$ if $\mu$ is Type II at $E_i$ in state $s$ whenever $f_i$ is active in state $s \in J^*_\varepsilon$. This contradiction rules out the second possibility, and completes the proof of the first case.

**Case 2:** Now, suppose that $f_i(s)$ is a maximizer for $p_i(s)\hat{u}_{i,s}(\cdot)$ on $\Delta(C \times X)$. If $\mu$ is of Type IIa on $E_i$ in state $s \in J^*_\varepsilon$, then $p_i(s)\hat{u}_{i,s}(\cdot)$ is not constant on $\Delta(C \times L)$. If $\mu$ is of Type IIb on $E_i$ in state $s \in J^*_\varepsilon$, then $p_i(s)\hat{u}_{i,s}(\cdot)$ is constant on $\Delta(C \times L)$. However, in either case, there exists $\ell \in L$ such that $p_i(s)\hat{u}_{i,s}(f_i(s)) > p_i(s)\hat{u}_{i,s}(\ell(s))$. (Such an $\ell$ exists because $f_i(s)$ is a maximizer of $p_i(s)\hat{u}_{i,s}(\cdot)$ and by hypothesis that $\mu$ is of Type II, there exists some $\beta \in \Delta(C \times L)$ that is not a maximizer.)

Consider the menu $\frac{1}{2}x + \frac{1}{2} \ell$. By Lemma K.3, we see that $\mu \in \mathcal{Y}(x)$ implies $\mu \in \mathcal{Y}(\frac{1}{2}x + \frac{1}{2} \ell)$. Because $x$ is nice, $x_0$, which satisfies $V(x_0) = V(x)$, is the unique generator set of $x$. L-Independence now implies that $V(\frac{1}{2}x_0 + \frac{1}{2} \ell) = V(\frac{1}{2}x + \frac{1}{2} \ell)$. Moreover, Lemma K.3 says that $\frac{1}{2}x + \frac{1}{2} \ell$ is nice. It follows immediately that $\frac{1}{2}x_0 + \frac{1}{2} \ell$ is a generator set for $\frac{1}{2}x + \frac{1}{2} \ell$.

Now consider the nice menu $\frac{1}{2}x + \frac{1}{2} \ell$ with generator $\frac{1}{2}x_0 + \frac{1}{2} \ell$, and let $\mu \in \mathcal{Y}(\frac{1}{2}x_0 + \frac{1}{2} \ell)$. By construction, $\frac{1}{2}f_i(s) + \frac{1}{2} \ell(s)$ is not a maximizer of $p_i(s)\hat{u}_{i,s}$ on $\Delta(C \times X)$ (although $f_i(s)$ is), which means that we now satisfy the hypotheses of Case 1. Lemma K.3 ensures that $\mathcal{Y}(\frac{1}{2}x + \frac{1}{2} \ell) \cap \mathcal{Y}(x) = \emptyset$ and that a partitional optimal strategy at $x$ is also optimal at $\frac{1}{2}x + \frac{1}{2} \ell$. These facts allow us to establish that even in this case, $\mu$ cannot be of Type II, which completes the proof.

Let $x$ be nice and let $\mu \in \mathcal{Y}_0(x)$. Let $(E_{\mu}^{ix})$ be a partitional representation of an optimal
strategy (for instance, one coming from the generators of \( x \)) given \( \mu \) and consider the mapping

\[
(\mu, E_i^{\mu,x}, s) \mapsto \tilde{p}_i(s)\tilde{u}_{i,s}(-) := \int_{E_i^{\mu,x}} p(s)u_{s}(-) \, d\mu(p, u)
\]

Let \( \{f_1, \ldots, f_k\} \) be the unique generator set of \( x \), and let \( J_i \) denote the set of states where \( f_i \) is active so \( (J_i) \) is a partition of \( S \). Now define

\[
\gamma_{\mu,x}^i := \sum_{s \in J_i} \tilde{p}_i(s)
\]

\[\square\]

\[
p_i(s) := \begin{cases} \tilde{p}_i(s)/\gamma_{\mu,x}^i & \text{if } s \in J_i \\ 0 & \text{otherwise} \end{cases}
\]

\[
\hat{u}_s := \gamma_{\mu,x}^i \tilde{u}_{i,s} \quad \text{where } i \text{ is such that } s \in J_i
\]

and let

\[
\hat{\mu}_p := \{\hat{\mu} \in \Delta(\Omega) : \text{supp}(\hat{\mu}) = \{(p_i, \hat{u}) : i = 1, \ldots, k \text{ where } k \leq n = |S|\}\}
\]

For each nice \( x \), fix a \( \mu \in \mathcal{Y}_0(x) \), and let \( \xi \in \mathcal{E}_x \) be such that \( \mathcal{J}(\xi) \sim x \). (Because \( x \) is nice, there is a unique such \( \xi \).) Let \( (J_i) \) be the partition of \( S \) generated by \( \xi \), and \( (E_i^{\mu,x}) \) a partition of \( \Omega \) that is consistent with the (unique) generator set of \( x \). Then, \( (E_i^{\mu,x}) \) represents a partitional optimal strategy for \( \mu \) at \( x \). Consider the mapping

\[
\mathcal{D}(\mu, x, (E_i^{\mu,x})) \mapsto \hat{\mu} \in \hat{\mathcal{M}}_p
\]

where \( \text{supp} \hat{\mu} = \{(p_i, \hat{u}) : i = 1, \ldots, k\} \), \( p_i \) for \( i = 1, \ldots, k \) and \( \hat{u} \) are defined in \( \square \), and \( \hat{\mu} \) itself is defined as

\[
\hat{\mu}((p_i, u)) = \mu(E_i^{\mu,x})
\]

We note that in the definition of the mapping \( \mathcal{D} \), we omit the partition \( (J_i) \) as it is uniquely pinned down if \( x \) is nice, which is the case here. (The domain of \( \mathcal{D} \) is easily defined, but notationally cumbersome, and because omitting it will not cause any confusion in the sequel, we refrain from a formal definition.)

A collection of probability measures \( \{p_1, \ldots, p_k\} \) on \( S \) (so each \( p_i \in \Delta(S) \)) forms a partitional system if (i) for all \( s \in S \), \( p_i(s) > 0 \) implies \( p_j(s) = 0 \) for all \( j \neq i \), and (ii) for all \( s \), \( \sum_{i=1}^{k} p_i(s) > 0 \). In other words, every state \( s \) is supported by exactly one \( p_i \) in the collection.

A positive measure \( \mu \) on \( \Omega \) is elementary if its support is Dirac (degenerate) on \( \Omega_{x, \ell^i(x)} \) and the support on \( \Delta(S) \) is a partitional system of probability measures on \( S \). In other words, there exist \( p_1, \ldots, p_k \in \Delta(S) \) and \( u_s \in \Omega_{x, \ell^i(x)} \) for all \( s \) such that \( \mu \) is supported on the finite collection \( (p_1, u), \ldots, (p_k, u) \) where \( u = (u_s)_{s \in S} \). Rather than saying that the marginal of \( \mu \) on \( \Delta(S) \) has support \( \{p_1, \ldots, p_k\} \), we will often say in the sequel that \( \mu \) supports the partitional system \( (p_i) \).

With these definitions, it is clear that each \( \hat{\mu} \in \hat{\mathcal{M}}_p \) is elementary. The following proposition says that it is without loss of generality to restrict attention to elementary measures. Towards this end, let us define \( \hat{V} : X_0 \to \mathbb{R} \) as

\[
\hat{V}(x) := \sup_{\mu \in \hat{\mathcal{M}}_p} \left[ \sum_i \left[ \max_{f \in \mathcal{E}_x} \sum_s p_i(s)u_s(f(s)) \right] \mu(p_i, u) \right]
\]

24
Proposition K.7. For all nice \( x \), \( \hat{V}(x) = V(x) \). Moreover, the supremum is attained in the definition of \( \hat{V} \).

Proof. Let \( x \) be nice, \( \mu \in \mathcal{Y}_0(x) \), and \( \{f_1, \ldots, f_k\} \) the unique generator set of \( x \). Let us first prove that \( V(x) \leq \hat{V}(x) \). Let \( (E_i^{\mu,x})_i \) be the optimal partitional strategy representing \( \mu \) at \( x \) and let \( \hat{\mu} = \mathcal{D}(\mu, x, (E_i^{\mu,x})) \). Then,

\[
V(x, \mu) = \sum_i \max_{f \in x} \left[ \sum_s \int_{E_i^{\mu,x}} p(s)u_s(f(s)) \, d\mu(p, u) \right] \\
= \sum_i \max_{f \in x} \sum_s \tilde{p}_i(s)\tilde{u}_{i,s}(f(s))
\]

Lemma K.6 says that \( \mu \) cannot be of Type II (a or b) if \( s \in J_i^\epsilon \), and hence must be either Type Ia or Type Ib. In either case, \( \tilde{p}_i(s)\tilde{u}_{i,s}(f(s)) \leq 0 = \tilde{p}_i(s)\tilde{u}_{i,s}(f^\dagger(s)) \) for all \( s \in J_i^\epsilon \). Therefore, it must be that

\[
V(x, \mu) \leq \sum_i \max_{f \in x} \sum_s p_i(s)\hat{u}_{s}(f(s)) = \hat{V}(x, \hat{\mu}) \leq \hat{V}(x)
\]

which proves one half of our claim.

We now prove that \( \hat{V}(x) \leq V(x) \) for all nice \( x \). Suppose, by way of contradiction, that \( \hat{V}(x, \hat{\mu}) > V(x) \) for some nice \( x \) and \( \hat{\mu} \in \hat{\mathcal{M}}_p \). Suppose the optimal strategy here is to choose \( f_i \in x \) whenever the ‘interim information’ is \( (p_i, u) \).

Now recall that \( \hat{\mu} = \mathcal{D}(\mu, y, (E_i^{\mu,y})) \) for some \( \mu \in \mathcal{Y}_0 \) and \( y \in X_0 \). Consider the strategy \( \xi^\mu \) that is constant on \( E_i^{\mu,y} \), ie, satisfies \( \xi^\mu(E_i^{\mu,y}) = f_i \in x \) for each \( i \) (where \( f_i \) is the optimal choice when presented with the interim information \((p_i, u)\)). The value of this strategy, \( V(x, \mu, \xi^\mu) \), is given by

\[
V(x, \mu, \xi^\mu) = \sum_i \left[ \sum_s \int_{E_i^{\mu,y}} p(s)u_s(f_i(s)) \, d\mu(p, u) \right] \\
= \sum_i \left[ \sum_s \tilde{p}_i(s)\tilde{u}_{i,s}(f_i(s)) \right]
\]

It follows from Lemma K.6 that \( \mu \) is not Type II (a or b) at \( E_i^{\mu,y} \) in state \( s \) for all \( s \in J_i^\epsilon \). (Note that the partition \( (J_i) \) is generated by the unique \( \xi_y \in \Xi_y \) such that \( J(\xi_y) \sim y \). Thus, \( (J_i) \) does not depend on \( x \).) Therefore, for all such \( s \in J_i^\epsilon \), it must be that \( \tilde{p}_i(s)\tilde{u}_{i,s}(f_i(s)) \leq 0 \). For such an \( s \in J_i^\epsilon \), if we replace \( f_i(s) \) by \( \ell_\ast \), we obtain the new menu \( x' \), which has the property that \( V(x', \mu, \xi^\mu_{x'}) = \hat{V}(x, \hat{\mu}) \). But this implies \( V(x') \geq \hat{V}(x, \hat{\mu}) > V(x) \), which violates Axiom IICC (Axiom 4) because \( x' \) is obtained from \( x \) by replacing payoffs in acts in \( x \) by \( \ell_\ast \). This proves that \( \hat{V}(x) = V(x) \) for all nice \( x \).

Now, to show that the maximum is achieved in the definition of \( \hat{V}(x) \), observe that for each nice \( x \), there exists \( \mu \in \mathcal{Y}_0(x) \), so that

\[
V(x) = V(x, \mu) \quad \text{definition of } \mu \\
\leq \hat{V}(x, \hat{\mu}) \quad \text{from proof of } V(x) \leq \hat{V}(x) \text{ above} \\
\leq \hat{V}(x) \quad \text{definition of } \hat{V} \\
\leq V(x) \quad \text{because } \hat{V}(x) \leq V(x) \text{ as proved above}
\]
where \( \hat{\mu} = \mathcal{D}(\mu, x, (E^\mu_i,x)) \), \( \mu \in \mathcal{Y}_0(x) \), and \( (E^\mu_i,x) \) is an optimal partition strategy for \( \mu \) at \( x \). Therefore, \( \hat{\mu} \) is \( \hat{V} \)-optimal for \( x \), as claimed.

Because \( V \) is Lipschitz, it follows immediately that \( \hat{V} \) is also Lipschitz on \( X_0 \). Because \( X_0 \) is dense in \( X \), \( \hat{V} \) uniquely extends to \( X \). It is easy to see that in the representation of \( \hat{V} \), this amounts to replacing \( \mathcal{A}_P \) with its closure. In what follows, we shall therefore assume that \( \mathcal{A}_P \) is closed and that \( \hat{V} \) is defined on \( X \).

Thus far, we have shown that \( \succeq \) is represented by a function \( V : X \to \mathbb{R} \) that has the form

\[
V(x) = \max_{\mu \in \mathcal{A}} V(x, \mu)
\]

where

- each \( \mu \in \mathcal{A} \) is a positive elementary measure,
- \( V(x, \mu) = \left[ \sum_{p \in \Delta(S)} \left( \max_{f \in X} \sum_{s \in S} p(s) u_s(f(s)) \right) \mu(p; u) \right] \), and
- \( V(\ell; \mu) = \hat{V}(\ell; \mu') \) for all \( \mu, \mu' \in \mathcal{A} \) and \( \ell \in L \).

Our first result establishes that we can replace an elementary measure by an elementary probability measure.

**Lemma K.8.** Let \( \mu \) be an elementary measure. Then, there exists an elementary probability measure \( \hat{\mu} \) such that for all \( x \in X \), \( V(x, \mu) = V(x, \hat{\mu}) \).

**Proof.** Let \( \mu \) be supported on \( (p_1, u), \ldots, (p_k, u) \), and let \( \|\mu\|_1 \) be the total weight of \( \mu \). (That is, \( \|\mu\|_1 := \sum_i \mu((p_i, u)) \).) For any \( s \in S \), define \( \hat{u}_s := \|\mu\|_1 u_s \), and for any \( p \in \Delta(S) \), let \( \hat{\mu}(p, \hat{u}) := \mu(p, u)/\|\mu\|_1 \) where \( \hat{u} = (\hat{u}_s)_{s \in S} \). It is easy to see that \( \hat{\mu} \) so defined is elementary and is also a probability measure.

Moreover, we have

\[
V(x, \hat{\mu}) = \sum_p \max_{f \in X} \sum_s \hat{\mu}(p, \hat{u}) p(s) \hat{u}_s(f(s))
= \sum_p \max_{f \in X} \sum_s \frac{\mu(p, u)}{\|\mu\|_1} p(s) \|\mu\|_1 u_s(f(s))
= V(x, \mu)
\]

which establishes the claim.

Two partitional systems of probability measures \( \{p_1, \ldots, p_k\} \) and \( \{q_1, \ldots, q_k\} \) are *similar* if there exists a permutation \( \sigma : \{1, \ldots, k\} \to \{1, \ldots, k\} \) such that for all \( i = 1, \ldots, k \), \( \text{supp}(p_i) = \text{supp}(q_{\sigma(i)}) \).

Every elementary probability measure \( \mu \) on \( \Delta(S) \) supports a partitional system. We now show that we can replace, ie, without affecting utility considerations, \( \mu \) by another elementary probability measure \( \hat{\mu} \) that supports another partitional system that is similar to the partitional system supported by \( \mu \).
**Lemma K.9.** Let $\mu$ be an elementary probability measure whose support is $(p_1, u), \ldots, (p_k, u)$. Let $\{\hat{p}_1, \ldots, \hat{p}_k\}$ be a partitional system on $\Delta(S)$ that is similar to $\{p_1, \ldots, p_k\}$. Then, there exists an elementary probability measure $\hat{\mu}$ with support $(\hat{p}_1, \hat{u}), \ldots, (\hat{p}_k, \hat{u})$ such that for all $x \in X$ we have $V(x, \mu) = V(x, \hat{\mu})$.

**Proof.** We will assume, without loss of generality, that for all $i$, supp$(p_i) = \text{supp}(\hat{p}_i)$. Now, define $\hat{u}_s := (p_i(s)/\hat{p}_i(s))u_s$, and set $\mu(p_i, u) = \hat{\mu}(\hat{p}_i, \hat{u})$, where $\hat{u} = (\hat{u}_s)_{s \in S}$. Then, we have

$$V(x, \hat{\mu}) = \sum_i \max_{f \in \mathcal{X}} \sum_s \hat{\mu}(\hat{p}_i, \hat{u}) \hat{p}_i(s) \hat{u}_s(f(s))$$

$$= \sum_i \max_{f \in \mathcal{X}} \sum_s \mu(p_i, u) p_i(s) u_s(f(s))$$

$$= V(x, \mu)$$

which completes the proof. \hfill \Box

Let $\mu$ be an elementary probability measure and define $\pi_{\mu} \in \Delta(S)$ as

$$\pi_{\mu}(s) := \sum_p \mu(p) p(s)$$

Let $\pi_0 \in \Delta(S)$ and $P := (J_i)$ be a partition of $S$. Then, the conditional probability induced by $J_i$ is $q_i(\cdot, \pi_0 \mid J_i)$ where

$$q_i(s; \pi_0 \mid J_i) := \pi_0(s \mid J_i)$$

for all $J_i \in P$. It is easy to see that $(q_i(\cdot, \pi_0 \mid J_i))$ is a partitional system of probabilities on $S$. Conversely, let $\mu$ be an elementary measure that supports the partitional system $(p_i)$. This induces the partition $P_{\mu} := (J_i)$ of $S$ where $J_i := \text{supp}(p_i)$.

**Lemma K.10.** Let $\pi_0 \in \Delta(S)$, $\mu$ an elementary probability measure that supports the partitional system $(p_i)$, and let $(J_i)$ be the partition of $S$ induced by $(p_i)$. Then, there exists an elementary probability measure $\hat{\mu}$ such that

(a) $\hat{\mu}$ supports the partitional system $(q_i(\cdot, \pi_0 \mid J_i))$,

(b) $\pi_{\hat{\mu}} = \pi_0$, and

(c) $V(x, \mu) = V(x, \hat{\mu})$ for all $x \in X$.

**Proof.** Let $\mu$ and $\pi_0$ be as hypothesized and consider the induced partitional system $q(\cdot; \pi_0 \mid P_{\mu})$. By Lemma K.9, there exists an elementary probability measure $\hat{\mu}$ that supports $(q_i(\cdot; \pi_0 \mid J_i))$ while keeping utilities unaltered.

For each $s$, define the utility function

$$\hat{u}_s := \left[ \frac{\sum_i \hat{\mu}(q_i(s; \pi_0 \mid J_i), \hat{u})}{\sum_i \pi_0(J_i) 1_{\{s \in J_i\}}} \right] \hat{u}_s$$

and observe that in the sums in both the numerator and denominator, only one term is non-zero. Now, define the elementary probability measure $\hat{\mu}$ as follows: If $s$ is supported by $q_i(\cdot; \pi_0 \mid J_i)$, set

$$\hat{\mu}(q_i(\cdot; \pi_0 \mid J_i), \hat{u}) := \pi_0(J_i)$$
which proves (a). With this definition, \( \pi_\mu(s) = \mu((q_i(\cdot \mid J_i), \hat{u})) \cdot q_i(s; \pi_0 \mid E_i) = \pi_0(s) \), as desired for the proof of (b). To see (c), notice that we have

\[
V(x, \hat{\mu}) = \sum_i \max_{f \in x} \sum_s \hat{\mu}(q_i(\cdot \mid J_i), \hat{u}) q_i(s; \pi_0 \mid J_i) \hat{u}_s(f(s))
\]

which completes the proof. \( \square \)

We are now in a position to prove Proposition C.2.

**Proof of Proposition C.2.** We shall first prove (a) implies (b). We have shown that given the representation \( \bullet \) in Theorem 4 and IICC (Axiom 4), \( V \) has the form in [K.1]. In [K.1], every \( \mu \in \mathcal{M} \) is an elementary (positive, but finite) measure. Lemma K.8 shows that it is without loss of generality to assume that every measure supports the partitional system \( (p_i) \). Let \( J_i = \text{supp}(p_i) \), and notice that \( (J_i) \) is a partition of \( S \). Lemma K.10 says that it is without loss of generality to assume that every \( \mu \) supports the partitional system \( (q_i(\cdot \mid J_i)) \) (recall that \( q_i(s; \pi_0 \mid J_i) = \pi_0(s \mid J_i) \)) and also has the feature that \( \pi_\mu(s) := \sum_i \mu(q_i(s; \pi_0 \mid J_i))q_i(s; \pi_0 \mid J_i) = \pi_0(s) \) for all \( s \). (To ease notational burden, in what follows we shall write \( q_i(s; \pi_0 \mid J_i) \) as \( q_i(s) \).)

In particular, this last property implies that \( \mu(q_i, u) = \pi_0(J_i) \) and \( \mu(q_i, u)q_i(s) = \pi_0(J_i)\pi_0(s \mid J_i) \). This implies

\[
V(x, \mu) := \sum_i \left[ \max_{f \in x} \sum_s q_i(s) u_s(f(s)) \right] \mu(q_i, u)
\]

\[
= \sum_{J_i \in P} \left[ \max_{f \in x} \sum_s \pi_0(s \mid J_i) u_s(f(s)) \right] \pi_0(J_i)
\]

\[
= \sum_{J_i \in P} \left[ \max_{f \in x} \sum_{s \in J_i} \pi_0(s \mid J_i) u_s(f(s)) \right] \pi_0(J_i)
\]

\[
=: V'(x, \pi_0, (P, u))
\]

In other words, the informational content of the elementary probability measure \( \mu \) is now encoded into the prior \( \pi_0 \), the partition \( P = (J_i) \), and the utility functions \( u = (u_s) \). Let \( \mathcal{M}' \) be the collection of all such pairs \( (P, u) \) induced by elementary probability measures in \( \mathcal{M} \). Then, we can write

\[
V(x) = \max_{\mu \in \mathcal{M}} V(x, \mu)
\]

\[
= \max_{(P, u) \in \mathcal{M}'} V'(x, \pi_0, (P, u))
\]

\[
=: V'(x)
\]
where $V'(x) = V(x)$ for all $x \in X$; this proves the representation part.

Observe now — see [K.1] — that for all $\ell \in L$ and $\mu, \mu' \in \mathcal{M}$, we have $V(\ell, \mu) = V(\ell, \mu')$. This implies, for all $\ell \in L$ and $(P, u), (P', u') \in \mathcal{M}'$, we have $V'(\ell, \pi_0, (P, u)) = V(\ell, \pi_0, (P', u'))$.

Recall that $\ell^\dagger \in L$ is such that $u_s(\ell^\dagger(s)) = 0$ for all $s \in$. For any $\alpha \in \Delta(C \times L)$, define $\hat{\ell}_\alpha^s \in L$ as

$$
\hat{\ell}_\alpha^s(s') = \begin{cases} 
\alpha & \text{if } s' = s \\
\ell^\dagger(s') & \text{otherwise}
\end{cases}
$$

For all $(P, u), (P', u') \in \mathcal{M}'$, we then have $V(\hat{\ell}_\alpha^s, \pi_0, (P, u)) = V(\hat{\ell}_\alpha^s, \pi_0, (P', u'))$. Notice that $V(\hat{\ell}_\alpha^s, \pi_0, (P, u)) = \pi_0(s)u_x(\alpha) = \pi_0(s)u'_x(\alpha) = V(\hat{\ell}_\alpha^s, \pi_0, (P', u'))$. Since this is true for all $\alpha \in \Delta(C \times L)$, it follows that $u_x$ and $u'_x$ are identical on $C \times L$ for all $(P, u), (P', u') \in \mathcal{M}'$. This proves that (a) implies (b).

That (b) implies (a) follows immediately from Lemma K.10 which shows how to construct an elementary measure $\mu$ given the prior $\pi_0$, the partition $P_{\mu} = (J_i)$, and the utility function $u = (u_x)$. \qed