

# Subjective Dynamic Information Constraints\*

DAVID DILLENBERGER<sup>†</sup>

R. VIJAY KRISHNA<sup>‡</sup>

PHILIPP SADOWSKI<sup>§</sup>

APRIL 2016

## Abstract

We axiomatize a new 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 feasible 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 Blackwell dominance. 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, D90

## 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 <ddill@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 recursive dynamic model that incorporates intertemporal information constraints and allows us to identify and quantify them from observable choice behavior.<sup>1</sup> Just as with intertemporal budget constraints, intertemporal information constraints have the property that information choice in one period can affect the set of feasible information choices in the future. However, these constraints need not be linear and can accommodate many patterns, such as developing expertise in processing information. Our framework unifies behavioral phenomena that arise in the presence of such constraints, regardless of their nature. For example, it applies whether the constraints are cognitive, so that individuals have limited ability to take into account available information (as is common in the literature on rational inattention); or physical, where the constraint reflects the cost of acquiring information.

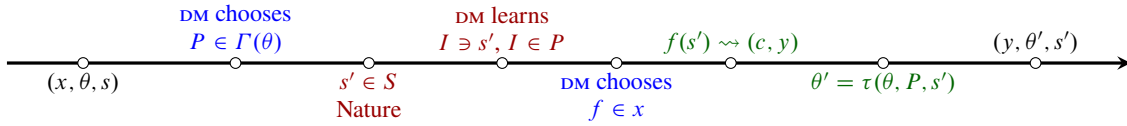
We axiomatize the behavior of a Decision Maker (henceforth DM) whose choice of information is constrained by a subjective *Markov Decision Process* (MDP), which specifies how future constraints depend on *unobservable* current and past choices of information. MDPs are the fundamental building blocks 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 MDP for *Information Choice* (MIC) is parametrized by an 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 of such MICs are at the end of this section.

We show that from observing DM's choice between appropriate infinite horizon decision 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 discount factor; and (iv) the MIC up to a recursive extension of Blackwell dominance. Here, identifying a subjective MDP from behavior is our main conceptual contribution.

The domain on which choice is observable is the space of Recursive Anscombe-Aumann Choice Problems (RACPS) that consists of menus of acts on  $S$ , where each act is a state-contingent lottery that yields current consumption and a continuation RACP (a new menu of acts for the next period).<sup>2</sup> Our representation suggests the following timing

- 
- (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 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

of events and decisions, as illustrated in Figure 1. DM enters a period facing an RACP  $x$ , and equipped with a prior belief  $\pi_s$  over  $S$  and an MDP state  $\theta$ . He first chooses a partition  $P \in \Gamma(\theta)$ . For any realization of a cell  $I \in P$ , DM updates his beliefs using Bayes' rule and chooses an act  $f$  from  $x$ . 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  $\theta' = \tau(\theta, P, s')$  and a new belief  $\pi_{s'}$  are determined for next period, where the latter is based on another subjectively perceived Markov process that governs the transition of the state, starting from  $s'$ .



**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:

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

Theorem 1 establishes that  $(u_s, \pi_s)_{s \in S}$  and  $\delta$  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 unobservable, its behavioral content lies entirely in 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.<sup>3</sup>

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

---

market.

(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.

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  $\theta$  — and so is not stationary — it is nonetheless recursive, meaning that preferences should satisfy the weaker notion of Self-Generation. Because  $\theta$  is subjective, observed behavior cannot condition on it. This 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 temporal 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 versatile tools that provide 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 others.

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 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 if the information constraint tightens over time, perhaps due to *search-fatigue*.

We now present some examples of MICS.

**Example 1.1.** Each period DM receives 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  $K' = \tau(K, P) = \beta[K + \kappa - c(P)]$  to determine the continuation constraint. The case with  $\beta = 0$  is a typical constraint in the literature on rational inattention.

**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  $c(P | Q) = (1 - b)H_\mu(P) + bH_\mu(P | Q)$ , where, given a probability  $\mu$  over  $S$ ,  $H_\mu(P)$  is the entropy of  $P$  and  $H_\mu(P | Q)$  is the relative entropy of  $P$  with respect to  $Q$ . Note that  $H_\mu(P | P) = 0$  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 DM can gain expertise. (Note that if  $\mu$  evolves over time it becomes part of the 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.<sup>4</sup> Formally, let  $\mathcal{P}_{\text{exp}} \subset \mathcal{P}$  be the space of all partitions of  $S$  that correspond to a possible experiment. If DM has chosen partitions  $P_1, \dots, P_n$  in the past, his current access to information is given by the partition  $P := P_1 \wedge \dots \wedge P_n$  (where  $P \wedge Q$  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

$$\Gamma(P, n) := \begin{cases} \{P \wedge Q : Q \in \mathcal{P}_{\text{exp}}\} & n < k \\ \{P\} & n = k \end{cases} \quad \text{and} \quad \tau((P, n), Q) := \begin{cases} (P \wedge Q, n + 1) & Q \neq P \\ (P, n) & Q = P \end{cases}$$

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 (ie, he can only learn the trivial partition of  $S$ ) today.<sup>5</sup>

**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.<sup>6</sup>

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 behavioral patterns that our model can address. Section 5 surveys

(4) For example, a growing startup may be able to gradually 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.

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; additional technical details are in the Supplementary Appendix.<sup>7</sup>

## 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  $\Delta(Y)$  the space of probability measures over  $Y$ , by  $\mathcal{F}(Y)$  the set of acts that map each  $s \in S$  to an element of  $Y$ , and by  $\mathcal{K}(Y)$  the space of closed and non-empty subsets of  $Y$ . The space of Recursive Anscombe-Aumann Choice Problems (RACPS) is  $X \simeq \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ .<sup>8</sup> 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  $\succsim$ , which is a binary relation on  $X$ , and denote its restriction to  $L$  by  $\succsim|_L$ .

The space  $X$  of RACPS, which embodies the *descriptive approach* of Kreps and Porteus (1978), 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 subspace  $L$  of consumption streams is also a subspace of the domain in Krishna and Sadowski (2014).<sup>9</sup>

### 2.2. MIC-Representation

Given 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 MDP for Information Choice (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;  $\Gamma : \Theta \rightarrow 2^{\mathcal{P}} \setminus \emptyset$  is a set of feasible partitions in a given MDP state  $\theta$ ; and  $\tau : \mathcal{P} \times \Theta \times S \rightarrow \Theta$  is a *transition operator* that determines the transition of the MDP state  $\theta$ , given a particular choice of partition and the realization of an objective state. Let  $\mathbf{M}$  be the space of MICs.

(7) The Supplementary Appendix, Dillenberger, Krishna, and Sadowski (2016), is available online at <http://people.duke.edu/~ps78/documents/ric-supp.pdf>.

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

(9) The domain in Krishna and Sadowski (2014) is  $Z \simeq \mathcal{F}(\mathcal{K}(\Delta(C \times Z)))$ , the space of *State-Contigent Infinite Horizon Consumption Problems*.

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<sup>10</sup> 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  $\succsim$  on the space of RACPS.

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

$$V(x, \theta, s) =$$

$$[\text{Val}] \quad \max_{P \in \Gamma(\theta)} \sum_{I \in P} \left[ \max_{f \in X} \sum_{s' \in I} \mathbf{E}^{f(s')} \left[ u_{s'}(c) + \delta V(y, \tau(P, \theta, s'), s') \right] \pi_s(s' | I) \right] \pi_s(I)$$

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  $\mathbf{E}^{f(s')}$  is the expectation with respect to this probability measure.<sup>11</sup>

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 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  $V : X \times \Theta \times S \cup \{s_0\} \rightarrow \mathbb{R}$  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  $\{P_1, \dots, P_m\}$  and  $\{Q_1, \dots, Q_n\}$ . If for every  $Q_i$  there is a  $P_j$  that is finer than it, we say that  $\{P_1, \dots, P_m\}$  *setwise Blackwell dominates*  $\{Q_1, \dots, Q_n\}$ . 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.

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

(11) One of the central properties of dynamic choice is *dynamic consistency*, which requires 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 DM and is not observed by the analyst.



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  $\mathcal{M}, \mathcal{M}^\dagger \in \mathbf{M}$ ,  $\mathcal{M}$  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).

**Theorem 1.** *Let  $((u_s), \delta, \Pi, \mathcal{M})$  be an MIC-representation of  $\succsim$ . 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.*<sup>12</sup>

A proof is in Appendix B. An immediate benefit of identifying all the parameters is that it allows a meaningful comparison of people.

Consider, then, two decision makers with preferences  $\succsim$  and  $\succsim^\dagger$ , respectively. We say that  $\succsim$  has a *greater affinity for dynamic choice* than  $\succsim^\dagger$  if for all  $x \in X$  and  $\ell \in L$ ,  $x \succsim^\dagger \ell$  implies  $x \succsim \ell$ .<sup>13</sup> The comparison in the definition implies that  $\succsim$  and  $\succsim^\dagger$  have the same ranking over consumption streams in  $L$ .<sup>14</sup> A typical RACP  $x$  may allow DM to wait for information to arrive over multiple periods before making a choice. This option should be more valuable the more information plans DM’s 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  $((u_s^\dagger), \delta^\dagger, \Pi^\dagger, \mathcal{M}^\dagger)$  be MIC-representations of  $\succsim$  and  $\succsim^\dagger$  respectively. The preference  $\succsim$  has a greater affinity for dynamic choice than  $\succsim^\dagger$  if, and only if,  $\Pi = \Pi^\dagger$ ,  $\delta = \delta^\dagger$ ,  $(u_s)_{s \in S}$  and  $(u_s^\dagger)_{s \in S}$  are identical up to the addition of constants and a common scaling, and  $\mathcal{M}$  recursively Blackwell dominates  $\mathcal{M}^\dagger$ .*

A proof is in Appendix B. The Proposition connects a purely behavioral comparison of 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 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 RICS, which is the canonical space of information constraints, and Section 2.5 formalizes the duality, which we call *strong alignment*, between the space of RICS and the domain  $X$  of RACPs.

---

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

(13) This definition is the analogue of notions of ‘greater preference for flexibility’ in the dynamic settings of Higashi, Hyogo, and Takeoka (2009) and Krishna and Sadowski (2014).

(14) That is,  $\ell \succsim \ell'$  if, and only if,  $\ell \succsim^\dagger \ell'$  for all  $\ell, \ell' \in L$ . This is Lemma 34 in Appendix F of Krishna and Sadowski (2014), and uses the fact that both  $\succsim$  and  $\succsim^\dagger$  satisfy Independence on  $L$ .



## 2.4. Recursive Information Constraints

The MICS  $\mathcal{M}$  and  $\mathcal{M}'$  are *indistinguishable* if they afford the same choices of partition in the first period and, for any choice in the first 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  $\Omega$  which is compact and metrizable. We now describe the construction of  $\Omega$ . Suppose DM can choose from the set of partitions  $\{P_1^{(1)}, \dots, P_m^{(1)}\}$  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  $\{P_1^{(2)}, \dots, P_k^{(2)}\}$  is available in the second period. If this process proceeds to infinity, with each history of choices and realized states determining the next set of feasible partitions, we get an element  $\omega$  of  $\Omega$ .

The description above suggests a *recursive* way to think of  $\Omega$ : 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  $\omega'_s \in \Omega$  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,  $\Omega$  is isomorphic to  $\mathcal{K}_b(\mathcal{P} \times \Omega^S)$ .<sup>15</sup> We call  $\Omega$  the space of *Recursive Information Constraints* (RICS) so that each  $\omega \in \Omega$  is an RIC. Conversely, every RIC is also an MIC. (Indeed, set  $\Gamma^*(\omega) = \{P : (P, \omega') \in \omega\}$  and  $\tau^*(\omega, P, s) = \omega'_s$  to obtain the MIC  $\mathcal{M}_\omega = (\Omega, \omega, \Gamma^*, \tau^*)$  which is indistinguishable from  $\omega$ .)

**Proposition 2.4.** The space  $\mathbf{M}$  of MICS is isomorphic to  $\Omega$  in the following sense.

- (a) Every  $\mathcal{M} \in \mathbf{M}$  is indistinguishable from a unique  $\omega_{\mathcal{M}} \in \Omega$ .
- (b) Every  $\omega \in \Omega$  induces an  $\mathcal{M}_\omega \in \mathbf{M}$  that is indistinguishable from  $\omega$ .

A proof is in Appendix A.6. Viewing  $\Omega$  as the canonical state space for MICS implies that the recursive Blackwell order on  $\Omega$  is the unique recursive order so defined that is continuous and satisfies our notion of dominance (see Section 2.3 for the definition of dominance and also Proposition A.4 in the appendix).

When considering a representation with a canonical MIC  $\omega$  (that is, an RIC), we often

---

(15) For metric spaces  $X$  and  $Y$ , we denote by  $\mathcal{K}_b(X \times Y)$  the space of all non-empty closed subsets of  $X \times Y$  with the property that a subset contains distinct  $(x, y)$  and  $(x', y')$  only if  $x \neq x'$ . A formal construction of  $\Omega$  is given in Appendix A.3. The metric on  $\Omega$  can be intuitively described as follows: Consider  $\omega, \omega' \in \Omega$  as MICS. 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,  $\omega, \omega' \in \Omega$  are indistinguishable if, and only if, they are identical. The isomorphism between  $\Omega$  and  $\mathcal{K}_b(\mathcal{P} \times \Omega^S)$  is a homeomorphism.

write [Val] as

$$V(x, \omega, s) =$$

$$[\text{Val}] \quad \max_{(P, \omega') \in \omega} \sum_{I \in P} \left[ \max_{f \in x} \sum_{s' \in I} \mathbf{E}^{f(s')} \left[ u_{s'}(c) + \delta V(y, \omega'_{s'}, s') \right] \pi_s(s' | I) \right] \pi_s(I)$$

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

## 2.5. Strong Alignment between RACPS and RICS

Notice that on the subdomain  $L$ ,  $V$  is independent of  $\omega$  and satisfies Independence. Indeed,  $V$  is then completely characterized by the parameters  $((u_s), \delta, \Pi)$ , meaning that  $\succsim|_L$  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  $(u_s)$ .<sup>16</sup> What remains then is to identify the RIC  $\omega$ .

To do so, we must show that for RICS  $\omega$  and  $\omega'$  that do not recursively Blackwell dominate each other, there is an RACP  $x$  that *separates* them, ie,  $V(x, \omega, \cdot) > V(x, \omega', \cdot)$ . We start by considering *simple* RICS which have the form  $(P, \omega)$ , ie, which offer no information choice in the first period. An RACP  $x \in X$  is *strongly aligned* with a simple RIC  $(P, \omega)$  if (i)  $V(x, (P, \omega), \cdot) \geq V(x, \omega', \cdot)$  for all  $\omega'$ , 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 RICS 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  $\omega'$ . Thus, 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 every state  $s$ .<sup>17</sup> Clearly,  $\ell^*$  is the best consumption stream

(16) More precisely, Krishna and Sadowski (2014) establish the uniqueness of an RAA representation when the consumption space  $C$  is finite. We establish in Proposition 5.5 of the Supplementary Appendix that the existence and uniqueness of the RAA representation also holds when  $C$  is a compact metric space.

(17) 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  $s \in S$ .

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.)

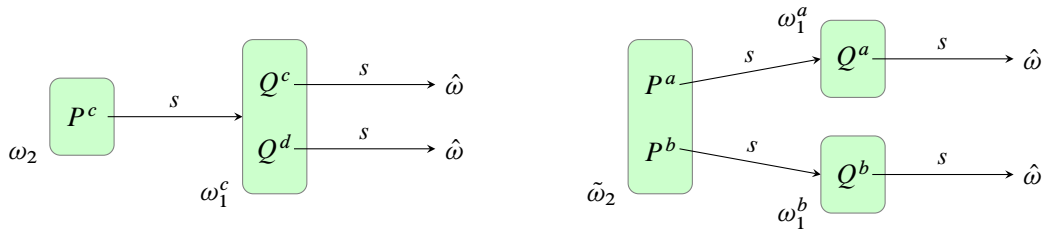
Let  $\hat{\omega}$  be the RIC that delivers the coarsest partition in every period, offering no choice at all. Every other RIC recursively Blackwell dominates  $\hat{\omega}$ .<sup>18</sup> Consider now an RIC  $\omega_1$  that only has non-trivial choice of partition in the first period, and provides 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

$$[\spadesuit] \quad x_1(P, \hat{\omega}) := \{f_{1,J} : J \in P\} \quad \text{where} \quad f_{1,J}(s) := \begin{cases} (c_s^+, \ell^*) & s \in J \\ (c_s^-, \ell_*) & s \notin J \end{cases}$$

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  $\omega_1$ .<sup>19</sup>

Our proof builds on this idea to construct an  $F_{\omega_2}$  that is uniformly strongly aligned with the 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 RIC of the form considered above. Given this extension, we can then proceed inductively to achieve strong alignment for any 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 RICs. In the rest of this section, we 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 example displayed in Figure 2.



**Figure 2:** RICS  $\omega_2$  and  $\tilde{\omega}_2$

(18) This follows immediately from Proposition A.4 in the Appendix.

(19) The construction of  $x_1(P, \hat{\omega})$  is as in Theorem 1 of Laffont (1989, p59), which concerns a static setting.

To construct the (two-period) RICS  $\omega_2$  and  $\tilde{\omega}_2$ , begin with the one-period RICS  $\omega_1^a := (Q^a, \hat{\omega})$ ,  $\omega_1^b := (Q^b, \hat{\omega})$ , and  $\omega_1^c := \{(Q^c, \hat{\omega}), (Q^d, \hat{\omega})\}$ , with  $\hat{\omega}$  as defined above. Then,  $\tilde{\omega}_2 := \{(P^a, \omega_1^a), (P^b, \omega_1^b)\}$  and  $\omega_2 := \{(P^c, \omega_1^c)\}$  where  $\omega_{1,s}^i = \omega_1^i$  for  $i \in \{a, b, c\}$ .

Suppose now that the partitions are strictly ordered in terms of fineness as follows (and are not ordered otherwise):

$$\begin{array}{ccc} P^a & \xrightarrow{\quad} & \\ P^b & \xrightarrow{\quad} & P^c \end{array} \qquad \begin{array}{ccc} Q^a & \longrightarrow & Q^c \\ Q^b & \longrightarrow & Q^d \end{array}$$

where  $P^b \longrightarrow P^d$ , for instance, indicates that  $P^b$  is strictly finer than  $P^d$ . Notice that neither  $\omega_2$  nor  $\tilde{\omega}_2$  recursively Blackwell dominates the other.

One might think that DM would always be better off with  $\tilde{\omega}_2$  as an RIC instead of  $\omega_2$ . After all, given  $\omega_2$ , the plan that says ‘Pick  $P^c$  in the first period and  $Q^c$  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  $\tilde{\omega}_2$ . Similarly, the plan ‘Pick  $P^c$  in the first period and  $Q^d$  in the second’ is dominated by ‘Pick  $P^b$  in the first period and  $Q^b$  in the second’.

Nevertheless, there are RACPS for which  $\tilde{\omega}_2$  is *not* as valuable to DM as  $\omega_2$  because, under  $\omega_2$ , choosing  $(P^c, \omega_1^c)$  in the first period allows DM to wait until the second period and for intervening uncertainty to resolve, before making a choice of partition for the second period. Define the two-period RACP  $x_2(P^c, \omega_1^c)$  as

$$\begin{aligned} x_2(P^c, \omega_1^c) &:= \{f_{2,J} : J \in P^c\} \quad \text{where} \\ f_{2,J}(s) &:= \begin{cases} (c_s^+, \text{Unif}((x_1(Q^c, \hat{\omega}), (x_1(Q^d, \hat{\omega})))) & s \in J \\ \ell_*(s) & s \notin J \end{cases} \end{aligned}$$

where  $\text{Unif}((x_1(Q^c, \hat{\omega}), (x_1(Q^d, \hat{\omega}))))$  is the equiprobable lottery over the RACPS  $x_1(Q^c, \hat{\omega})$  and  $x_1(Q^d, \hat{\omega})$  defined in [♣], 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 (singleton) set of RACPS  $F_{\omega_2} := \{x_2(P^c, \omega_1^c)\}$  is (uniformly) strongly aligned with  $\omega_2$ . Clearly,  $V(x_2(P^c, \omega_1^c), \omega_2, \cdot) = V(\ell^*, \omega', \cdot) \geq V(x', \omega', \cdot)$  for all  $x'$  and  $\omega'$ , as the plan ‘Pick  $P^c$  in the first period and then pick  $Q^c$  or  $Q^d$  in the second when facing  $x_1(Q^c, \hat{\omega})$  or  $x_1(Q^d, \hat{\omega})$ , respectively’ is at least as good as any other plan.

We now argue that  $x_2(P^c, \omega_1^c)$  separates  $\omega_2$  and  $\tilde{\omega}_2$ . To see this, consider the value of  $x_2(P^c, \omega_1^c)$  for the two feasible plans under  $\tilde{\omega}_2$ . 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  $Q^c$  or  $Q^d$ , but still receives, with probability  $\frac{1}{2}$ , an RACP that is strongly aligned with either  $Q^c$  or  $Q^d$ . Our earlier (static) argument shows that this entails a loss of utility, relative to  $\ell^*$ .

Similarly,  $F_{\tilde{\omega}_2} := \{x_2(P^a, \omega_1^a), x_2(P^b, \omega_1^b)\}$  is uniformly strongly aligned with  $\tilde{\omega}_2$ . In particular,  $x_2(P^a, \omega_1^a)$  is strongly aligned with  $(P^a, \omega_1^a)$  and thus separates  $\tilde{\omega}_2$  and  $\omega_2$ :  $P^c$  generates lower expected value than does  $P^a$ . 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 resolution of uncertainty over choice problems (a notion first introduced by Kreps and Porteus (1978)) which may render the *option* to delay information choice valuable.

**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. We can then define 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 resolution of uncertainty over choice problem is *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  $\succsim$  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  $\mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ , ignoring the fact that  $X$  is itself again the domain of our preferences (that is,  $X \simeq \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ ). Axiom 2 imposes assumptions on  $\succsim|_L$ , the restriction of  $\succsim$  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  $X$ .

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, DM 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 temporal separability, Strategic Rationality, Independence, and Stationarity are satisfied when the analyst is not able to condition observed behavior on information choice.<sup>20</sup> 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  $\succsim$  that are common in the menu-choice literature.

**AXIOM 1** (Basic Properties).

- (a) Order:  $\succsim$  is non-trivial, complete, and transitive.
- (b) Continuity: The sets  $\{y : y \succsim x\}$  and  $\{y : x \succsim y\}$  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  $(1 - t)x + t\ell^\# \succ (1 - t)y + t\ell_\#$ .
- (d) Monotonicity:  $x \cup y \succsim x$  for all  $x, y \in X$ .
- (e) Aversion to Randomization: If  $x \sim y$ , then  $x \succsim \frac{1}{2}x + \frac{1}{2}y$  for all  $x, y \in X$ .

Items (a)–(d) are standard.<sup>21</sup> Item (e) is familiar from Ergin and Sarver (2010) and De Oliveira et al. (2016) and relaxes Independence in order to accommodate unobserved information choice: Suppose DM 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, DM 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.

(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).



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  $\succsim$  to satisfy additional standard assumptions when consumption streams are involved. 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  $\ell^*$  and  $\ell_*$ , 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  $y$ .<sup>22</sup>

Let  $\ell_s := \ell_* \oplus_{(1, s)} \ell \in L$ , so that we can define the induced binary relation  $\succsim_s$  on  $L$  by  $\ell \succsim_s \hat{\ell}$  if  $\ell_s \succsim|_L \hat{\ell}_s$ .

**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  $tx + (1 - t)\ell \succ ty + (1 - t)\ell$ .
- (b) *L*-History Independence: For all  $\ell, \hat{\ell} \in L$ ,  $c \in C$ , and  $s, s', s'' \in S$ ,  $(c, \ell_s) \succsim_{s'} (c, \hat{\ell}_s)$  implies  $(c, \ell_s) \succsim_{s''} (c, \hat{\ell}_s)$ .
- (c) *L*-Stationarity: For all  $\ell, \hat{\ell} \in L$  and  $c \in C$ ,  $\ell \succsim|_L \hat{\ell}$  if, and only if,  $(c, \ell) \succsim|_L (c, \hat{\ell})$ .
- (d) *L*-Indifference to Timing:  $\frac{1}{2}(c, \ell) + \frac{1}{2}(c, \ell') \sim|_L (c, \frac{1}{2}\ell + \frac{1}{2}\ell')$ .

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. Temporal Separability

Whether or not 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 DM faces. That is, DM's optimal learning will not change

(22) Because 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.



when substituting act  $f$  with  $g$  as long as they induce, on each state  $s$ , the same marginal distributions over  $C$  and  $X$ . 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.

**AXIOM 3** (State-Contingent Indifference to Correlation). For a finite menu  $x$ , if  $f \in x$  and  $g \in \mathcal{F}(\Delta(C \times X))$  are such that  $g_1(s) = f_1(s)$  and  $g_2(s) = f_2(s)$  for all  $s \in S$ , then  $[(x \setminus \{f\}) \cup \{g\}] \sim x$ .<sup>23</sup>

### 3.3. Strategic Rationality

Suppose that, contingent on a sequence of actions and realizations, DM is offered a chance to replace a certain continuation problem with another. DM's attitude towards such replacements may depend on his previous information choices, which are 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 DM no worse off than receiving the worst consumption stream,  $\ell_*$ . In particular, because the best consumption stream,  $\ell^*$ , leaves DM strictly better off than  $\ell_*$  in every state, optimal choice from a menu  $(1-t)x + t\ell^*$  should give rise to a consumption stream that is also strictly better than  $\ell_*$ .

Formally, let  $X^* := \{(1-t)x + t\ell^* : x \in X \text{ is finite, } t \in (0, 1)\}$ . For a mapping  $e : x \rightarrow (0, 1]$ , let  $(x \oplus_{(e,s)} y) := \{f \oplus_{(e(f),s)} y : f \in x\}$ , which perturbs the continuation lottery in state  $s$  for any act  $f$  in  $x$  by giving weight  $e(f)$  to  $(c_s^-, y)$ . For  $x \in X^*$  we then require  $x \succ [x \oplus_{(e,s)} \ell_*]$  and  $[x \oplus_{(e,s)} y] \succsim [x \oplus_{(e,s)} \ell_*]$  for all  $s \in S$  and  $y \in X$ . This is part (a) of Axiom 4 below.

Part (b) investigates the conditions under which DM is actually *indifferent* to replacing continuation lotteries with the worst consumption stream. Recall that the RIC requires 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), DM's choice of partition determines which act he will choose from a given menu, contingent on the state. DM should then be willing to commit to this choice. In other words, there should be a contingent plan that specifies which act DM will choose for each state, such that he will be indifferent between the original menu and one where he is penalized whenever 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 all functions  $\xi : S \rightarrow x$ . An *Incentivized Contingent Commitment* to  $\xi \in \mathcal{E}_x$ , is then the set

$$\mathcal{F}(\xi) = \{f \oplus_{(1,I^c)} \ell_* : f \in x \text{ and } I = \{s : f = \xi(s)\}\}$$

(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.

which replaces the outcome of  $f$  with the worst outcome  $(c_s^-, \ell_*)$  in any state where  $f$  should not be chosen according to  $\xi$ . Obviously  $x \succsim \mathcal{F}(\xi)$  for all  $\xi \in \mathcal{E}_x$ . However, if for no  $s \in S$  is it ever optimal to choose an act outside  $\xi(s)$ , then  $x \sim \mathcal{F}(\xi)$  should hold.

**AXIOM 4** (Indifference to Incentivized Contingent Commitment).

- (a) If  $x \in X^*$  and  $e : x \rightarrow (0, 1]$ , then  $x \succ [x \oplus_{(e,s)} \ell_*]$  and  $[x \oplus_{(e,s)} y] \succsim [x \oplus_{(e,s)} \ell_*]$  for all  $s \in S$  and  $y \in X$ .
- (b) For all  $x \in X$ , there is  $\xi \in \mathcal{E}_x$  such that  $x \sim \mathcal{F}(\xi)$ .<sup>24</sup>

### 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 RACPS that lead to different optimal initial information choices,  $\succsim|_{X'}$  should satisfy Independence if  $X' \subset X$  consists only of concordant RACPS. 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 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  $x_1(P) := \{\ell^* \oplus_{(1,I^c)} \ell_* : I \in P\} \in \mathcal{K}(L)$ , we have  $x_1(P) \sim \ell^*$  if, and only if,  $x_1(P)$  is evaluated under a partition that is finer than  $P$ . (See also Section 2.5.)

Given these two observations, consider two RACPS  $x$  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  $\lambda \in (0, 1)$  small enough, such that  $(1 - \lambda)x + \lambda z \sim (1 - \lambda)y + \lambda z$  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  $(1 - \lambda)x + \lambda x_1(P_x) \sim (1 - \lambda)x + \lambda \ell^* \succ (1 - \lambda)y + \lambda x_1(P_x)$  for any  $\lambda \in (0, 1)$ . We will say that  $x$  and  $y$  are concordant in case (i) but not in (ii).<sup>25</sup> To extend this notion to  $x$  and  $y$  with  $x \approx y$ , 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  $x$ .

(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.

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

**AXIOM 5** (Concordant Independence). If  $x$  and  $y$  are  $\lambda$ -concordant, so are  $x$  and  $\frac{1}{2}x + \frac{1}{2}y$ . Furthermore, if  $X' \subset X$  consists of concordant RACPS, then  $\succsim|_{X'}$  satisfies Independence.<sup>26</sup>

### 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.<sup>27</sup> *Self-Generation* is satisfied in any recursive model precisely because it embodies the dynamic programming principle.

Because continuation preferences in our model 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:

$\succsim_{(P,s)}$  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 current preferences on  $X$  and induced preferences over the next period's continuation problems (again on  $X$ ) to satisfy the same axioms. This includes the Self-Generation Axiom itself, thereby connecting preferences over next period'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 next period's preferences are required to coincide with the current ones, and therefore those for two periods from now also coincide with next period'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.<sup>28</sup>

(26) If  $x, y, z, (1 - t)x + tz, (1 - t)y + tz \in X'$ ,  $t \in (0, 1)$ , and  $x \succ y$ , then  $(1 - t)x + tz \succ (1 - t)y + tz$ .

(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) Yet another formulation is that  $\succsim$  must belong to the recursively defined set  $\Psi^*$ , where  $\Psi^* := \{\succsim \text{ on } X : \text{(i) } \succsim \text{ satisfies Axioms 1–5, and (ii) } \succsim_{(P,s)} \in \Psi^*\}$ . 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.

Clearly,  $\succsim_{(P,s)}$  must be inferred from the initial ranking of RACPS, 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 RACP  $x$ . Because there are only finitely many partitions of  $S$ , we can perturb each 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. Contingent on  $s \in S$ , DM must anticipate choosing *some* act  $f \in x$ . Hence, if he prefers perturbing  $f(s)$  by  $y$  rather than  $y'$  simultaneously for *each*  $f \in x$ , we can infer that  $y \succsim_{(P,s)} y'$ . Based on this intuition, we now define an induced binary relation  $\succsim_{(x,s)}$  which coincides with  $\succsim_{(P,s)}$ .

**Definition 3.2.** If for  $y, y' \in X$ ,  $s \in S$ , and finite  $x$  there is  $\varepsilon \in (0, 1]$  such that  $x \oplus_{(\varepsilon,s)} y$ ,  $x \oplus_{(\varepsilon,s)} y'$ , and  $x$  are pairwise concordant, then  $y \succsim_{(x,s)} y'$  if  $[x \oplus_{(\varepsilon,s)} y] \succsim [x \oplus_{(\varepsilon,s)} y']$ .<sup>29</sup>

We verify in Appendix C.4 that  $\succsim_{(x,s)}$  is well defined. Further, for all  $x$  in a dense subset of  $X$ , it is complete (on  $X$ ). In that case  $\succsim_{(x,s)} = \succsim_{(P,s)}$ , where  $P$  is an optimal information choice given  $x$ . Conversely, for every  $P$  and  $s$  there is a finite  $x \in X$ , such that  $\succsim_{(x,s)} = \succsim_{(P,s)}$  on  $X$ .

**AXIOM 6 (Self-Generation).** If  $\succsim$  is such that  $\succsim_{(x,s)}$  (induced by  $\succsim$  as in Definition 3.2) is complete on  $X$ , then  $\succsim_{(x,s)}$  satisfies Axioms 1–6.

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 MIC-representation of  $\succsim$ .

The proof of Theorem 2 is quite involved. In Appendices C.1–C.3 we establish the following representation of  $\succsim$ :

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

where  $\mathcal{Q} \subset \mathcal{P}$  is a set of partitions of  $S$ , the measure  $\pi(s | I)$  is the probability of  $s$  conditional on the event  $I \subset S$ , and utilities ( $v_s$ ) over continuation problems depend only on the partition  $P$ . We say that  $V$  is *implemented* by  $((u_s), \mathcal{Q}, (v_s(\cdot, P)), \pi)$ . This representation

(29) Slightly abusing notation, we write  $x \oplus_{(\varepsilon,s)} y$  to denote  $x \oplus_{(\varepsilon \mathbf{1}, s)} y$ , where  $\mathbf{1}(f) = 1$  for all  $f \in x$ .

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.

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  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  in a way that each  $v_s(\cdot, P)$  is itself in  $\Phi^*$  (see Appendix A.8). In Appendix C.4, we rely on Self-Generation (Axiom 6) to show that the representation  $V$  of  $\succsim$  can be made self-generating.<sup>30</sup> Clearly, a self-generating value function may not be recursive.

The remainder of our construction in Appendix C.5 has two main components. First, we construct an RIC  $\omega_0$  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; temporal 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).

---

(30) See also Footnote 28.

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.

**Proposition 3.4.** If  $\succsim$  has MIC-representation  $((u_s), \delta, \Pi, \mathcal{M})$ , 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  $\succsim$ -best and  $\succsim_{(y,s)}$ -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, DM faces a trivial choice between information plans, that is, he can not influence the arrival of information about the state of the world.<sup>31</sup>

**Axiom 8 (Independence).** If  $x \succ y$ , then  $tx + (1-t)z \succ ty + (1-t)z$  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  $\succsim$  has MIC-representation  $((u_s), \delta, \Pi, \mathcal{M})$ , 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  $x_1(P) \sim x_1(P') \sim \ell^* \succ \frac{1}{2}x_1(P) + \frac{1}{2}x_1(P')$ , 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 DM 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 temporal 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 DM faces non-trivial unobserved choice, and full temporal separability, in the sense that preferences over continuation problems are independent of the initial choice problem, cannot hold if the subjective constraint is *not* time invariant.

---

(31) This parallels the representation in Krishna and Sadowski (2014), where DM faces a fixed stream of information about his own taste, rather than the state of the world.



## 4. Applications

We propose two applications that build on examples of MICS outlined in the [Introduction](#).

### 4.1. Expertise in learning and familiarity bias

Suppose 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 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.<sup>32</sup>

Let  $\mathcal{K}^\infty \subset X$  collect all separable consumption problems, where  $(F_t) \in \mathcal{K}^\infty$  denotes the RACP for which DM expects to choose from  $F_t \in \mathcal{K}(\mathcal{F}(\Delta(C)))$  in period  $t$ . For  $F, G \in \mathcal{K}(\mathcal{F}(\Delta(C)))$ , denote by  $F_T G$  the problem  $(F_t) \in \mathcal{K}^\infty$  with

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

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

**Definition 4.1.**  $\succsim$  is *familiarity biased* if

- (a)  $F_T G \succsim F_\infty$  implies  $G_\infty \succsim G_T F$  for all  $F, G \in \mathcal{K}(\mathcal{F}(\Delta(C)))$  and  $T > 0$ .
- (b)  $F_\infty \succ F_T G$  and  $G_\infty \succ G_T F$  for some  $F, G \in \mathcal{K}(\mathcal{F}(\Delta(C)))$  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 DM to familiarize himself with any RACP  $x$  before having to choose from it. For  $\ell \in L$  and  $x \in X$ , denote by  $\ell \diamond_{T^*} x \in X$  the RACP that pays according to  $\ell$  in every period until  $T^* - 1$  and then delivers  $x$  in every state in period  $T^*$ . For an arbitrary but fixed  $\ell \in L$ , define  $\succsim_{T^*}$  on  $X$  by  $x \succsim_{T^*} y$  if

---

(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 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.



$\ell \diamond_{T^*} x \succsim \ell \diamond_{T^*} y$ . That is,  $\succsim_{T^*}$  compares  $x$  and  $y$  after  $T^* - 1$  periods that require no choice.

Consider  $\succsim$  with the MIC-representation  $((u_s), \delta, \Pi, \mathcal{M})$ . We now verify that for  $T^*$  large enough,  $\succsim_{T^*}$  will be familiarity biased when the MIC  $\mathcal{M} = (\Theta, \theta_0, \Gamma, \tau)$  relies on expertise in the sense of the following two criteria. First, after learning  $P$  it is always possible to learn  $P$  again. Second, there are  $P \neq Q$  that are maximal among those accessible under  $\mathcal{M}$ ,<sup>33</sup> 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  $T^*$  large enough  $\succsim_{T^*}$  is familiarity biased.

A proof is in Appendix D. Intuitively, for  $T^*$  large enough DM can use the first  $T^*$  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  $F_T G \succ_{T^*} F_\infty$  and  $G_T F \succ_{T^*} G_\infty$ . 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 relies on expertise, provided DM can never learn the finest partition of  $S$ . In that case, it can be shown that  $T^* = 0$  in Proposition 4.3, that is,  $\succsim$  itself is familiarity biased.<sup>34</sup>

## 4.2. Search for Wages and Optimal Stopping Rule

Consider the following standard search problem.<sup>35</sup> An unemployed worker seeks to maximize  $\mathbf{E} \sum_{t=0}^{\infty} \delta^t a_t$ , 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

(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  $c(P | Q)$  is constant, as would be the case for states that are distributed independently over time.

(35) See for example Ljungqvist and Sargent (2004, p 161).

period is  $w_s \in W := \{w_1, \dots, w_n\}$ . 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 RACPS as follows. The stopping problem is the menu  $x(w) := \{\text{accept}, \text{continue}\}$ , 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  $\text{continue}(w) = (0, x)$  for all  $s$ .

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 based on Example 1.6: at the beginning of each period he can either learn nothing or pay attention and 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) DM 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.<sup>36</sup> 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.

---

(36) Brown, Flinn, and Schotter (2011) discuss this evidence and also document and investigate declining reservation wages in a laboratory experiment.

## 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, 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 corresponds to uncertainty about future beliefs about the objective state of the world. Ergin and Sarver (2010) and De Oliveira et al. (2016) weaken Independence in these respective models in order to accommodate subjective uncertainty that is not fixed, but a choice variable.<sup>37</sup> The former studies costly contemplation about future tastes, while the latter studies rational inattention to information about the state.

None of the models discussed so far are recursive or let DM react to information arriving over multiple periods. Krishna and Sadowski (2014) provide a dynamic extension of Dekel, Lipman, and Rustichini (2001) where the flow of information is taken as given by 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, RACPS 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 RIC (or equivalently the MIC).<sup>38</sup>

As a consequence of DM controlling the 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 RACPS, 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.

Piermont, Takeoka, and Teper (2015) study a decision maker who learns about his uncertain, but time invariant, consumption taste (only) through consumption, and so has some control over the flow of information.

For static choice situations, the literature based on ex post choice partly parallels the

---

(37) The insight that weakening Independence is essential in order to allow unobserved actions can be traced back at least to Markowitz (1959, Chapters 10 and 11); see also Gilboa and Schmeidler (1989) who consider an Anscombe-Aumann style setting and allow for the choice of beliefs to vary with the act.

(38) To be sure, Dillenberger et al. (2014) and De Oliveira et al. (2016) permit more general information structures than partitions, and the latter also allows for explicit costs of acquiring information.

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.

## 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  $\mathcal{DM}$  assigns to choice problems is consistent with the value optimal (static) choice at each instance will generate over the infinite horizon.

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

### Costly Information Acquisition

An alternative way to model limitations on information acquisition is via direct information costs, measured in consumption ‘utils’ (see, for example, Ergin and Sarver (2010), Woodford (2012), Caplin and Dean (2014), and De Oliveira et al. (2016)). We confine our attention to information constraints — and leave the analysis of information costs to future research — for a number of reasons.

First, in the  $\mathcal{RIC}$ -representation (or equivalently the  $\mathcal{MIC}$ -representation), the  $\mathcal{RIC}$  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.<sup>39</sup>

Finally, information costs or constraints can at best be identified 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. In contrast, the RIC in our model can be identified from preferences up to recursive Blackwell dominance, which ranks RICS *independently* of preferences.<sup>40</sup>

One approach to dealing with identification in the presence of costs would be to confine attention to representations that are minimal, in the sense that they only include undominated information strategies (see, for example, Ergin and Sarver (2010)). We could follow the same approach here.<sup>41</sup> Furthermore, for a minimal representation with information costs, our proof strategy would need only minor adaptations to establish unique identification.<sup>42</sup> However, most intuitive examples of information costs or constraints (including all those in Section 1) will *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

- 
- (39) 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 may be sensitive to the stakes in a given period, thereby reducing the gap between the two models.
  - (40) 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.
  - (41) Section 3 of the Supplementary Appendix provides a recursive notion of minimality for sets of RICS that is based on recursive Blackwell dominance, and is therefore independent of preferences.
  - (42) 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.

take into account the information about the state of nature encoded in its realization or, more generally, in the realization of the pair  $(c, y)$  of consumption and continuation problem.<sup>43</sup> But if  $\text{DM}$  learns *only* the information encoded in the act, then he should be willing to pay a premium (eg in terms of current consumption) to avoid acts with state-independent payoffs, leading to implausible violations of continuity. Moreover, if it is possible to place side bets with arbitrarily small stakes,  $\text{DM}$  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  $\text{DM}$  always learns the state at the end of the period.<sup>44</sup>

### 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  $\text{DM}$ '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  $\text{DM}$ .<sup>45</sup> 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  $\text{DM}$  is indifferent between  $x(P, \omega')$  and  $\ell^*$ , 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  $\text{DM}$  to follow a particular information plan with just one binary choice question.<sup>46</sup>

---

(43) 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 *less* accurate than  $\pi_s$ , which seems unreasonable.

(44) An alternative model could assume that  $\text{DM}$  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).

(45) 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  $y$ .

(46) 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  $\succsim|_L$  is discussed in Krishna and Sadowski (2014).



## MICS in Strategic Situations

This paper provides a recursive dynamic model of choice under intertemporal 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, it will be interesting to think about the *strategic interaction* of so-constrained agents. To suggest just one instance, consider a monopolistic competition setting where each firm faces an MICS 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 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?<sup>47</sup>

## Appendices

### A. Preliminaries

#### A.1. Metrics on Probability Measures

Let  $(Y, d_Y)$  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  $f^1$

(47) 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.



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  $\Delta(C \times X_{n-1})$ .

For acts  $f^n, g^n \in \mathcal{F}(\Delta(C \times X_{n-1}))$ , 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  $d_H^{(n)}$  on  $X_n$  as

$$d_H^{(n)}(x_n, y_n) := \max \left[ \max_{f^n \in x_n} d^{(n)}(f^n, y_n), \max_{g^n \in y_n} d^{(n)}(g^n, x_n) \right]$$

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, endow  $\prod_{n=1}^{\infty} X_n$  with the product topology. The Tychonoff metric induces this topology and is given as follows: For  $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots) \in \prod_{n=1}^{\infty} X_n$ ,

$$d(x, y) := \sum_n \frac{d_H^{(n)}(x_n, y_n)}{2^n}$$

It is easy to see that for all  $x, y \in \prod_{n=1}^{\infty} X_n$ ,  $d(x, y) \leq 1$ . Moreover, and this is easy to verify (because it holds for  $d_H^{(n)}$  for each  $n$ ),  $d(\frac{1}{2}x + \frac{1}{2}y, y) = \frac{1}{2}d(x, y)$ .

The space of RACPS  $X$  is all members of  $\prod_{n=1}^{\infty} X_n$  that are *consistent*. Intuitively,  $x = (x_1, x_2, \dots)$  is consistent if deleting the last period in the  $n$ -period problem  $x_n$  results in the  $(n-1)$ -period problem  $x_{n-1}$ .<sup>48</sup> The space of RACPS,  $X$ , is our domain for choice, and it follows from standard arguments that  $X$  is (linearly) homeomorphic to  $\mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ . We denote this homeomorphism by  $X \simeq \mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ . In what follows, we shall abuse notation and use  $d$  as a metric both on  $X$  as well as  $\mathcal{K}(\mathcal{F}(\Delta(C \times X)))$ . 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

(48) See also Gul and Pesendorfer (2004) for a more formal definition in a related setting.

its entropy  $H(P)$  as  $H(P) := -\sum_{J \in \mathcal{P}} \mu(J) \log \mu(J)$ . Then, we can define a metric  $d$  on  $\mathcal{P}$  as  $d(P, Q) := 2H(P \wedge Q) - H(P) - H(Q)$ , where  $P \wedge Q$  is the coarsest refinement of  $P$  and  $Q$ . In Section 4 of the Supplementary Appendix, we show that  $d$  is indeed a metric. Thus,  $(\mathcal{P}, d)$  is a metric space.

Let  $\Omega_1 := \mathcal{K}(\mathcal{P})$ , and define recursively for  $n > 1$ ,  $\Omega_n := \mathcal{K}_b(\mathcal{P} \times \Omega_{n-1}^S)$  (see Section 2.4 for a definition of  $\mathcal{K}_b$ ). Then, we can set  $\Omega' := \prod_{n=1}^{\infty} \Omega_n$ . A typical member of  $\Omega_n$  is  $\omega_n$ , while  $\omega_n = (\omega_{n,s})_{s \in S}$  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  $\Psi_1 : \Omega_2 \rightarrow \Omega_1$  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  $\Psi_n(\omega_{n+1}) := \{\psi_n(P, \omega_n) : (P, \omega_n) \in \omega_{n+1}\}$ .

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* (RICS) is

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

that is, the set of 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,  $\mathcal{P} \times \Omega_{n-1}^S$  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  $\mathcal{K}_b(\mathcal{P} \times \Omega^S)$ .*

We write the homeomorphism as  $\Omega \simeq \mathcal{K}_b(\mathcal{P} \times \Omega^S)$ . 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 MICS, ie, RICS. The extension to the case of general MICS is straightforward. In what follows, let  $\mathbf{C}(X \times \Omega \times (S \cup \{s_0\}))$  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  $V \in \mathbf{C}(X \times \Omega \times (S \cup \{s_0\}))$  satisfying [Val] that represents DM's preference over RACPS. Moreover, there is an optimal dynamic information plan.

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

$$TW(x, \omega, s') = \max_{(P, \omega') \in \omega} \sum_{I \in P} \left[ \max_{f \in X} \sum_{s \in S} \mathbf{E}^{f(s)} \left[ u_s(c) + \delta W(y, \omega'_s, s) \right] \pi_{s'}(s | I) \right] \pi_{s'}(I)$$

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,  $T(W + a) \leq TW + \delta a$ ), so  $T$  is a contraction mapping with modulus  $\delta \in (0, 1)$ . Therefore, for each DM who is characterized by  $((u_s)_{s \in S}, \Pi, \delta, \omega)$ , there exists a unique  $V \in \mathbf{C}(X \times \Omega \times (S \cup \{s_0\}))$  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*.<sup>49</sup> 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).  $\square$

## A.5. Recursive Blackwell Order

Let  $\hat{\omega} \in \Omega$  denote the RIC that delivers the coarsest partition in each period in every state. Define  $\hat{\Omega}_0 := \mathcal{K}_b(\mathcal{P} \times \{\hat{\omega}\})$ , and inductively define  $\hat{\Omega}_{n+1} := \mathcal{K}_b(\mathcal{P} \times \hat{\Omega}_n)$  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)  $\omega_{n-1,s} \succsim_{n-1} \omega'_{n-1,s}$  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,  $\succsim_{n+1} \upharpoonright_{\hat{\Omega}_n} = \succsim_n$ .

*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,  $\hat{\omega} \succsim_0 \hat{\omega}$ . 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)  $\hat{\omega} \succsim_0 \hat{\omega}$  for all  $s \in S$ . But this implies  $\omega_0 \succsim_0 \omega'_0$ , which proves that  $\succsim_{n+1} \upharpoonright_{\hat{\Omega}_n} = \succsim_n$  when  $n = 0$ .

As our inductive hypothesis, we suppose that  $\succsim_n \upharpoonright_{\hat{\Omega}_{n-1}} = \succsim_{n-1}$ . Let  $\omega_n \succsim_n \omega'_n$ . Then, for all  $(P', \tilde{\omega}'_{n-1}) \in \omega'_n$ , there exists  $(P, \tilde{\omega}_{n-1}) \in \omega_n$  such that (i)  $P$  is finer than  $P'$ , and (ii)  $\tilde{\omega}_{n-1,s} \succsim_{n-1} \tilde{\omega}'_{n-1,s}$  for all  $s \in S$ . But by the induction hypothesis, this is equivalent to  $\tilde{\omega}_{n-1,s} \succsim_n \tilde{\omega}'_{n-1,s}$  for all  $s \in S$ , which implies that  $\omega_n \succsim_{n+1} \omega'_n$ .

Conversely, let  $\omega_n \succsim_{n+1} \omega'_n$ . Then, for all  $(P', \tilde{\omega}'_{n-1}) \in \omega'_n$ , there exists  $(P, \tilde{\omega}_{n-1}) \in \omega_n$  such that (i)  $P$  is finer than  $P'$ , and (ii)  $\tilde{\omega}_{n-1,s} \succsim_n \tilde{\omega}'_{n-1,s}$  for all  $s \in S$ . However, the induction hypothesis implies  $\tilde{\omega}_{n-1,s} \succsim_{n-1} \tilde{\omega}'_{n-1,s}$  for all  $s \in S$ , proving that  $\omega_n \succsim_n \omega'_n$  and therefore  $\succsim_{n+1} \upharpoonright_{\hat{\Omega}_n} = \succsim_n$ .  $\square$

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  $\omega, \omega' \in \hat{\Omega}_n$  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.

(49) 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)  $\omega \succsim \omega'$ .
- (b) for all  $(P', \tilde{\omega}') \in \omega'$ , there exists  $(P, \tilde{\omega}) \in \omega$  such that (i)  $P$  is finer than  $P'$ , and (ii)  $\tilde{\omega}_s \succsim \tilde{\omega}'_s$  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 exists  $n$  such that  $\omega, \omega' \in \hat{\Omega}_n$  and  $\omega \succsim_n \omega'$ . This implies that for all  $(P', \tilde{\omega}'_{n-1}) \in \omega'_n$ , there exists  $(P, \tilde{\omega}_{n-1}) \in \omega_n$  such that (i)  $P$  is finer than  $P'$ , and (ii)  $\tilde{\omega}_{n-1,s} \succsim_{n-1} \tilde{\omega}'_{n-1,s}$  for all  $s \in S$ . But the latter property implies  $\tilde{\omega}_s \succsim \tilde{\omega}'_s$  for all  $s \in S$ , which establishes (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  $n \geq 0$ .  $\square$

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

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

*Proof.* Because  $\Omega = \text{cl}(\hat{\Omega})$ , we simply extend  $\succsim$  to  $\Omega$  by re-defining it to be  $\text{cl}(\succsim)$ . It is easy to see that  $\succsim$  so defined is continuous and non-trivial. That  $\succsim$  is a unique recursive Blackwell order follows immediately from the facts that  $\hat{\Omega}$  is dense in  $\Omega$ , the continuity of  $\succsim$ , 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  $\hat{\omega}$ , ie, 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  $\omega, \omega' \in \Omega$ ,  $\omega \succsim \omega'$  if, and only if, for all  $n \in \mathbb{N}$ ,  $\omega_n \succsim \omega'_n$ .

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

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  $m \geq 1$ ,  $\omega_m \succsim \omega'_m$  implies  $\omega_n \succsim \omega'_n$  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  $n \geq 1$ ,  $\text{proj}_n \omega_m \succsim \text{proj}_n \omega'_m$ . For  $n \geq m$ ,  $\text{proj}_n \omega_m = \omega_m$ , but for  $n \leq m$ ,  $\text{proj}_n \omega_m = \omega_n$ , which implies that for all  $n \leq m$ ,  $\omega_n \succsim \omega'_n$ .  $\square$

## 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

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

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

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  $\Theta = \Omega$ ,  $\theta_0 = \omega$ ,  $\Gamma^*(\theta) = \{P : P \text{ is supported by } \theta\}$ , and  $\tau^*(P, \omega, s) = \omega'_s$  where  $\omega' \in \Omega^S$  is the unique collection of RICS such that  $(P, \omega') \in \omega$ . This results in the MIC  $\mathcal{M}_\omega = (\Theta, \Gamma^*, \tau^*, \theta_0 = \omega)$  that is uniquely determined by  $\omega$ .  $\square$

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  $\mathbf{M}$ . Define the function  $D : \mathbf{M} \times \mathbf{M} \rightarrow \mathbb{R}$  as follows:

$$\begin{aligned} & D(\mathcal{M}(\theta_0), \mathcal{M}'(\theta'_0)) := \\ \text{[A.1]} \quad & \max \left[ d_H(\Gamma(\theta_0), \Gamma'(\theta'_0)) \wedge 1, \frac{1}{2} \max_{P \in \Gamma(\theta_0), s \in S} D(\mathcal{M}(\tau(\theta_0, P, s)), \mathcal{M}'(\tau'(\theta'_0, P, s))) \right] \end{aligned}$$

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  $\mathbf{B}(\mathbf{M} \times \mathbf{M})$  denote the space of real-valued bounded functions defined on  $\mathbf{M} \times \mathbf{M}$  with the supremum norm.

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

*Proof.* Consider the operator  $T : \mathbf{B}(\mathbf{M} \times \mathbf{M}) \rightarrow \mathbf{B}(\mathbf{M} \times \mathbf{M})$  defined as

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

for all  $D' \in \mathbf{B}(\mathbf{M} \times \mathbf{M})$ . 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, ie,  $T(D + a) \leq TD + \frac{1}{2}a$  for all  $a \geq 0$ . This implies that  $T$  has a unique fixed point in  $\mathbf{B}(\mathbf{M} \times \mathbf{M})$ , 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  $\mathcal{M}, \mathcal{M}' \in \mathbf{M}$ . 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  $\text{proj}_n(\omega) \not\geq \text{proj}_n(\omega')$  for some  $n \geq 1$ , but for all  $m < n$ ,  $\text{proj}_m(\omega) \geq \text{proj}_m(\omega')$ . Then, there exists finite sequences  $(P_k)_{k=1}^{n-1}$  and  $(s_k)_{k=1}^{n-1}$  which induce RICS  $\omega_{(n-k)}^i := \tau^*(\omega_{(n-k+1)}^i, P_k, s_k) \in \hat{\Omega}_{n-k}$  where  $P_k \in \Gamma^*(\omega_{n-k+1}^i)$ , such that  $\Gamma^*(\omega_1^1)$  does not setwise-Blackwell dominate  $\Gamma^*(\omega_1^2)$ .

*Proof.* If not, we would have  $\omega_n^1 \geq \omega_n^2$ , a contradiction.  $\square$

Let  $\triangleright$  be a recursive order on  $\mathbf{M}$  defined as follows: For MICs  $\mathcal{M} = (\Theta, \Gamma, \tau, \theta_0)$  and  $\mathcal{M}' = (\Theta', \Gamma', \tau', \theta'_0)$ ,

[ $\boxtimes$ ]  $\mathcal{M} \triangleright \mathcal{M}'$  if for every  $P' \in \Gamma'(\theta'_0)$ , there exists  $P \in \Gamma(\theta_0)$  such that (i)  $P$  is finer than  $P'$ , and (ii)  $(\Theta, \Gamma, \tau, \tau(\theta_0, P, s)) \triangleright (\Theta', \Gamma', \tau', \tau'(\theta'_0, P', s))$  for all  $s \in S$ .

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

**Proposition A.11.** Let  $\triangleright$  be a recursive order on  $\mathbf{M}$  that satisfies [ $\boxtimes$ ]. If  $\mathcal{M} \triangleright \mathcal{M}'$ , then  $\mathcal{M}$  recursively Blackwell dominates  $\mathcal{M}'$ .

*Proof.* We will prove the contrapositive. If  $\mathcal{M}$  does not recursively Blackwell dominate  $\mathcal{M}'$ , then  $\omega_{\mathcal{M}} \not\geq \omega_{\mathcal{M}'}$ . Corollaries A.5 and A.6 imply that there is a smallest  $n$  such that  $\omega_{\mathcal{M},n} \not\geq \omega_{\mathcal{M}',n}$  but that for all  $m < n$ ,  $\omega_{\mathcal{M},m} \geq \omega_{\mathcal{M}',m}$  (where  $\omega_{\mathcal{M},n} = \text{proj}_n \omega_{\mathcal{M}}$  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  $\mathcal{M}$  is indistinguishable from  $\omega_{\mathcal{M}}$ , and so is  $\mathcal{M}'$  from  $\omega_{\mathcal{M}'}$ . 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  $\mathcal{M} \not\triangleright \mathcal{M}'$ .  $\square$



## 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  $u_s \in \mathbf{C}(C)$  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  $u_s(c_s^\dagger) = 0$  for all  $s \in S$  for some fixed  $c_s^\dagger \in C$ .

We show in Section 5 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 -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  $c_s^+$  and  $c_s^-$  must exist. Now, define  $f^+ \in \mathcal{F}(\Delta(C))$  to be the act such that  $f^+(s) := c_s^+$  — 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  $\alpha_1 \in \Delta(C)$ ,  $u_s(c_s^-) \leq u_s(\alpha_1) \leq u_s(c_s^+)$ . An immediate consequence of this is that for any  $c \in C$ ,  $\ell \in L$  and  $s \in S$ ,  $(c, \ell^*) \succsim_s (c, \ell) \succsim_s (c, \ell_*)$ . Lipschitz Continuity (Axiom 1(c)) implies that  $\ell^* \succ \ell_*$  (see Corollary 1.3 in the Supplementary Appendix), so  $(c, \ell^*) \succ_s (c, \ell_*)$ .

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

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

Suppose  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  is a tuple where

- $u_s \in \mathbf{C}(C)$  for all  $s \in S$ ,
- $\mathbb{Q} \subset \mathcal{P}(S)$ ,
- $v_s(\cdot, P) \in \text{BL}(X)$  for all  $s \in S$  and  $P \in \mathbb{Q}$ ,<sup>50</sup>

(50) The space  $\text{BL}(X)$  consists of all bounded Lipschitz functions on  $X$ ; see Appendix A.2.

- $\pi \in \Delta(S)$ ,
  - $u_s(c_s^\dagger) = v_s(\ell^\dagger, P) = 0$  for all  $s \in S$  and  $P \in \mathbb{Q}$ ,
  - $v_s(\cdot, P)$  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  $P \in \mathbb{Q}$ ,
- and  $v \in \mathbb{R}^X$  is such that

$$v(x) = \max_{P \in \mathbb{Q}} \sum_{E \in \mathcal{P}} \pi(E) \max_{f \in x} \sum_{s \in S} \pi(s | E) [u_s(f_1(s)) + v_s(f_2(s), P)]$$

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

More generally, for any subset  $\Phi \subset \mathbf{C}(X)$ , define the operator  $\mathbf{A} : 2^{\mathbf{C}(X)} \rightarrow 2^{\mathbf{C}(X)}$  as follows:

$$\mathbf{A}\Phi := \left\{ v \in \mathbf{C}(X) : \exists ((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi) \text{ that implements } v \right. \\ \left. \text{and } v_s(\cdot, P) \in \Phi \text{ for all } s \in S \text{ and } P \in \mathbb{Q} \right\}$$

**Proposition A.12.** The operator  $\mathbf{A}$  is well defined and has a largest fixed point  $\Phi^* \neq \{\mathbf{0}\}$ . Moreover,  $\Phi^*$  is a cone.

*Proof.* It is easy to see that for all nonempty  $\Phi \subset \mathbf{C}(X)$ ,  $\mathbf{A}\Phi$  is nonempty. (Simply take any  $\mathbb{Q}$ , any  $\mathbf{0} \neq v_s(\cdot, P) \in \Phi$  for all  $P \in \mathbb{Q}$ , and any  $u_s$ , so that  $\mathbf{A}\Phi \neq \emptyset$ .) The operator  $\mathbf{A}$  is *monotone* in the sense that  $\Phi \subset \Phi'$  implies  $\mathbf{A}\Phi \subset \mathbf{A}\Phi'$ . Thus, it is a monotone mapping from the lattice  $2^{\mathbf{C}(X)}$  to itself, where  $2^{\mathbf{C}(X)}$  is partially ordered by inclusion. The lattice  $2^{\mathbf{C}(X)}$  is *complete* because any collection of subsets of  $2^{\mathbf{C}(X)}$  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,  $\mathbf{A}$  has a largest fixed point  $\Phi^* \in 2^{\mathbf{C}(X)}$ .

To see that  $\Phi^* \neq \{\mathbf{0}\}$ , ie,  $\Phi^*$  does not contain only the trivial function  $\mathbf{0}$ , fix  $\mathbb{Q} = \{\{s\} : s \in S\}$  so that it contains only the finest partition of  $S$ . For the value function  $V$  in [Val], take any  $u_s \in \mathbf{C}(C) \setminus \{\mathbf{0}\}$  with  $u_s(c_s^\dagger) = 0$  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  $((u_s), \mathbb{Q}, \delta V, \pi)$ , while  $\delta V$  is implemented by  $((\delta u_s), \mathbb{Q}, \delta^2 V, \pi)$ , and so on. Therefore, the set  $\Phi_V := \{\delta^n V : n \geq 0\}$  is clearly a fixed point of  $\mathbf{A}$ . 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  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  implements  $v$ . Then, for all  $\lambda \geq 0$ ,  $((\lambda u_s), \mathbb{Q}, (\lambda v_s(\cdot, P)), \pi)$  implements  $\lambda v$ , ie,  $\lambda \Phi^*$  is also a fixed point of  $\mathbf{A}$ . Because  $\Phi^*$  is the largest fixed point, it must be a cone.  $\square$

Notice that each  $v \in \Phi^*$  is implemented by a tuple  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  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  $\varphi^* : \Phi^* \rightarrow \Omega$  that satisfies for some implementation  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  of  $v$ , that

$$\varphi^*(v) := \left\{ (P, \varphi^*(v_s(\cdot, P))) : P \in \mathbb{Q} \right\}$$

and is independent of the implementation chosen.

*Proof.* Let  $v^{(1)} \in \Phi_1$ , and suppose  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$  implements  $v^{(1)}$ . In this implementation,  $\mathbb{Q}$  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  $\varphi_1 : \Phi_1 \rightarrow \Omega_1$  as

$$\varphi_1(v^{(1)}) := \mathbb{Q}, \quad \text{where } ((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi) \text{ implements } v^{(1)}$$

Proceeding iteratively, we define  $\varphi_n : \Phi_n \rightarrow \Omega_n$  as

$$\varphi_n(v^{(n)}) := \left\{ (P, \varphi_{n-1}(v_s^{(n-1)}(\cdot, P))) : \exists ((u_s), \mathbb{Q}, (v_s^{(n-1)}(\cdot, P)), \pi) \text{ that} \right. \\ \left. \text{implements } v^{(n)} \text{ and } P \in \mathbb{Q} \right\}$$

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  $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$ , where each  $v_s(\cdot, P)$  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

$$\omega_0 := (\omega^{(1)}, \omega^{(2)}, \dots, \omega^{(n)}, \dots) \in \Omega$$

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

To recapitulate, we can now extract an RIC from a self-generating representation. In other words, the identification of the 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

$$\mathfrak{h}_n = ((x^{(0)}, \omega^{(0)}, \dots, (P^{(n-1)}, f^{(n-1)}, s^{(n-1)}, x^{(n-1)}, \omega^{(n-1)}))$$

Let  $\mathfrak{H}_n$  denote the collection of all  $n$ -period histories.

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

A dynamic plan  $\sigma = (\sigma_i, \sigma_c)$  with initial states  $x^{(0)} := x$  and  $\omega^{(0)} := \omega_0$  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  $s, s' \in I$ ,  $\sigma_c^{(n)}(h_n)(s) = \sigma_c^{(n)}(h_n)(s')$ .

Each dynamic plan along with initial states  $(x, \omega_0, \pi_0)$  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 in period  $n$ . A dynamic plan is *stationary* if  $\sigma^{(n)}(h_n)$  only depends on  $(x^{(n-1)}, \omega^{(n-1)}, s^{(n-1)})$ .<sup>51</sup>

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

$$[\text{A.2}] \quad \lim_{n \rightarrow \infty} \left\| \mathbf{E}^{\sigma, \pi} v^{(n)}(\cdot, \omega^{(n)}, s^{(n)}, \sigma) \right\|_\infty = 0$$

*Proof.* Consider  $V \in \Phi^*$  with Lipschitz rank  $\lambda$ . Recall that for any  $x \in X$ ,  $\ell^\dagger \diamond_n x \in X$  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  $n \geq N$ ,  $d(\ell^\dagger \diamond_n x, \ell^\dagger \diamond_n y) < \varepsilon/\lambda$ . Lipschitz continuity of  $V$  then implies  $|V(\ell^\dagger \diamond_n x) - V(\ell^\dagger \diamond_n y)| < \varepsilon$ .

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

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

for all  $n \geq N$ . Recall that

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

Moreover, we have

$$\begin{aligned} & \sup_x \left| \mathbf{E}^{\sigma, \pi} v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) \right| \\ &= \sup_x \left| \mathbf{E}^{\sigma, \pi} [v^{(n)}(x, \omega^{(n)}, s^{(n)}, \sigma) - v^{(n)}(\ell^\dagger, \omega^{(n)}, s^{(n)}, \sigma)] \right| < \varepsilon \end{aligned}$$

which completes the proof.  $\square$

(51) Of course, the choice of plan doesn't affect the evolution of the objective states ( $s^{(n)}$ ).

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  $\times_n \mathfrak{H}_n$ , the space of all histories. It also induces a *unique* consumption stream  $\ell_{\sigma(x, \omega)}$  that delivers consumption  $\sigma_c(\mathfrak{h}_n)(s')$  after history  $\mathfrak{h}_n$  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  $x$ .

**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  $V(x) = \max_{\sigma \in \Sigma} V(x, \sigma) = \max_{\ell \in L_{x, \omega}} V(\ell)$ .

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).<sup>52</sup>

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 49). 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).  $\square$

## 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  $(P, \omega') \in \omega$ .

(52) 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.

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

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

where  $\text{Unif}(\cdot)$  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

$$f_{n,J} := \begin{cases} (c_s^+, \text{Unif}_{n-1}(\{x_{n-1}(Q, \tilde{\omega}) : (Q, \tilde{\omega}) \in \omega'_s\})) & \text{if } s \in J \\ \ell_*(s) & \text{if } s \notin J \end{cases}$$

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 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  $J \in P$

$$f_J := \begin{cases} (c_s^+, \text{Unif}(\{x(Q, \tilde{\omega}) : (Q, \tilde{\omega}) \in \omega'_s\})) & \text{if } s \in J \\ \ell_*(s) & \text{if } s \notin J \end{cases}$$

as claimed. □

It is straightforward to verify that

$$V(\ell^*, \omega, \pi_0) = V(x(P, \omega'), \omega, \pi_0) \geq V(x(P, \omega'), \tilde{\omega}, \pi_0)$$

for all  $\tilde{\omega} \in \Omega$ . Indeed,  $V(x(P, \omega'), \omega, \pi_0) = V(x(P, \omega'), (P, \omega'), \pi_0)$ .

**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  $[\star]$  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  $[\star]$ . 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).) □



**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  $m < n$ ,  $\text{proj}_m(\omega')$  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  $\Gamma^*(\tau^{*(n)}(\omega', (P'_k), (s_k)))$  does not setwise Blackwell dominate the set  $\Gamma^*(\tau^{*(n)}(\omega, (P_k), (s_k)))$ , where  $\tau^{*(n)}(\theta_0, (P_k), (s_k))$  represents the  $n$ -stage transition following the sequence of choices  $(P_k)$  and states  $(s_k)$ ,  $\omega_{n-k}^i = \tau^*(\omega_{n-k+1}^i, P_k, s_k)$  where  $P_k \in \Gamma^*(\omega_{n-k+1}^i)$ , and  $\Gamma^*(\omega_1^1)$ .

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.  $\square$

*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  $\omega$ .  $\square$

This allows us to characterize the recursive Blackwell order in terms of the instrumental value of information.

**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.  $\square$

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  $\succ^{\dagger} \upharpoonright_L = \succ \upharpoonright_L$ . This, and the uniqueness of the RAA representation (Proposition 5.5) 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 follows immediately from Corollary B.4.  $\square$

## C. Existence

As always,  $\mathbf{C}(C \times X)$  is the space of all uniformly continuous functions on  $C \times X$  and for  $\alpha \in \Delta(C \times X)$  and  $u \in \mathbf{C}(C \times X)$ ,  $u(\alpha) := \int_{C \times X} u(c, x) d\alpha(c, x) =: \langle \alpha, u \rangle$ . For each  $s \in S$ , fix  $\ell^{\dagger}(s) \in \Delta(C \times X)$ , and define  $\mathfrak{U}_{s, \ell^{\dagger}(s)} := \{u_s \in \mathbf{C}(C \times X) : u_s(\ell^{\dagger}(s)) = 0, \|u_s\|_{\infty} = 1\}$ . Finally, define  $\mathfrak{U} := \{(p_1 u_1, \dots, p_n u_n) : u_s \in \mathfrak{U}_{s, \ell^{\dagger}(s)}, p_i \geq 0, \sum_i p_i = 1\}$ . The space  $\mathfrak{U}$  will serve as our

subjective state space below. It is useful to reconsider  $\mathfrak{U}$  as  $\mathfrak{U} := \{(p, u) : p := (p_1, \dots, p_n) \in \Delta(S), u := (u_1, \dots, u_n) \in \prod_{s \in S} \mathfrak{U}_{s, \ell^\dagger(s)}\}$ .

Throughout this section we assume that  $\succsim$  is a binary relation on  $X$  and has a static representation  $V : X \rightarrow \mathbb{R}$  as follows:

$$[\diamond] \quad V(x) := \max_{\mu \in \mathfrak{M}} \left[ \int_{\mathfrak{U}} \max_{f \in x} \sum_s p_s u_s(f(s)) \, d\mu(p, u) \right]$$

where the set  $\mathfrak{M} \subset ba_n(\mathfrak{U})$  is weak\* compact<sup>53</sup> and  $\int_{\mathfrak{U}} \max_{\alpha \in x} \sum_i p_i u_i(\alpha_i) \, d\mu(p, u)$  is independent of  $\mu$  for all  $\ell \in L$ . Theorem 1 of Section 1 in the Supplementary Appendix shows that  $\succsim$  satisfies Basic Properties (Axiom 1) and Axiom 2 (a) if, and only if, it has a representation of the form  $[\diamond]$ .

For each  $\mu \in \mathfrak{M}$ , let  $V(x, \mu) := \int_{\mathfrak{U}} \max_{f \in x} \sum_s p_s u_s(f(s)) \, d\mu(p, u)$  be the utility from choosing the measure  $\mu$ . Let  $\mathcal{Y} : X \rightrightarrows \mathfrak{M}$  be the mapping selecting the maximizing  $\mu$  for each  $x$ ; that is,  $\mathcal{Y}(x) := \arg \max_{\mu \in \mathfrak{M}} V(x, \mu)$ . It is easy to see that  $V(x, \mu)$  is continuous in  $\mu$ , so it follows that  $\mathcal{Y}$  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  $x \in X$ ,  $\ell \in L$  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  $\mathfrak{M}$ , eventually leading us to the desired representation theorem.

### C.1. Partitional Representation

In this section, we consider the representation in  $[\diamond]$  of  $\succsim$  and impose Indifference to Incentivized Contingent Commitment (henceforth IICC, Axiom 4).

The main consequence of assuming IICC (Axiom 4) is that instead of considering arbitrary finitely additive measures  $\mu \in \mathfrak{M}$  over  $\mathfrak{U}$  in the representation  $[\diamond]$ , 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  $u \in \mathbf{C}(C \times X)$ .

**Proposition C.1.** Consider a preference relation  $\succsim$  on  $X$ , and suppose  $V : X \rightarrow \mathbb{R}$  represents  $\succsim$  and has the form in  $[\diamond]$ . Then, the following are equivalent:

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

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

where  $\mathfrak{M}_p$  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  $C \times X$  with the property that for all  $s \in S$ ,  $u_s(\alpha) = u'_s(\alpha)$  for all  $(P, u), (P', u') \in \mathfrak{M}_p$  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  $\mathfrak{U}$ , both of which are countably additive. Thus, an

(53) Here,  $ba_n(\mathfrak{U})$  is the space of bounded additive (or finitely additive) measures (ie, charges) on  $\mathfrak{U}$  that are also normal (ie, inner and outer regular).

essential part of the proof of Proposition C.1 is to show that ICC (Axiom 4) allows us to replace each  $\mu \in \mathfrak{M}$  by a countably additive measure without affecting the representation. The proof is lengthy precisely due to the complications that arise from dealing with  $\mu \in \mathfrak{M}$  in  $\blacklozenge$  that are finitely additive. If we knew beforehand that each  $\mu$  was countably additive, the proof would simply formalize the intuition behind ICC (Axiom 4) and be considerably shorter. The complete proof can be found in Section 6 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  $V : X \rightarrow \mathbb{R}$  represents  $\succsim$  and takes the form [C.1]. For each  $(P, u)$ , define

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

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

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

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

and observe that  $u'_s : \Delta(C \times X) \rightarrow \mathbb{R}$  is continuous and linear<sup>54</sup> so that  $u'_s \in \mathbf{C}(C \times X)$ . Moreover, for all  $(P, u'), (\hat{P}, \hat{u}') \in \mathfrak{M}_p$ ,  $u'_s|_{C \times L} = \hat{u}'_s|_{C \times L}$ . This implies that  $V(\ell, (P, u'))$  is independent of  $(P, u') \in \mathfrak{M}'_p$ .

Now define  $V' : X \rightarrow \mathbb{R}$  as

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

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  $\succsim$ .

**Lemma C.2.** 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  $P = \{J_1, \dots, J_m\}$ , for  $i = 1, \dots, m$  we can define the acts

$$f_i := \arg \max_{f \in x} \sum_s \pi_0(s | J_i) u_s(f(s))$$

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

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

$$\begin{aligned} V(x) &= V(\{f_1, \dots, f_m\}) && \text{definition of } \{f_1, \dots, f_m\} \\ &= V(\{\hat{f}_1, \dots, \hat{f}_m\}) && \text{established in (a) above} \\ &= V(\{\hat{f}_1, \dots, \hat{f}_m\}, (\hat{P}, \hat{u}')) && \text{established in (c) above} \\ &= V(\{\hat{f}_1, \dots, \hat{f}_m\}, (\hat{P}, \hat{u}')) && \text{established in (b) above} \\ &= V(\{f_1, \dots, f_m\}, (\hat{P}, \hat{u}')) && \text{established (d) above} \\ &\leq V'(\{f_1, \dots, f_m\}) && \text{definition of } V' \\ &\leq V'(x) && \text{monotonicity of } V' \end{aligned}$$

which completes the proof. □

We can now state the main result of this section.

**Proposition C.3.** Let  $V$  be as in [C.1] and suppose  $V$  represents  $\succsim$ . Then, the following are equivalent.

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

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

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

Lemma C.2 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  $(P, u), (P', u') \in \mathfrak{M}'_p$  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

- $u_s(\alpha_1) := u_s(\alpha_1 \otimes \ell_2^\dagger(s))$  (and notice  $u_s(\alpha) = u'_s(\alpha)$  for all  $(P, u), (P', u') \in \mathfrak{M}'_p$  and for all  $\alpha \in \Delta(C \times L)$  because  $\alpha_1 \otimes \ell_2^\dagger(s) \in \Delta(C \times L)$ ); and

- $v_s(\alpha_2) := u_s(\ell_1^\dagger(s) \otimes \alpha_2)$ .

With these definitions,  $u_s \in \mathbf{C}(C)$  while  $v_s(\cdot) \in \mathbf{C}(X)$ . Notice that the lotteries  $\frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^\dagger(s)$  and  $\frac{1}{2}(\alpha_1 \otimes \ell_2^\dagger(s)) + \frac{1}{2}(\ell_1^\dagger(s) \otimes \alpha_2)$  have identical marginals, which implies that for every  $(P, u)$ ,

$$u_s \left( \frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^\dagger(s) \right) = u_s \left( \frac{1}{2}(\alpha_1 \otimes \ell_2^\dagger(s)) + \frac{1}{2}(\ell_1^\dagger(s) \otimes \alpha_2) \right)$$

This means we can write

$$\begin{aligned} \frac{1}{2}u_s(\alpha_1 \otimes \alpha_2) + \frac{1}{2}u_s(\ell^\dagger(s)) &= u_s \left( \frac{1}{2}(\alpha_1 \otimes \alpha_2) + \frac{1}{2}\ell^\dagger(s) \right) = u_s \left( \frac{1}{2}(\alpha_1 \otimes \ell_2^\dagger(s)) + \frac{1}{2}(\ell_1^\dagger(s) \otimes \alpha_2) \right) \\ &= \frac{1}{2}u_s(\alpha_1 \otimes \ell_2^\dagger(s)) + \frac{1}{2}u_s(\ell_1^\dagger(s) \otimes \alpha_2) = \frac{1}{2}u_s(\alpha_1) + \frac{1}{2}v_s(\alpha_2) \end{aligned}$$

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  $u_s(\ell^\dagger(s)) = 0$ , 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.2 gives us [C.3], as desired.  $\square$

As always, for each  $(P, (v_s)) \in \mathfrak{M}_p''$ , define  $V(x, (P, (v_s)))$  as

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

### 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

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

where  $\mathfrak{M}_p^\#$  is a finite collection of partitions  $P$  of  $S$ ,  $u_s \in \mathbf{C}(C)$ , and  $v_s(\cdot, P) \in \mathbf{C}(X)$  for each  $s \in S$  and  $P \in \mathfrak{M}_p^\#$ , with the property that for all  $P, P' \in \mathfrak{M}_p^\#, s \in S, v_s(\cdot, P)|_L = v_s(\cdot, P')|_L$ .

For a fixed  $P$  in the representation in [C.3], let  $X'_P$  and  $\hat{X}_P$  be defined as follows:

$$\begin{aligned} X'_P &:= \{x : V(x) = V(x, (P, (v_s))) \text{ for some } (P, (v_s)) \in \mathfrak{M}_p'' \text{ and} \\ &\quad V(x) > V(x, (Q, (v'_s))) \text{ for all } (Q, (v'_s)) \in \mathfrak{M}_p'' \text{ such that } P \neq Q\} \\ \hat{X}_P &:= \{x : V(x) = V(x, (P, (v_s))) \text{ for some } (P, (v_s)) \in \mathfrak{M}_p'' \text{ and} \\ &\quad V(x) \geq V(x, (Q, (v'_s))) \text{ for all } (Q, (v'_s)) \in \mathfrak{M}_p'' \text{ such that } P \neq Q\} \end{aligned}$$

Recall that  $x_1(P) := x(P, \hat{\omega})$  as in [A] 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.4.** Let  $x \in X'_P$ . Then, for all  $\lambda \in (0, 1)$ ,  $(1 - \lambda)x + \lambda x_1(P) \in X'_P$ . Moreover,  $V((1 - \lambda)x + \lambda x_1(P)) = V((1 - \lambda)x + \lambda \ell^*) > V((1 - \lambda)x + \lambda x_1(Q))$  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 RAA representation of Appendix A.7,  $[\ell^* \oplus_{(1, S \setminus s)} \ell_*] \succ [\ell_* \oplus_{(1, S \setminus s)} \ell_*] = \ell_*$  for all  $s \in S$ . Because we have  $v_s(\ell) = v'_s(\ell)$  for all  $\ell \in L$  and  $(P, (v_s)), (P', (v'_s)) \in \mathfrak{M}''_P$  in [C.3],  $v_s(\ell^*) > v_s(\ell_*)$  follows for all  $s \in S$ .

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

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

$$\begin{aligned} V(x_1(Q), (P, (v_s))) &= \sum_{J \in P} \pi_0(J) \max_{f \in x_1(Q)} \sum_s \pi_0(s | J) [u_s(f_1(s)) + v_s(f_2(s))] \\ &= \sum_{J \in P} \pi_0(J) \sum_s \pi_0(s | J) [u_s(c_s^+) + v_s(\ell^*)] \\ &= V(\ell^*, (P, (v_s))) = V(\ell^*) \end{aligned}$$

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

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

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

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

$$\begin{aligned} V((1 - \lambda)x + \lambda x_1(P), (P, (v_s))) &= (1 - \lambda)V(x) + \lambda V(x_1(P), (P, (v_s))) \\ &= (1 - \lambda)V(x) + \lambda V(\ell^*) = V((1 - \lambda)x + \lambda \ell^*) \end{aligned}$$

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

$$\begin{aligned} V((1 - \lambda)x + \lambda x_1(P), (Q, (v'_s))) &= (1 - \lambda)V(x, (Q, (v'_s))) + \lambda V(x_1(P), (Q, (v'_s))) \\ &< (1 - \lambda)V(x) + \lambda V(\ell^*) \end{aligned}$$

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  $(1 - \lambda)x + \lambda x_1(P) \in X'_P$ . Moreover, it now follows immediately that  $V((1 - \lambda)x + \lambda x_1(P)) = V((1 - \lambda)x + \lambda \ell^*)$ .

Finally, suppose  $P$  is not finer than  $Q$ . Consider the menu  $(1 - \lambda)x + \lambda x_1(Q)$  and suppose  $(P', \cdot) \in \mathfrak{M}''_P$  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,

$$\begin{aligned} V((1-\lambda)x + \lambda x_1(Q)) &= V((1-\lambda)x + \lambda x_1(Q), (P', \cdot)) \\ &= (1-\lambda)V(x, (P', \cdot)) + \lambda V(x_1(Q), (P', \cdot)) \\ &< (1-\lambda)V(x, (P, \cdot)) + \lambda V(\ell^*) = V((1-\lambda)x + \lambda \ell^*) \end{aligned}$$

which completes the proof.  $\square$

**Lemma C.5.**  $X'_P$  is convex and consists of concordant RACPS.

*Proof.* Because  $V$  is  $L$ -affine, any  $(P, (v_s))$  that is optimal for  $x$  is also optimal for  $(1-t)x + t\ell$  for all  $t \in [0, 1]$  and  $\ell \in L$ , and vice versa. Thus,  $x \in X'_P$  if, and only if,  $(1-t)x + t\ell \in X'_P$ .

Let  $x, y \in X'_P$  be such that  $x \succsim y$ . It follows from IICC (Axiom 4) that  $x \succsim y \succsim \ell_*$ . By Continuity of  $\succsim$  (Axiom 1 (b)), there exists a  $t \in [0, 1]$  such that  $(1-t)x + t\ell_* \sim y$  and  $(1-t)x + t\ell_* \in X'_P$  (as we observed above). Thus, it is without loss of generality to consider  $x, y \in X'_P$  such that  $x \sim y$ .

By Lemma C.4,  $(P, (\tilde{v}_s)) \in \mathfrak{M}''_P$  remains optimal for both  $(1-\lambda)x + \lambda x_1(P)$  and  $(1-\lambda)y + \lambda x_1(P)$ , for all  $\lambda \in (0, 1)$ . It follows that  $x$  and  $y$  are  $\lambda$ -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  $(Q, (v'_s)) \in \mathfrak{M}''_P$  is optimal for  $\frac{1}{2}x + \frac{1}{2}y$ . Then,  $V(x, (Q, (v'_s))) \leq V(x)$  and  $V(y, (Q, (v'_s))) \leq V(y)$ . Because  $x, y$ , and  $\frac{1}{2}x + \frac{1}{2}y$  are  $\lambda$ -concordant,  $V(x) = V(y) = V(\frac{1}{2}x + \frac{1}{2}y)$ , ie,  $(Q, (v'_s))$  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 that  $Q = P$  for any maximizer  $(Q, (v'_s))$  at  $z$ , ie,  $X'_P$  is convex.  $\square$

**Lemma C.6.** For each  $x \in X$ , there exists  $(P, (v_s)) \in \mathfrak{M}''_P$  such that  $x \in \text{cl}(X'_P)$ .

*Proof.* Let  $x \in \hat{X}_{P_1} \cap \dots \cap \hat{X}_{P_n}$  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  $P_2, \dots, P_n$  are finer than  $P_1$ . In analogy to the arguments in the proof of Lemma C.4, we find that  $V((1-\lambda)x + \lambda x_1(P_1), (P_1, (v_s^1))) = V((1-\lambda)x + \lambda \ell^*) > V((1-\lambda)x + \lambda x_1(P_1), (P_i, (v_s^i)))$  for some  $v_s^1$  with  $(P, (v_s^1)) \in \mathfrak{M}''_P$  and all  $(P_i, (v_s^i)) \in \mathfrak{M}''_P$  for  $i = 2, \dots, n$ . That is,  $(1-\lambda)x + \lambda x_1(P_1) \in X'_{P_1}$  for all  $\lambda \in (0, 1)$ , which implies  $x \in \text{cl}(X'_{P_1})$  as claimed.  $\square$

**Lemma C.7.** Let  $x \in X'_P$  and let  $Y_x$  denote the set of RACPS 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.5.) If  $P$  is not finer than  $Q$ , by Lemma C.4,  $(1-\lambda)y + \lambda x_1(Q) \succ (1-\lambda)x + \lambda x_1(Q)$ , which contradicts our assumption that  $x$  and  $y$  are concordant. Conversely, if  $Q$  is not finer than  $P$ , then an analogous argument establishes that  $(1-\lambda)x + \lambda x_1(P) \succ (1-\lambda)y + \lambda x_1(P)$ , 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.5.  $\square$

Notice that replacing  $\mathfrak{M}_p''$  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  $\mathfrak{M}_p''$  is weak\*-compact.

**Lemma C.8.** Let  $x \in \text{cl}(X'_p)$ . Then, there exists  $(v_s)$  such that  $(P, (v_s)) \in \mathfrak{M}_p''$  is optimal for all  $y \in \text{cl}(X'_p)$ .

*Proof.* By Lemma C.7,  $Y_x \subset X'_p$  which, by Lemma C.5, is convex. By Concordant Independence,  $\succsim|_{X'_p}$  satisfies Independence. That is,  $V|_{X'_p}$  is linear. It follows from Lemma 2.5 in the Supplementary Appendix that there exists  $(v_s)$  such that  $(P, (v_s))$  is optimal for all  $x \in X'_p$ . Continuity now implies that  $(P, (v_s))$  is optimal for all  $x \in \text{cl}(X'_p)$ .  $\square$

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

**Corollary C.9.** Let  $V$  be as in [C.3] and suppose  $V$  represents  $\succsim$ . Then, the following are equivalent.

- (a)  $\succsim$  satisfies Concordant Independence (Axiom 5).
- (b)  $V$  can be written as in [C.4].

As always, for any partition  $P \in \mathfrak{M}_p^\#$ , we define

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

## 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  $\Phi^*$ ). Starting from the representation in [C.4], we show in this section that imposing Self-Generation (Axiom 6) on  $\succsim$  implies it has a self-generating representation.

**Proposition C.10.** Let  $\succsim$  be a binary relation on  $X$ . Then, the following are equivalent.

- (a)  $\succsim$  satisfies Axioms 1–6.
- (b)  $\succsim$  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  $\succsim_{(x,s)}$  from Definition 3.2 is well defined. We begin with a preliminary lemma.

**Lemma C.11.** Let  $x = \{f_1, \dots, f_m\}$ , and  $x' = \{f'_1, f'_2, \dots, f'_m\}$ . Suppose  $d(f_i, f'_i) < \varepsilon$ . Then,  $d(x, x') < \varepsilon$ .

*Proof.* Recall that  $d(f_i, x') := \min_j d(f_i, f'_j) < \varepsilon$ . Therefore,  $\max_{f_i \in x} d(f_i, x') < \varepsilon$ . A similar calculation yields  $\max_{f'_j \in x'} d(f'_j, x) < \varepsilon$ , which implies that  $d(x, x') < \varepsilon$  from the definition of the Hausdorff metric.  $\square$

Notice that  $\mathfrak{M}_p^\#$  in [C.4] is finite and can be taken to be minimal (in the sense that if  $\mathfrak{N}_p^\#$  is another set that represents  $V$  as in [C.4], then  $\mathfrak{M}_p^\# \subset \mathfrak{N}_p^\#$ ) without affecting the representation.

**Lemma C.12.** Let  $\succsim$  have a representation as in [C.4]. For all  $P \in \mathfrak{M}_p^\sharp$ , there exists a finite  $x \in X'_p \cap X^*$ , where  $X'_p$  is defined in Section C.3.

*Proof.* The finiteness and minimality of  $\mathfrak{M}_p^\sharp$  in [C.4] implies that for any  $P \in \mathfrak{M}_p^\sharp$ , there exists an open set  $O \subset X'_p$ . Because the space  $X^*$  is dense in  $X$ , there exists  $x \in O \cap X^*$ .  $\square$

**Lemma C.13.** Let  $\succsim$  have a representation as in [C.4]. For all  $P \in \mathfrak{M}_p^\sharp$ ,  $v_s(y, P) \geq v_s(\ell_*, P)$ .

*Proof.* Suppose instead that  $v_s(y, P) < v_s(\ell_*, P)$ . Consider  $x \in X'_p \cap X^*$  which exists by Lemma C.12. Then, for  $\varepsilon > 0$  small enough such that  $P$  remains optimal,  $[x \oplus_{\varepsilon, s} \ell_*] \succ [x \oplus_{\varepsilon, s} y]$ . To see this, suppose  $f \oplus_{\varepsilon, s} y$  is chosen optimally from the menu  $x \oplus_{\varepsilon, s} y$ . Then,  $v_s(y, P) < v_s(\ell_*, P)$  implies

$$\begin{aligned} (1 - \varepsilon)[u_s(f_1(s)) + v_s(f_2(s), P)] + \varepsilon[u_s(c_s^-) + v_s(y, P)] \\ < (1 - \varepsilon)[u_s(f_1(s)) + v_s(f_2(s), P)] + \varepsilon[u_s(c_s^-) + v_s(\ell_*, P)] \end{aligned}$$

This implies  $V(x \oplus_{\varepsilon, s} \ell_*) > V(x \oplus_{\varepsilon, s} y)$ . But this contradicts part (a) of ICC (Axiom 4), which requires that  $[x \oplus_{(\varepsilon, s)} y] \succsim [x \oplus_{(\varepsilon, s)} \ell_*]$  for all  $y \in X$ .  $\square$

**Lemma C.14.** Let  $\succsim$  have a representation as in [C.4]. Fix  $P \in \mathfrak{M}_p^\sharp$ . For any finite  $x \in X'_p$  and  $s \in S$ ,  $\succsim_{(x, s)}$  is independent of the choice of  $\varepsilon \in (0, 1)$  for which Definition 3.2 applies. In particular,  $\succsim_{(x, s)}$  is represented by  $v_s(\cdot, P)$  on any  $X$ . Finally, if  $x'$  is finite, has a unique optimal partition, and is concordant with  $x$ , then  $\succsim_{(x, s)} = \succsim_{(x', s)}$ .

*Proof.* Let  $x \in X'_p$  be finite, so that  $V(x) = V(x, P)$ . Fix  $s \in S$ . Because  $V$  in [C.4] is continuous, there is  $\varepsilon > 0$  such that  $P$  is the unique optimal partition for all  $x' \in B(x; \varepsilon)$ , and hence all  $x', x'' \in B(x; \varepsilon)$  are concordant with each other (see Lemma C.7). By Lemma C.11,  $[x \oplus_{(\varepsilon, s)} y], [x \oplus_{(\varepsilon, s)} y'] \in B(x; \varepsilon)$ . Then,  $[x \oplus_{(\varepsilon, s)} y] \succsim [x \oplus_{(\varepsilon, s)} y']$  if, and only if,  $V(x \oplus_{(\varepsilon, s)} y) \geq V(x \oplus_{(\varepsilon, s)} y')$ . Suppose  $f \oplus_{\varepsilon, s} y$  is optimally chosen from  $x \oplus_{\varepsilon, s} y$  in the state  $s$ . Then, it must be that

$$\begin{aligned} (1 - \varepsilon)[u_s(f_1(s)) + v_s(f_2(s), P)] + \varepsilon[u_s(c_s^-) + v_s(y, P)] \\ \geq (1 - \varepsilon)[u_s(f_1(s)) + v_s(f_2(s), P)] + \varepsilon[u_s(c_s^-) + v_s(y', P)] \end{aligned}$$

which implies  $v_s(y, P) \geq v_s(y', P)$ . Conversely,  $v_s(y, P) \geq v_s(y', P)$  implies that if  $f \oplus_{\varepsilon, s} y'$  is optimally chosen from  $x \oplus_{\varepsilon, s} y'$  in state  $s$ , then the inequality displayed above holds, which implies  $[x \oplus_{(\varepsilon, s)} y] \succsim [x \oplus_{(\varepsilon, s)} y']$ . But this is independent of our choice of  $\varepsilon > 0$  as long as it maintains concordance.

Finally, if  $x$  and  $x'$  are concordant and  $x'$  has a unique optimal partition, then by Lemma C.7  $x' \in X'_p$ . It follows that  $\succsim_{(x', s)}$  is also represented by  $v_s(\cdot, P)$ , and hence  $\succsim_{(x, s)} = \succsim_{(x', s)}$ , which completes the proof.  $\square$

**Lemma C.15.** The binary relation  $\succsim_{(P, s)}$  on  $X$  which is represented by  $v_s(\cdot, P)$  satisfies Axioms 1–6.

*Proof.* By Lemma C.14,  $\succsim_{(P, s)} = \succsim_{(x, s)}$  for some  $x \in X'_p$ . By Self-Generation (Axiom 6),  $\succsim_{(x, s)}$  satisfies Axioms 1–6 on  $X$ .  $\square$

Before we prove Proposition C.10, an interlude.

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

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

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

**Lemma C.16.** The set  $\tilde{X} \subset X$  is dense in  $X$ .

*Proof.* Recall that  $X$  is the space of all *consistent* sequences in  $\prod_{n=1}^{\infty} X_n$ , 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  $x = (x_1, x_2, \dots, x_n, \dots)$ , and the metric on  $X$  is the product metric.

For any  $x = (x_1, x_2, \dots) \in X$  and  $n \geq 1$  set  $\tilde{x}_n \in \tilde{X}_n$  to be  $x_n$  concatenated with  $\ell_*$ . It follows from the product metric on  $X$  — see Appendix A.2 — that for any  $\varepsilon > 0$ , there exists  $n \geq 1$  such that  $d(x, \tilde{x}_n) < \varepsilon$ , as claimed.  $\square$

**Lemma C.17.** Let  $\succsim$  satisfy Axioms 1–5. Then, for any  $s \in S$  and  $P \in \mathfrak{M}_P^\#$ ,  $\ell^* \succsim_{(P,s)} \ell \succsim_{(P,s)} \ell_*$  for all  $\ell \in L$ .

*Proof.* The preference  $\succsim$  has a separable and partitional representation as in [C.4]. Therefore,  $\succsim_s$  on  $L$  is represented by  $u_s(\cdot) + v_s(\cdot, Q)$  for all  $Q$ . Moreover,  $\succsim|_L$  has an RAA representation. As observed in Section A.7,  $\succsim_s$  on  $L$  is separable and has the property that for all  $c \in C$ ,  $\ell \in L$  and  $s \in S$ ,  $(c, \ell^*) \succsim_s (c, \ell) \succsim_s (c, \ell_*)$ . 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  $Q \in \mathfrak{M}_P^\#$  in the representation [C.4]. But  $v_s(\cdot, P)$  represents  $\succsim_{(P,s)}$  which implies that  $\ell^* \succsim_{(P,s)} \ell \succsim_{(P,s)} \ell_*$  for all  $\ell \in L$ ,  $s \in S$ .  $\square$

**Proposition C.18.** Let  $\succsim$  satisfy Axioms 1–6. Then, for all  $x \in X$ ,  $\ell^* \succsim x$ .

*Proof.* By the continuity of  $\succsim$  and by Lemma C.16, it suffices to show that for all  $\tilde{x} \in \tilde{X}$ ,  $\ell^* \succsim \tilde{x}$ .

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 the representation in [C.4] is equivalent to  $\succsim$  satisfying Axioms 1–5. But  $\succsim$  satisfies Axiom 6, so that  $\succsim_{(P,s)}$  also satisfies Axioms 1–6 for any  $P \in \mathfrak{M}_P^\#$ , which implies that there exists  $\ell_{(P,s)}^*$  such that  $v_s(\ell_{(P,s)}^*) \geq v_s(\tilde{x}_1, P)$  for all  $\tilde{x}_1 \in \tilde{X}_1$ . By Lemma C.17, we may take  $\ell_{(P,s)}^* = \ell^*$ , so that  $v_s(\ell^*, P) \geq v_s(\tilde{x}_1, P)$  for all  $\tilde{x}_1 \in \tilde{X}_1$ .

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

Because  $\succsim_{(P,s)}$  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.17 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  $P$ ,

$$\begin{aligned} V(\tilde{x}_{n+1}, P) &= \max_{f \in \tilde{X}_{n+1}} \sum_{J \in P} \pi_0(s | J) [u_s(f_1(s)) + v_s(f_2, P)] \\ &\leq \sum_{J \in P} \pi_0(s | J) [u_s(c_s^+) + v_s(\ell^*, P)] \\ &= V(\ell^*, P) = V(\ell^*) \end{aligned}$$

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  $\tilde{x}_{n+1} \in \tilde{X}_{n+1}$ ,  $\ell^* \succ \tilde{x}_{n+1}$ , which completes the proof.  $\square$

#### C.4.2. Proof of Proposition C.10

*Proof.* To see that (b) implies (a), suppose  $\succsim$  has the representation [C.4]. By Proposition C.9  $\succsim$  satisfies Axioms 1–5. All that remains to establish is that  $\succsim$  also satisfies Axiom 6.

Given a representation as in [C.4] that is *also* self-generating, let  $x \in X$  be finite and  $P \in \mathfrak{M}_p^\#$  be an optimal partition for  $x$ . Observe first that if  $x \in X'_P$ , then by Lemma C.14  $\succsim_{(x,s)}$  is represented by  $v_s(\cdot, P)$ . Because the representation is self-generating,  $\succsim_{(x,s)}$  must satisfy Axioms 1–5 on  $X$ .

In general,  $P$  may not be uniquely optimal for  $x$ . By definition,  $\succsim_{(x,s)}$  is complete on  $X$  only if for all  $y, y' \in X$  there is  $\varepsilon \in (0, 1]$  such that  $[x \oplus_{(\varepsilon,s)} y]$ ,  $[x \oplus_{(\varepsilon,s)} y']$ , and  $x$  are pairwise concordant.

As in the proof of Lemma C.7 we assume, without loss of generality, that  $[x \oplus_{(\varepsilon,s)} y] \sim [x \oplus_{(\varepsilon,s)} y']$ . To build intuition, suppose the only optimal partitions for  $x$  are  $P, Q \in \mathfrak{M}_p^\#$  with  $P \neq Q$ . Suppose, further, that  $P$  is optimal for  $x \oplus_{(\varepsilon,s)} y$  and  $Q$  is not, while  $Q$  is optimal for  $x \oplus_{(\varepsilon,s)} y'$  and  $P$  is not. Again without loss of generality, suppose that  $Q$  is not finer than  $P$ . In that case

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

violating concordance of  $x \oplus_{(\varepsilon,s)} y$  and  $x \oplus_{(\varepsilon,s)} y'$ . 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  $\mathfrak{M}_p^\#$  such that for every  $y \in X$  there is  $\varepsilon > 0$  with  $P$  optimal for  $x \oplus_{(\varepsilon,s)} y$ . Therefore,  $\succsim_{(x,s)}$  is represented on  $X$  by  $v_s(\cdot, P)$  for some  $P \in \mathfrak{M}_p^\#$ . 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.15 has two implications. First,  $\succsim_{(P,s)}$  has a separable and partitional representation  $v'_s(\cdot, P)$  as in [C.4]. Because  $v_s(\cdot, P)$  also represents  $\succsim_{(P,s)}$  it follows that  $v_s(\cdot, P)$  and  $v'_s(\cdot, P)$  are identical up to a monotone transformation. But, by L-Indifference to Timing (Axiom 2(d)), it must be that  $v_s(\cdot, P)$  and  $v'_s(\cdot, P)$  are unique up to a positive affine transformation on  $L$ . Let us re-normalize  $v'_s(\cdot, P)$  so that  $v_s(\cdot, P) = v'_s(\cdot, P)$  on  $L$ .

Second, because  $\succsim_{(P,s)}$  satisfies Axioms 1–6, it satisfies the hypotheses of Proposition C.18. Together with Lemma C.17 and IICC (Axiom 4), this implies that  $\ell^* \succ_{(P,s)} y \succ_{(P,s)} \ell_*$  for all  $y \in X$ . Because  $v_s(\cdot, P)$  and  $v'_s(\cdot, P)$  both represent  $\succsim_{(P,s)}$ , they must agree on  $X$  because they agree on  $L$ . It follows that  $v_s(\cdot, P)$  also has a representation as in [C.4], that is, it can be written as

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

Then, because  $\succsim_{(x,s)}$  satisfies Axioms 1–6, it follows from the reasoning above that each  $v'_s(\cdot, P')$  in the above representation of  $v_s(\cdot, P)$  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  $((u_s), \delta, \Pi)$ . That is, there exist functions  $V_L^*(\cdot, s) : L \rightarrow \mathbb{R}$  such that  $V_L^*(\ell, \pi_0) := \sum_s \pi_0(s) V_L^*(\ell, s)$  represents  $\succsim|_L$ , and

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

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.10  $\succsim$  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  $\succsim|_L$  on  $L$ . Because  $\succsim|_L$  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  $L$ . We 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.19.** Let  $V$  be a self-generating representation of  $\succsim$  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  $X$ .<sup>55</sup>

*Proof.* In this proof, we frequently refer to objects defined in Appendix A.8. 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  $\succsim$  and  $V \in \Phi^*$ . Then, there exists an implementation of  $V$ , given by  $((u_s), \mathbb{Q}, (v_s^{(1)}(\cdot, P)), \pi)$ . For ease of exposition, we shall say that the collection  $(v_s^{(1)}(\cdot, P))$  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 ric, and so is entirely forward looking.

(55) It follows immediately from Proposition C.19 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).



### D. Proof of Proposition 4.3

Let  $V(F, P)$  be the consumption value  $F$  generates under partition  $P$ . Let  $\mathcal{P}_{\mathfrak{M}}$  be the set of partitions accessible under  $\mathfrak{M}$ . Since  $S$  is finite, so is  $\mathcal{P}_{\mathfrak{M}}$ . Therefore, there is  $\hat{T} < \infty$  by which any  $P \in \mathcal{P}_{\mathfrak{M}}$  can be reached via some information plan. By part (a) of Definition 4.2, for any such  $P$  there is an information plan that features learning  $P$  every period after  $\hat{T}$ . For  $H \in \mathcal{K}(\mathcal{F}(\Delta(C)))$  let  $P_H \in \arg \max_{P \in \mathcal{P}_{\mathfrak{M}}} V(H, P)$ . Consider now  $T^* > \hat{T}$ ; for  $\ell \diamond_{T^*} F_\infty$ , it must then be optimal to follow an information plan that features learning  $P_F$  in each period after  $T^*$ . If  $F_T G \succ_{T^*} F_\infty$ , 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  $G_\infty \succ_{T^*} G_T F$ . 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  $f_J(s) = c_s^+$  if  $s \in J$  and is  $c_s^-$  otherwise. Consider  $P$  and  $Q$  from part (b) of Definition 4.2. Without loss of generality, there exists  $\alpha \in (0, 1]$  such that  $V(F^P, P) = V(F_\alpha^Q, Q)$ , where  $F_\alpha^Q := \alpha F^Q + (1 - \alpha) f_\emptyset$ . 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  $\mathcal{P}_{\mathfrak{M}}$ , only an information plan that features  $P$  in every period after  $T^*$  is optimal for  $\ell \diamond_{T^*} F_\infty^P$ . Similarly, only an information plan that features  $Q$  in every period after  $T^*$  is optimal for  $\ell \diamond_{T^*} F_\infty^Q$ . Further, because  $Q \notin \Gamma(\tau(P, \theta, s))$  for any  $\theta \in \Theta$  and  $s \in S$ ,  $F_\infty^P \succ_{T^*} F_T^P F^Q$ . Analogously, because  $P \notin \Gamma(\tau(Q, \theta, s))$ ,  $(F^Q)_\infty \succ_{T^*} F_T^Q F^P$ . This establishes part (b) of Definition 4.1.

### References

- Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring". *Econometrica* 58 (5): 1041–1063. (Cit. on pp. 18, 20, 37).
- Bertsekas, Dimitri P., and Steven Shreve. 2000. *Stochastic Optimal Control: The Discrete-Time Case*. Belmont, MA: Athena Scientific. (Cit. on p. 53).
- Blackwell, David. 1953. "Equivalent Comparison of Experiments". *Annals of Mathematical Statistics* 24:265–272. (Cit. on p. 41).
- Blanchard, Olivier Jean, and Stanley Fischer. 1989. *Lectures on Macroeconomics*. Cambridge, Ma: MIT Press. (Cit. on p. 40).
- Brown, Meta, Christopher J. Flinn, and Andrew Schotter. 2011. "Real-Time Search in the Laboratory and the Market". *American Economic Review* 101 (2): 948–974. (Cit. on p. 24).
- Caplin, Andrew, and Mark Dean. 2014. *Revealed Preference, Rational Inattention, and Costly Information Acquisition*. Tech. rep. New York University. (Cit. on p. 26).

- De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and M. Kemal Ozbek. 2016. “Rationally Inattentive Preferences and Hidden Information Costs”. *Theoretical Economics*: forthcoming. (Cit. on pp. 2, 14, 25–27).
- Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 2001. “Representing Preferences with a Unique Subjective State Space”. *Econometrica* 69 (4): 891–934. (Cit. on p. 25).
- Dekel, Eddie, Barton L. Lipman, Aldo Rustichini, and Todd Sarver. 2007. “Representing Preferences with a Unique Subjective State Space: A Corrigendum”. *Econometrica* 75 (2): 591–600. (Cit. on p. 14).
- Dillenberger, David, R. Vijay Krishna, and Philipp Sadowski. 2016. *Supplement to ‘Subjective Dynamic Information Constraints’*. Tech. rep. Duke University. (Cit. on p. 6).
- Dillenberger, David, Juan S. Lleras, Philipp Sadowski, and Norio Takeoka. 2014. “A Theory of Subjective Learning”. *Journal of Economic Theory* 153:287–312. (Cit. on pp. 17, 25).
- Dubins, Lester E., and Leonard J. Savage. 1976. *Inequalities for Stochastic Processes: How to Gamble if You Must*. Mineola, New York: Dover Publications Inc. (Cit. on p. 40).
- Dudley, Richard Mansfield. 2002. *Real Analysis and Probability*. 2nd edition. Cambridge University Press. (Cit. on p. 29).
- Ellis, Andrew. 2016. *Foundations for Optimal Inattention*. Tech. rep. London School of Economics. (Cit. on pp. 2, 26).
- Ergin, Haluk, and Todd Sarver. 2010. “A Unique Costly Contemplation Representation”. *Econometrica* 78 (4): 1285–1339. (Cit. on pp. 14, 25–27).
- Gilboa, Itzhak, and David Schmeidler. 1989. “Maxmin Expected Utility with a Non-Unique Prior”. *Journal of Mathematical Economics* 18:141–153. (Cit. on pp. 15, 25).
- Gul, Faruk, and Wolfgang Pesendorfer. 2004. “Self-Control and the Theory of Consumption”. *Econometrica* 72 (1): 119–158. (Cit. on pp. 6, 19, 30).
- Herstein, Israel Nathan, and John Milnor. 1953. “An Axiomatic Approach to Measurable Utility”. *Econometrica* 21 (2): 291–297. (Cit. on p. 53).
- Higashi, Youichiro, Kazuya Hyogo, and Norio Takeoka. 2009. “Subjective random discounting and intertemporal choice”. *Journal of Economic Theory* 144 (3): 1015–1053. (Cit. on p. 8).
- Karatzas, Ioannis, and William Sudderth. 2010. “Two Characterizations of Optimality in Dynamic Programming”. *Applied Mathematics and Optimization* 61:421–434. (Cit. on p. 40).
- Koopmans, Tjalling C. 1964. “On the Flexibility of Future Preferences”. In *Human Judgments and Optimality*, ed. by M W Shelly and G L Bryan. John Wiley and Sons. (Cit. on p. 26).

- Kreps, David M. 1979. “A Representation Theorem for ‘Preference for Flexibility’”. *Econometrica* 47 (3): 565–577. (Cit. on pp. 14, 25).
- . 2012. *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton, NJ: Princeton University Press. (Cit. on p. 32).
- Kreps, David M, and Evan L Porteus. 1978. “Temporal Resolution of Uncertainty and Dynamic Choice Theory”. *Econometrica* 46 (1): 185–200. (Cit. on pp. 6, 13).
- Krishna, R. Vijay, and Philipp Sadowski. 2014. “Dynamic Preference for Flexibility”. *Econometrica* 82 (2): 655–703. (Cit. on pp. 6, 8, 10, 15, 16, 20, 21, 25, 28, 36, 42).
- Laffont, Jean-Jacques. 1989. *The Economics of Uncertainty and Information*. The MIT Press. (Cit. on pp. 11, 41).
- Ljungqvist, Lars, and Thomas J. Sargent. 2004. *Recursive Macroeconomic Theory*. 2nd edition. Cambridge, Ma: MIT Press. (Cit. on p. 23).
- Maćkowiak, Bartosz, and Mirko Wiederholdt. 2009. “Optimal Sticky Prices under Rational Inattention”. *American Economic Review* 99 (3): 769–803. (Cit. on p. 29).
- Mariotti, Thomas, Martin Meier, and Michele Piccione. 2005. “Hierarchies of Beliefs for Compact Possibility Models”. *Journal of Mathematical Economics* 41 (3): 303–324. (Cit. on p. 31).
- Markowitz, Harry M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York, NY: Wiley. (Cit. on p. 25).
- Massa, Massimo, and Andrei Simonov. 2006. “Hedging, Familiarity and Portfolio Choice”. *Review of Financial Studies* 19 (2): 633–685. (Cit. on p. 22).
- Matějka, Filip, and Alisdair McKay. 2014. “Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model”. *American Economic Review* forthcoming. (Cit. on p. 26).
- Myatt, David P., and Christopher Wallace. 2012. “Endogenous Information Acquisition in Coordination Games”. *Review of Economic Studies* 79:340–374. (Cit. on p. 29).
- Orkin, Michael. 1974. “On Stationary Policies—The General Case”. *The Annals of Statistics* 2 (1): 219–222. (Cit. on pp. 32, 53).
- Ozdenoren, Emre, Stephen W. Salant, and Dan Silverman. 2012. “Willpower and the Optimal Control of Visceral Urges”. *Journal of the European Economic Association* 10 (2): 342–368. (Cit. on p. 5).
- Persico, Nicola. 2000. “Information Acquisition in Auctions”. *Econometrica* 68 (1): 135–148. (Cit. on p. 2).
- Piermont, Evan, Norio Takeoka, and Roe Teper. 2015. *Learning the Krepsian State: Exploration Through Consumption*. Tech. rep. University of Pittsburgh. (Cit. on p. 25).

- Samuelson, William, and Richard Zeckhauser. 1988. “Status Quo Bias in Decision Making”. *Journal of Risk and Uncertainty* 1:7–59. (Cit. on p. 22).
- Sims, Christopher A. 2003. “Implications of Rational Inattention”. *Journal of Monetary Economics* 50 (3): 665–690. (Cit. on p. 2).
- . 1998. “Stickiness”. *Carnegie-Rochester Conference Series on Public Policy* 49:317–356. (Cit. on p. 2).
- Steiner, Jakub, Colin Stewart, and Filip Matějka. 2015. *Rational Inattention Dynamics: Inertia and Delay in Decision-Making*. Tech. rep. CERGE-EI. (Cit. on p. 28).
- Stigler, George. 1961. “The Economics of Information”. *Journal of Political Economy* 69 (3): 213–225. (Cit. on p. 2).
- Stokey, Nancy L., Robert E. Lucas Jr, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, Ma: Harvard University Press. (Cit. on p. 40).
- Woodford, Michael. 2012. *Inattentive Valuation and Reference-Dependent Choice*. Tech. rep. Columbia University. (Cit. on p. 26).