Subjective Information Choice Processes*

DAVID DILLENBERGER† R. VIJAY KRISHNA‡ PHILIPP SADOWSKI§

JANUARY 2017

Abstract

We propose 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 an Information Choice Process (ICP). An ICP specifies the information that can be acquired about the payoff-relevant state in the current period, and how this choice affects what can be learned in the future. The constraint imposed by the ICP is identified up to a recursive dynamic extension of Blackwell dominance. All the other parameters of the model are also uniquely identified. Additionally, we provide behavioral foundations, ie axioms, for the model.

KEY WORDS: Dynamic Preferences, Information Choice Process, Recursive Information Constraints, Recursive Blackwell Dominance, Rational Inattention, Subjective Markov Decision Process, Familiarity Bias

JEL Classification: D80, D81, D90

(*) We thank Attila Ambrus, Alexander Bloedel, Andrew Ellis, Itzhak Gilboa, Jay Lu, George Mailath, Robert Miller, 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>

1. Introduction

1.1. Motivation and Preview of Results

In a typical dynamic choice problem, a decision maker (henceforth DM) must choose an action that, contingent on the state of the world, determines a payoff for the current period as well as the collection of actions available in the next period. A standard example is a 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 market.

Faced with a dynamic choice problem, DM wants to acquire information about the state of the world, but often is constrained in the amount of information he 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 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. 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, unlike budget constraints, these constraints need not be linear and can accommodate many patterns, such as developing expertise in processing information. Indeed, our information constraints can encode arbitrary history dependence. Our framework unifies behavioral phenomena that arise in the presence of such constraints, regardless of their nature. In particular, 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.

To fix ideas, suppose DM has to manage his portfolio. In each period, there is a set of possible investments he can make, given the value of the portfolio. Depending on the state of the economy, each choice of investment results in an instantaneous payoff (eg, a dividend) and a continuation value of the portfolio for the next period. Further suppose that in order to improve his portfolio choice, DM must also take an unobservable action to acquire information about the true state. The choice of information may affect the feasible information choices in the next period. For instance, it may be that DM is subject to fatigue, and so can acquire information only if he did not do so in the last period. Alternatively, he may gain expertise, so that acquiring a particular piece of information in one period makes it easier to acquire that same information in subsequent periods. These information constraints become increasingly complicated as the length of DM's planning horizon grows. The difficulty for the analyst is that while portfolio choice is observable, DM's information choice, and it's impact on the feasibility of subsequent information choices, is not. Identifying these constraints is important, for example, for policy decisions; Mullainathan and Shafir (2013) suggest that poorer people make suboptimal investment choices, because they face too many demands on their time or cognitive resources to fully inform themselves. In order to ameliorate the effects of information constraints, a policy maker must first understand them. It is natural to ask, Can (unobservable) information constraints be identified by looking at DM's preferences over portfolio problems? And if so, can useful inference be made from a small number of initial portfolio choices, without the need to observe investment decisions over time?

Our main result is Theorem 1, which shows that the class of dynamic choice problems we consider is sufficiently rich to infer the *entire* set of parameters governing DM's preferences over those problems from observed behavior; the parameters being (i) state dependent utilities, (ii) (time varying) beliefs about the state, (iii) the discount factor, and (iv) the information constraint up to a recursive extension of informativeness (Blackwell dominance).[1]

Formally, DM faces an entirely subjective control problem, which specifies how future constraints depend on current and past choices of information that are *unobservable* (to the analyst), and which we refer to as an *Information Choice Process* (ICP). The payoff-relevant state $s \in S$ changes over time, and we focus on partitional learning in every period. The ICP is parametrized by a control state θ, a function $\Gamma(\theta)$ that determines the set of feasible partitions of S, and an operator τ that governs the transition of θ in response to the choice of partition and the realization of $s \in S$. Examples of such ICPs are in Section 1.2.

The domain of choice consists of sets (or *menus*) of actions, where each action (an *act* on S) results in a state-contingent lottery that yields current consumption and a new menu of acts for the next period. Our model suggests the following timing of events and decisions, as illustrated in Figure 1. DM enters a period facing a menu x, while being equipped with a prior belief π_s over S and an (information) control state θ. 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 to obtain $\pi_s(\cdot \mid I)$ and then 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 current consumption c and continuation menu y for the next period. At the same time, a new control state $\theta' = \tau(\theta, P, s')$ and a new belief $\pi_{s'}$ are determined for next period. (The collection of measures $(\pi_s)_{s \in S}$ implicitly defines the transition operator for a Markov process on S.)

(1) Recent work 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.

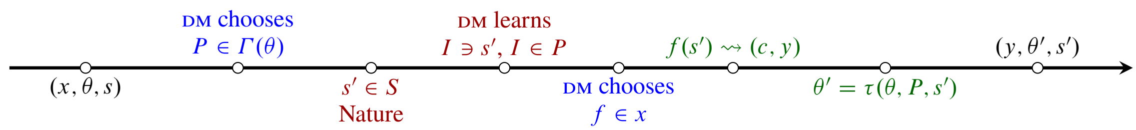

Figure 1: *Timeline*

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

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

Theorem 1 establishes that $(u_s, \pi_s)_{s \in S}$ and δ are essentially uniquely identified, and that the remaining preference parameter, the ICP, is identified up to the addition or deletion of information choices that are dominated in terms of a recursive extension of Blackwell's order comparing information structures. Identifying a subjectively controlled process (the ICP) from behavior is our main conceptual contribution. The unobservability of both the control state θ and the information choice makes DM's control problem a subjective Markov Decision Process with partially unobservable states, actions, and transitions. To the best of our knowledge, ours is the first result on identifying such a decision process.

The proof of Theorem 1 relies on a notion of duality — which we term strong alignment — between the space of ICPs and our domain of observable dynamic choice problems. In Section 3.3 we discuss some general lessons to be learned from our results for identifying parameters in other models with unobservable choice processes. In particular, in Section 6 we consider a related model that features history-dependent information costs instead of a time-varying constraint, and discuss how, with slight modifications, our approach can be used to identify all the parameters of such a model.

In Section 4 we discuss an axiomatic foundation for our model, with the formal statement of the axioms being deferred to Appendix C. We first establish the extent to which aspects of standard properties, such as Independence and Temporal Separability — that are central to virtually all existing axiomatic models of dynamic choice — hold even when behavior depends on contemporaneous unobservable information choice. The unobservability of the control state θ implies that observed behavior is not stationary, even though the value function is recursive. Our most important axiomatic innovation is then to generate a recursive axiomatic structure: We assume that preferences over dynamic choice problems are *self-generating*, that is, they have the same properties as each of the preferences over continuation problems that together generate them (and thus embody the dynamic programming principle).

The paper is organized as follows. The next subsection presents examples of ICPs and some of the behavioral patterns they can generate. In Section 2 we introduce the analytical framework, state our utility representation, and describe our notion of comparative informativeness for ICPs. Section 3 establishes our identification result, provides a sketch of the proof, and discusses some general insights from our identification strategy that extend to other settings with unobservable actions. Section 4 discusses behavioral foundations, namely the axioms, and provides a representation theorem, Theorem 3. Section 5 surveys related decision-theoretic literature, while other related literature is discussed in the relevant sections. Section 6 extends our analysis to a model with direct information costs. Section 7 offers some concluding remarks. Proofs can be found in the Appendix; additional technical details are in the Supplementary Appendix.[2]

1.2. Examples of ICPs and Patterns of Behavior

ICPs can accommodate *any* dependence of the information constraint on the history of information choices and state realizations. We now give a few simple and illustrative examples. Our first example draws analogy between optimal choice of information and a standard consumption-investment problem.

Example 1.1. In 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 β. 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).[3] Formally, with 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. An ICP of this sort is parametrized by the 4-tuple (K_0, κ, c, β) where K_0 is the initial stock of attention. The case with $\beta = 0$ corresponds to a typical per period constraint in the literature on rational inattention, as for example in Maćkowiak and Wiederholdt (2009).

The next two examples can be used to capture *fatigue* in learning, where paying attention in the current period diminishes the ability to pay further attention.

Example 1.2. 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,

(2) The Supplementary Appendix, Dillenberger, Krishna, and Sadowski (2017), is available online at <http://people.duke.edu/~ps78/documents/ICP-supp.pdf>.

(3) 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.

he can only learn the trivial partition of S) in the current one.[4] In this case, we may set $\Theta = \{0, 1\}$, $\theta_0 \in \Theta$,

$$\Gamma(\theta) := \begin{cases} \{S\} & \text{if } \theta = 0 \\ \mathscr{P} & \text{if } \theta = 1 \end{cases} \quad \text{and} \quad \tau\big(P, \theta, s\big) := \begin{cases} 0 & \text{if } \theta = 1,\, P \neq \{S\} \\ 1 & \text{if } P = \{S\} \end{cases}$$

where $\mathscr{P}$ is the collection of all partitions of S.

In Section 7.3 we briefly comment on the coordination problem that may arise when firms that are constrained as in Example 1.2 interact strategically.

Example 1.3 (Resource exhaustion)**.** DM is endowed with an initial attention stock K_0, which he draws down every time he chooses to learn. The amount of attention stock drawn is equal to the cost of choosing a partition. Formally, this corresponds to the ICP in Example 1.1 with parameters $(K_0, \kappa = 0, c, \beta = 1)$.

Conceptually, this type of ICP 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. Consider 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, it is easy to show that our model as specified in Example 1.3 will generate a reservation wage that decreases over time, because the expected value of continuing the search decreases as the information constraint tightens over time, due to search-fatigue.

The evolution of the constraint may also depend on the evolution of the state, as in the next example.

Example 1.4 (State dependence)**.** The feasible set of partitions at any period solely depends on the realization of the state in the previous period. Thus, in this example, $\Theta = S$ and $\tau(s, P, s') = s'$ for all $s \in S$ and $P \in \Gamma(s)$.

Our last example captures the notion of *expertise* in learning, that is, a complementarity between information acquired in different periods.

Example 1.5 (Example 1.1 continued)**.** The cost of learning a partition depends on past choices. In particular, if partition Q was chosen yesterday, then the cost of learning P today is $c(P \mid Q) = (1 - b)H_\mu(P) + bH_\mu(P \mid Q)$, where, given a probability μ over S, $H_\mu(P)$

(4) This example suggests that periods in which individuals pay careful attention are usually followed by periods in which they should rest. In addition to the cognitive interpretation, 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.

is the entropy of P and $H_\mu(P \mid Q)$ is the relative entropy of P with respect to Q. Note that $H_\mu(P \mid P) = 0$ and hence $c(P, P) = (1-b)H_\mu(P)$. That is, while learning P initially costs $H_\mu(P)$, learning P again in the subsequent period costs only a fraction $(1-b)$ thereof. The parameter b measures the degree to which DM can gain expertise.

To see how expertise can help explain a preference pattern that is difficult to reconcile in the absence of dynamic information constraints, suppose DM has become familiar with a certain set of alternatives 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* from the same set of alternatives, as he may find it too attention-intensive to discern the value of 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.[5] [6] The benefit of avoiding such a switch is amplified by, for example, a greater value of the parameter b of Example 1.5.

Another intuitive ICP that can capture expertise is as follows. From a set of possible experiments, each of which corresponds to a different partition of S, DM might be able to 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. This might correspond to an individual building human capital through a sequence of courses, each of which imparts him with expertise in possibly different areas. Unlike Example 1.5, this constraint does not rely on the specification of an attention cost function, and captures permanent expertise.

We note that *every* ICP has a finite horizon truncation (see Definition 2.3), that mimics the original ICP for t periods, after which it admits only the coarsest partition. One of our technical contributions lies in metrizing the space of ICPs and showing that every ICP can be approximated by its finite horizon truncations. (See Section 3.2 and especially the discussion following Proposition 3.3.)

(5) If we denote by FtG the menu that allows DM to choose a consumption act (without any continuation component) from menu F for the first t periods and from G thereafter — where H_∞ denotes choosing from menu H forever — then we may say that DM is familiarity biased between F and G after t periods whenever he strictly prefers both F_∞ to FtG and G_∞ to GtF. Intuitively, DM is likely to display such preferences if the optimal choices of acts from F and G require a different type of learning, for example, if they involve bets on different types of events.

(6) Among investors who manage their own portfolio (rather than use index funds), evidence that the home bias in portfolio choice persists in favor of the old home even after a move to a new location — see Massa and Simonov (2006) — nicely illustrates this pattern.

2. Representation with Information Choice Processes

2.1. Domain

Let S be a finite set of *objective* or *observable* states. 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.

Let C be a compact metric space representing consumption. A one-period consumption problem is $x_1 \in X_1 := \mathcal{K}\big(\mathcal{F}(\Delta(C))\big)$. It consists of a menu of acts, each of which results in a state-dependent lottery over instantaneous consumption prizes. Then, the space of two-period consumption problems is $X_2 := \mathcal{K}\big(\mathcal{F}(\Delta(C \times X_1))\big)$, so that each two-period problem consists of a menu of acts, each of which results in a lottery over consumption and a one-period problem for the next period. Proceeding inductively, we may similarly define t-period problems as $X_t := \mathcal{K}\big(\mathcal{F}(\Delta(C \times X_{t-1}))\big)$.

Our domain consists of *dynamic choice problems* (henceforth, choice problems) and is denoted by X which is itself homeomorphic to $\mathcal{K}\big(\mathcal{F}(\Delta(C \times X))\big)$.[7] Note that *both* the current and the continuation choice problems are now in X. For any $x, y \in X$ and $t \in [0,1]$, we let $tx + (1-t)y := \{tf + (1-t)g : f \in x,\ g \in y\} \in X$.

A consumption stream is a degenerate choice problem that does not involve choice at any point in time. The space L of all consumption streams can be written recursively as $L \simeq \mathcal{F}(\Delta(C \times L))$. Thus, each $\ell \in L$ is an act that yields a state-dependent lottery over instantaneous consumption and continuation consumption streams (an $\ell' \in L$). There is a natural embedding of L in X. We analyze DM's preferences $\succsim$ over choice problems, which is a binary relation on X, and denote its restriction to L by $\succsim|_L$.

The space X of choice problems, 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).

2.2. ICP-Representation

Given a choice problem, 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 Information Choice Process (ICP). Formally, an ICP is a tuple $\mathcal{M} = (\Theta, \theta_0, \Gamma, \tau)$, where Θ is a set of ICP (or *control*)

(7) Formally, the homeomorphism is written as $X \simeq \mathcal{K}\big(\mathcal{F}(\Delta(C \times X))\big)$. See Appendix A.2 for details. We also refer to X as the space of Recursive Anscombe-Aumann Choice Problems.

states; θ_0 is the *initial* state; $\Gamma : \Theta \to 2^{\mathcal{P}} \setminus \varnothing$ is a set of *feasible partitions* in a given ICP or control state θ; and $\tau : \mathcal{P} \times \Theta \times S \to \Theta$ is a *transition operator* that determines the transition of the ICP state θ, given a particular choice of partition and the realization of an objective state. Let **M** be the space of ICPs.

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 Π be a fully connected transition operator[8] for a Markov process on 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 π_{s_0} the unique invariant measure of Π.

We consider the following utility representation of $\succsim$ on the space X of choice problems.

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

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

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 σ-algebra), so that $\mathbf{E}^{f(s')}$ is the expectation with respect to this probability measure.[9]

A *dynamic information plan* prescribes a choice of $P \in \Gamma(\theta)$ for each tuple (x,θ,s). Thus, an ICP 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 ICP-representation $\big((u_s)_{s\in S}, \delta, \Pi, \mathcal{M}\big)$ induces a unique function $V : X \times \Theta \times S \cup \{s_0\} \to \mathbb{R}$ that is continuous on X and satisfies [**Val**]. Moreover, an optimal dynamic information plan exists.

A proof is in Appendix A.4.

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

(9) 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 Recursive Anscombe-Aumann Choice Problems, 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,θ,s), so that *implied* behavior is dynamically consistent contingent on those state variables. The novel aspect is that the MDP state θ is controlled by DM and is not observed by the analyst.

2.3. Comparative Informativeness of ICPs

As noted above, an ICP can be viewed as circumscribing the set of available dynamic information plans. We now show that the space of ICPs has a natural order.

We call any ICP that affords only the coarsest partition after t periods a *t-period* ICP. We first formalize the notion of a finite horizon truncation of an arbitrary ICP.

Definition 2.3. Let $\mathscr{M} = (\Theta, \theta_0, \Gamma, \tau)$ be an ICP. For each $t \geq 1$, its finite horizon truncation is denoted by $\mathscr{M}_{(t)} := \big(\Theta_{(t)}, (\theta_0, t), \Gamma_{(t)}, \tau_{(t)}\big)$ defined as follows: $\Theta_{(t)} := \Theta \times \{0, \dots, t\}$,

$$\Gamma_{(t)}(\theta, j) := \begin{cases} \Gamma(\theta) & \text{if } j \geq 1 \\ \{S\} & \text{if } j = 0 \end{cases} \quad \text{and} \quad \tau_{(t)}\big(P, (\theta, j), s\big) := \begin{cases} \big(\tau(P, \theta, s), j-1\big) & \text{if } j > 1 \\ (\theta_0, 0) & \text{if } j = 0, 1 \end{cases}$$

Thus, the t-period ICP $\mathscr{M}_{(t)}$ mimics $\mathscr{M}$ for t periods.

A natural way to compare partitions is in terms of fineness, which coincides with Blackwell's comparison of informativeness. To extend this idea to ICPs, first consider two one-period ICPs $\mathscr{M}$ and $\mathscr{M}'$. Notice that as far as dynamic information plans are concerned, all that matters are the partitions each ICP renders feasible. This suggests the following order on one-period ICPs: *$\mathscr{M}$ (one-period) Blackwell dominates $\mathscr{M}'$* if for every $P' \in \Gamma'(\theta_0')$, there exists $P \in \Gamma(\theta_0)$ such that P is finer than P'.

In turn, this suggests a natural extension to two-period ICPs. *$\mathscr{M}$ (two-period) Blackwell dominates $\mathscr{M}'$* if for every $P' \in \Gamma'(\theta')$, there exists $P \in \Gamma(\theta_0)$ such that (i) P is finer than P', and (ii) for all $s \in S$ and for every $Q' \in \Gamma'(\tau'(P, \theta_0', s))$, there exists $Q \in \Gamma(\tau(P, \theta_0, s))$ such that Q is finer than Q'. Thus, for any information plan in $\mathscr{M}'$, there is another plan in $\mathscr{M}$ that is more informative in every period and state.

To extend our construction to more than two periods, we note that requirement (ii) amounts to the one-period continuation ICP $(\Theta, \tau(P, \theta_0, s), \Gamma, \tau)$ being more informative than $(\Theta', \tau'(P, \theta_0', s), \Gamma', \tau')$. In a similar fashion, we then inductively define an order extending Blackwell dominance to all t-period ICPs, whereby one t-period ICP *(t-period) Blackwell dominates* another if for each information plan from the second, there is another plan from the first that is more informative in the first period and, for all $s \in S$, leads to a more informative $(t-1)$-period plan starting in the second period.

As an example, consider the ICPs introduced in Example 1.1 where attention stock is drawn down, decays, and is renewed with attention income. Such an ICP is parametrized by the 4-tuple (K_0, κ, c, β). Consider now two ICPs $\mathscr{M}^i$, for $i = a, b$, parametrized by $(K_0, \kappa, c^i, \beta)$ that only differ in their costs of acquiring information. It is easy to see that $\mathscr{M}^a_{(1)}$ (the one-period truncation of $\mathscr{M}^a$) Blackwell dominates $\mathscr{M}^b_{(1)}$ if, and only if, $c^a \leq c^b$ (ie, $c^a(P) \leq c^b(P)$ for all $P \in \mathscr{P}$). Similarly, $\mathscr{M}^a_{(t)}$ (t-period) Blackwell dominates $\mathscr{M}^b_{(t)}$ if, and only if, $c^a \leq c^b$.

Ordering arbitrary ICPs is more delicate, because unlike finite horizon ICPs, arbitrary ICPs may not have a final period of non-trivial information choice, and hence may not permit backwards induction. Instead, we exploit the recursive structure of ICPs, so that our ordering of informativeness for arbitrary ICPs is also recursive.

Proposition 2.4. The *recursive Blackwell order* is non-trivial and is the *largest* order that satisfies the following: For all $\mathscr{M}, \mathscr{M}^\dagger \in \mathbf{M}$, $\mathscr{M}$ *dominates* $\mathscr{M}^\dagger$ 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 $s \in S$.

Proposition 2.4 follows from Propositions A.3 and A.10 in Appendices A.5 and A.6. In particular, it relies on a metrization of $\mathbf{M}$ which implies that (i) $\mathscr{M}$ recursively Blackwell dominates $\mathscr{M}'$ if, and only if, $\mathscr{M}_{(t)}$ (t-period) Blackwell dominates $\mathscr{M}'_{(t)}$ for all $t \geq 1$, and (ii) $\mathscr{M}_{(t)}$ converges to $\mathscr{M}$ as $t \to \infty$.

We note that there are other ways to define dynamic versions of the static Blackwell order; see, for instance, De Oliveira (2016) and the references therein. Our approach differs from these in that instead of comparing signal processes, we compare *controlled* signal processes that allow the decision maker to choose his signal (in our case, partition) as a function of the past. That our approach is particularly well suited to our problem is demonstrated by our main identification result, Theorem 1, in the next section.

3. Unique Identification

Theorem 1. *Let $((u_s), \delta, \Pi, \mathscr{M})$ be an ICP-representation of $\succsim$. Then, the functions $(u_s)_{s \in S}$ are unique up to the addition of constants and a common scaling, δ and Π are unique, and $\mathscr{M}$ is unique up to recursive Blackwell equivalence.*[10]

The formal proof is in Appendix B. On the subdomain L, V is independent of $\mathscr{M}$ and satisfies Independence. Indeed, V is then completely characterized by the parameters $((u_s), \delta, \Pi)$. Krishna and Sadowski (2014) show that such a representation on L is unique up to the addition of constants and a common scaling of (u_s). Our challenge then is to identify the ICP $\mathscr{M}$. In Section 3.1 we discuss the main ideas behind the identification strategy for finitely many periods, in Section 3.2 we construct the space of canonical infinite horizon constraints that allows us to extend our analysis to the infinite horizon, and in Section 3.3 we emphasize aspects of our identification strategy that generalize beyond the specifics of our model.

Notice that our model is a Markov decision process with state (x, θ, s), where x and s are observable by the analyst, while θ is not. Actions consist of choice of partition and an act that depends on the information received. While the choice of act is observed, the

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

choice of partition and subsequent information received is not. Thus, the Markov decision process is subjective with partially unobservable actions, states and transitions. Theorem 1 achieves full identification of this subjective Markov decision process. To the best of our knowledge, this is the first result of this sort in the literature.[11]

An immediate benefit of identifying all the parameters is that it allows a meaningful comparison of decision makers. The next result demonstrates that recursive Blackwell dominance plays the same role in our dynamic environment as does standard Blackwell dominance in a static setting.

Consider 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 $x \succsim \ell$.[12] The comparison in the definition implies that $\succsim$ and $\succsim^\dagger$ have the same ranking over consumption streams in L.[13] While any consumption stream requires no choice of information, a typical choice problem 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 ICP renders feasible. The uniqueness established in Theorem 1 allows us to formalize this intuition.

Theorem 2. *Let $((u_s), \delta, \Pi, \mathcal{M})$ and $\big((u_s^\dagger), \delta^\dagger, \Pi^\dagger, \mathcal{M}^\dagger\big)$ be* ICP*-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 Theorem connects a purely behavioral comparison of preferences to recursive Blackwell dominance of ICPs, 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 ICPs, a theme we will return to when we sketch the proof of Theorem 1 in Section 3.1. A useful corollary of Theorem 2 is the following characterization of the recursive Blackwell order: $\mathcal{M}$ recursively Blackwell dominates $\mathcal{M}^\dagger$ if, and only if, every discounted, expected utility maximizer prefers to have the ICP $\mathcal{M}$ instead of $\mathcal{M}^\dagger$.

The remainder of this section provides intuition for the proof of Theorem 1.

3.1. Strong Alignment between Choice Problems and ICPs

The choice problem x *separates* the ICPs $\mathcal{M}$ and $\mathcal{M}'$ if DM derives different values from x when constrained by (the information plans in) $\mathcal{M}$, than by $\mathcal{M}'$. Intuitively, if the value is

(11) For an econometrician's (as opposed to decision theorist's) perspective on identification, see Rust (1994).

(12) 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).

(13) 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.

higher under $\mathscr{M}$, which we write as $V(x;\mathscr{M}) > V(x;\mathscr{M}')$, then there is a dynamic information plan that is feasible under $\mathscr{M}$, which does strictly better for the choice problem x than *every* dynamic information plan feasible under $\mathscr{M}'$.

To identify $\mathscr{M}$, we need to show that for any ICP $\mathscr{M}'$ that does not recursively Blackwell dominate $\mathscr{M}$, there is a choice problem that separates $\mathscr{M}$ and $\mathscr{M}'$.[14] Our identification strategy for $\mathscr{M}$ is to find a finite collection of choice problems, such that for any such ICP $\mathscr{M}'$, each problem in the collection generates at least as much value under $\mathscr{M}$ as under $\mathscr{M}'$,[15] and at least one problem in the collection separates $\mathscr{M}$ and $\mathscr{M}'$. We call such a collection of choice problems *uniformly strongly aligned* with $\mathscr{M}$.

We first discuss how to construct collections of choice problems that are uniformly strongly aligned with one- and two-period ICPs. We then extend these ideas to arbitrary ICPs. Throughout, we emphasize the duality between the space of ICPs, **M**, and the space of collections of choice problems in X. In particular, to separate two t-period ICPs, it is necessary and sufficient to consider t-period choice problems, and these choice problems must resolve temporal uncertainty gradually, as we describe below.

3.1.1. *One-period* ICPs

Consider the one-period ICP $\mathscr{M}_{(1)}$. Notice that behaviorally all that matters is the set of partitions available for choice in the first period, namely the set $\Gamma(\theta_0)$. For each $P \in \Gamma(\theta_0)$, consider the one-period choice problem $x_1(P,\mathscr{M}_{(1)})$ defined as follows:

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

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

Notice that given choice problem $x_1(P,\mathscr{M}_{(1)})$ and the ICP $\mathscr{M}_{(1)}$, an optimal choice of partition in the first period is P, ie, $P \in \arg\max_{Q\in\Gamma(\theta_0)} V(x_1(P,\mathscr{M}_{(1)}), Q;\mathscr{M}_{(1)})$ (where $V(x,Q;\mathscr{M})$ is the utility from choosing Q in the first period given the ICP $\mathscr{M}$ and the choice problem x.) Notice also that, for any ICP $\mathscr{M}'$,

$$V(x_1(P,\mathscr{M}_{(1)}), P;\mathscr{M}_{(1)}) = V(\ell^*;\mathscr{M}_{(1)}) \geq V(x_1(P,\mathscr{M}_{(1)});\mathscr{M}')$$

(14) Of course, if $\mathscr{M}$ and $\mathscr{M}'$ recursively Blackwell dominate each other, then no choice problem can separate them.

(15) Intuitively, given any choice problem x in the collection, there is an information plan in $\mathscr{M}$ such that no other information plan (in any $\mathscr{M}'$) can give a strictly higher utility.

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

Moreover, if $\mathscr{M}'$ does not recursively Blackwell dominate $\mathscr{M}_{(1)}$, then there exists $P \in \Gamma(\theta_0)$ such that there is no $P' \in \Gamma'(\theta_0')$ that is finer than P. In this case, we must necessarily have

$$V(x_1(P, \mathscr{M}_{(1)}); \mathscr{M}_{(1)}) = V(x_1(P, \mathscr{M}_{(1)}), P; \mathscr{M}_{(1)}) > V(x_1(P, \mathscr{M}_{(1)}); \mathscr{M}')$$

In other words, with $x_1(P, \mathscr{M}_{(1)})$ as defined in [♠], the collection of choice problems $\{x_1(P, \mathscr{M}_{(1)}); P \in \Gamma(\theta_0)\}$ is uniformly strongly aligned with $\mathscr{M}_{(1)}$, that is, $\mathscr{M}_{(1)}$ can be separated from any ICP that does not recursively Blackwell dominate it by at least one choice problem in that collection.[17] We now extend this construction to two periods.

3.1.2. *Two-period* ICPs

Now consider a two-period ICP $\mathscr{M}_{(2)}$. As with one-period ICPs, all that is behaviorally relevant for the description of $\mathscr{M}_{(2)}$ is the set $\Gamma(\theta_0)$ of partitions available for choice in the first period, and the partitions available in the next period, given by $\Gamma(\tau(P, \theta_0, s))$, for each first-period choice of partition P and realized state s. As with one-period ICPs, we can establish the following.

Lemma 3.1. For each two-period ICP $\mathscr{M}_{(2)}$, there is a collection of two-period choice problems that are strongly aligned with it.

The Lemma is a special case of Lemma B.1 in the Appendix. We illustrate the main ideas behind the construction via the simple example displayed in Figure 2.

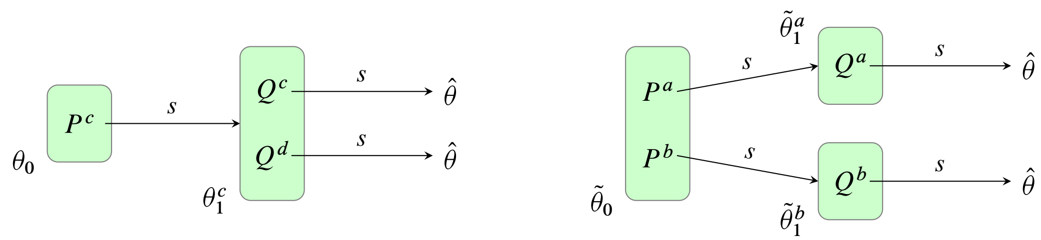

Figure 2: ICPs *$\mathscr{M}_{(2)}$ and $\tilde{\mathscr{M}}_{(2)}$*

Consider $\mathscr{M}_{(2)} := \{\Theta, \theta_0, \tau, \Gamma\}$, where $\Theta = \{\theta_0, \theta_1^c, \hat\theta\}$, and the mappings τ and Γ are summarized on the left in Figure 2: $\Gamma(\theta_0) = P^c$, $\Gamma(\theta_1^c) = \{Q^c, Q^d\}$, and $\tau(\theta_0, P^c, \cdot) = \theta_1^c$. Note also that $\Gamma(\hat\theta) = \{S\}$ and $\tau(\hat\theta, \{S\}, \cdot) = \hat\theta$, because $\mathscr{M}_{(2)}$ is a 2-period ICP. The ICP $\tilde{\mathscr{M}}_{(2)} := \{\tilde\Theta, \tilde\theta_0, \tilde\tau, \tilde\Gamma\}$ is defined similarly on the right in the figure. It is easy to see that $\mathscr{M}_{(2)}$ and $\tilde{\mathscr{M}}_{(2)}$ do not recursively Blackwell dominate each other.

Suppose now that the partitions are strictly ordered in terms of fineness as in Figure 3 where $P^b \longrightarrow P^c$, for instance, indicates that P^b is strictly finer than P^c.

---

(17) The construction of the act $f_{1,J}$ is standard. A textbook rendition is Theorem 1 of Laffont (1989, p59), which also concerns a static setting. The size of the set $\{x_1(P, \mathscr{M}_{(1)}); P \in \Gamma(\theta_0)\}$ is bounded above by the number of possible partitions of the finite state space S.

Figure 3: *Partitions ordered by fineness*

One might think that DM would always be better off with $\tilde{\mathcal{M}}_{(2)}$ as an ICP instead of $\mathcal{M}_{(2)}$. After all, given $\mathcal{M}_{(2)}$, the plan σ_1 that says 'Pick P^c in the first period and Q^c in the second' is strictly less informative than the plan $\tilde{\sigma}_1 =$ 'Pick P^a in the first period and Q^a in the second', and $\tilde{\sigma}_1$ is feasible under $\tilde{\mathcal{M}}_{(2)}$. Similarly, the plan $\sigma_2 =$ 'Pick P^c in the first period and Q^d in the second' is (strictly) less informative than the plan $\tilde{\sigma}_2 =$ 'Pick P^b in the first period and Q^b in the second'.

Consider, however, the information plan

$\sigma_3 =$ 'Pick P^c in the first period, wait for uncertainty to resolve, and then pick either Q^c or Q^d in the second period'

which is feasible under $\mathcal{M}_{(2)}$.

Claim 3.2. There exists a choice problem for which σ_3 is the optimal information plan, and strictly dominates every plan in $\tilde{\mathcal{M}}_{(2)}$.

Proof of Claim. Let $\mathcal{M}^c_{(1)} := \{\Theta, \theta^c_1, \tau, \Gamma\}$ denote a one-period ICP as depicted on the left panel of Figure 2. Let $x_1(Q^c, \mathcal{M}^c_{(1)})$ and $x_1(Q^d, \mathcal{M}^c_{(1)})$ be one-period choice problems as defined in [♠].

Define the two-period choice problem $x_2(P^c, \mathcal{M}_{(2)})$ as

$$x_2(P^c, \mathcal{M}_{(2)}) := \{f_{2,J} : J \in P^c\} \quad \text{where}$$
$$f_{2,J}(s) := \begin{cases} \left(c^+_s, \text{Unif}\left\{(x_1(Q^c, \mathcal{M}^c_{(1)}), (x_1(Q^d, \mathcal{M}^c_{(1)}))\right\}\right) & s \in J \\ \ell_*(s) & s \notin J \end{cases}$$

where Unif $\left\{(x_1(Q^c, \mathcal{M}^c_{(1)}), (x_1(Q^d, \mathcal{M}^c_{(1)}))\right\}$ is the equiprobable lottery over the one period choice problems $x_1(Q^c, \mathcal{M}^c_{(1)})$ and $x_1(Q^d, \mathcal{M}^c_{(1)})$.

We now argue that $x_2(P^c, \mathcal{M}_{(2)})$ separates $\mathcal{M}_{(2)}$ and $\tilde{\mathcal{M}}_{(2)}$ and that $x_2(P^c, \mathcal{M}_{(2)})$ is uniformly strongly aligned with $\mathcal{M}_{(2)}$. To see this, notice first that the value of $x_2(P^c, \mathcal{M}_{(2)})$ under the plan σ_3, where the choice depends on the outcome of the lottery over $x_1(Q^c, \mathcal{M}^c_{(1)})$ and $x_1(Q^d, \mathcal{M}^c_{(1)})$, gives maximal utility, ie, the same utility as ℓ^*. Thus, we have $V(\ell^*; \cdot) = V\left(x_2(P^c, \mathcal{M}_{(2)}); \mathcal{M}_{(2)}\right) \geq V\left(x_2(P^c, \mathcal{M}_{(2)}); \mathcal{M}'\right)$ for all $\mathcal{M}'$.

Moreover, the utility from following either plan $\tilde{\sigma}_1$ or $\tilde{\sigma}_2$, the only feasible plans under $\tilde{\mathcal{M}}_{(2)}$ (depicted on the right in Figure 2), is strictly less than ℓ^*. Thus, $x_2(P^c, \mathcal{M}_{(2)})$ separates $\mathcal{M}_{(2)}$ and $\tilde{\mathcal{M}}_{(2)}$ as claimed, and in fact the singleton $\left\{x_2(P^c, \mathcal{M}_{(2)})\right\}$ is uniformly strongly aligned with $\mathcal{M}_{(2)}$. □

In a similar fashion, we can construct two-period choice problems, $x_2(P^a, \tilde{\mathscr{M}}_{(2)})$ and $x_2(P^b, \tilde{\mathscr{M}}_{(2)})$, such that $\left\{x_2(P^a, \tilde{\mathscr{M}}_{(2)}), x_2(P^b, \tilde{\mathscr{M}}_{(2)})\right\}$ is uniformly strongly aligned with $\tilde{\mathscr{M}}_{(2)}$.[18]

There is an obvious sense in which $\mathbf{M}$ and X are dual to each other: Given $\big((u_s), \delta, \Pi\big)$, each $\mathscr{M}$ corresponds to a functional on X, and each x to a functional on $\mathbf{M}$. Moreover, $\mathbf{M}$ separates points in X (which is immediate from the representation), while X separates points in $\mathbf{M}$, which is the content of Theorem 1. Uniform strong alignment then amounts to a notion of direction in these spaces.[19] The example illustrates two central features of this duality. First, because ICPs may require DM to trade off coarser partitions at one date with finer partitions at another date, it is essential that choice problems allow for sufficient variation in acts at different dates. Second, because ICPs can accommodate information plans that allow DM to delay the choice of information (partition) until a later date, choice problems must feature *temporal resolution* of uncertainty over choice problems (a notion first introduced by Kreps and Porteus (1978)) which renders the option to delay information choice valuable.

The arguments above readily extend to t-period ICPs. That is, given any t-period ICP $\mathscr{M}_{(t)}$, there is a finite collection of t-period choice problems that is uniformly strongly aligned with it.

3.2. Infinite Horizon ICPs and Recursive Information Constraints

Extending the arguments in Section 3.1.2 to arbitrary (infinite) ICPs is more difficult because these arguments rely crucially on backward induction. The space $\mathbf{M}$ of ICPs, however, does not have a natural topology which would allow us to approximate arbitrary ICPs by their finite horizon truncations.

Nevertheless, as in the case of one and two periods, all that is relevant for the description of $\mathscr{M}$ is the set of partitions that are available at each instance. If we are willing to identify ICPs that permit the same choice of partition after every history, then we can metrize $\mathbf{M}$ and prove that finite horizon ICPs approximate arbitrary ICPs. Because all payoffs are bounded, the construction described above then establishes Theorem 1 via a simple continuity argument. We now describe the (pseudo-) metrization of $\mathbf{M}$.

Two ICPs $\mathscr{M}$ and $\mathscr{M}'$ are *indistinguishable* if they afford the same choices of partition in the first period and, for any choice in the first period, the same state-contingent choices in the second period, and so on. Intuitively, indistinguishable ICPs differ only up to a relabeling

(18) As in the case of one period, the number of choice problems needed in a uniformly strongly aligned set is bounded above by the number of possible partitions of S.

(19) Alignment is naturally seen as a property of Hilbert spaces, where direction is an obvious concept, and the dual and primal spaces are identical. A natural extension to normed spaces can be found on page 116 of Luenberger (1969). Our definition can be seen as an extension of this idea to arbitrary spaces.

of the control states, and up to the addition of control 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.7.

It is convenient to consider canonical ICPs that are defined on a control state space Ω which is compact and metrizable. We now describe Ω which, like the space X, is constructed inductively. Up to indistinguishability, the space of all one-period ICPs (see Definition 2.3) is $\Omega_1 := \mathcal{K}(\mathcal{P})$, ie, the space of all non-empty subsets of partitions of S. Inductively, we define the space of t-period ICPs (up to indistinguishability) $\Omega_t := \mathcal{K}_\flat\big(\mathcal{P} \times \Omega_{t-1}^S\big)$,[20] whereby each t-period ICP yields a state-contingent $(t-1)$-period ICP in the subsequent period. Because Ω_1 is metrizable, it follows that each Ω_t is also metrizable.

In the (projective) limit, we obtain the *canonical* space of all ICPs $\Omega \simeq \mathcal{K}_\flat\big(\mathcal{P} \times \Omega^S\big)$. This formalization suggests a recursive way to think of Ω: 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, ω is a finite collection of pairs $(P, \boldsymbol{\omega}')$, where $\boldsymbol{\omega}' = (\omega'_s)_{s \in S}$, so that Ω is isomorphic to $\mathcal{K}_\flat(\mathcal{P} \times \Omega^S)$.[21] We call Ω the space of *Recursive Information Constraints* (RICS) so that each $\omega \in \Omega$ is an RIC. Conversely, every RIC is also an ICP. (Indeed, set $\Gamma^*(\omega) = \{P : (P, \boldsymbol{\omega}') \in \omega\}$ and $\tau^*(\omega, P, s) = \omega'_s$ to obtain the ICP $\mathcal{M}_\omega = (\Omega, \omega, \Gamma^*, \tau^*)$ which is indistinguishable from ω.)

Proposition 3.3. The space $\mathbf{M}$ of ICPs is isomorphic to Ω 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 ω.

A proof is in Appendix A.6. The space Ω is a compact metric space, and has the property that finite horizon RICS approximate arbitrary RICS. Equivalently, this allows us to define a pseudo-metric on $\mathbf{M}$ as follows: The (pseudo-) distance between $\mathcal{M}$ and $\mathcal{M}'$ is precisely the distance between $\omega_{\mathcal{M}}$ and $\omega_{\mathcal{M}'}$.

Because finite horizon RICS approximate arbitrary RICS, we can define notions such as the Blackwell order for finite horizon ICPS and then take the appropriate limits. Moreover, viewing Ω as the canonical state space for ICPS implies that the recursive Blackwell order on Ω is the unique recursive order so defined that is continuous and satisfies our definition for comparing the informativeness of ICPS (see Section 2.3 for the definition and also Proposition A.3 in the appendix).

(20) For metric spaces X and Y, we denote by $\mathcal{K}_\flat(X \times Y)$ 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 $x \neq x'$.

(21) A formal construction of Ω is given in Appendix A.3. The metric on Ω can be intuitively described as follows: Consider $\omega, \omega' \in \Omega$ as ICPS. 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 ω and ω' is at most $1/2^n$. Thus, $\omega, \omega' \in \Omega$ are indistinguishable if, and only if, they are identical. The isomorphism between Ω and $\mathcal{K}_\flat(\mathcal{P} \times \Omega^S)$ is a homeomorphism.

3.3. Portable Lessons for other Problems with Unobservable Controls

Our insights are potentially valuable for other environments where DM faces dynamic choice problems, and must make an unobservable choice from a collection of subjective dynamic plans that an analyst would like to infer. For example, investments in human capital or maintenance of health stock are economically relevant (controlled) variables, but are chronically hard to observe directly.

Recycling notation from the previous sections, consider another environment where X represents the space of observable dynamic choice problems, each $\mathscr{M}$ represents the collection of unobservable plans available to DM, and $V(x, \theta)$ (which may also depend on objective state variables) is a function that evaluates the choice problems under the best plan from $\mathscr{M}$. 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 3.2. The following observations apply in all such settings.

- The set of plans constrained by $\mathscr{M}$ 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 $\mathscr{M}$. Clearly, identification of the set $\mathscr{M}$ 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 problems 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.

In Section 6 we apply our strategy to the identification of an alternative model to the one analyzed thus far, in which information bears a direct subjective intertemporal cost, rather than being subject to a constraint.

4. Behavioral Characterization

In this section we present an axiomatic foundation for our model. Given that the main focus of our paper is on identifying the parameters of an ICP-representation, in this section we describe the behavioral content of the axioms imposed on the preference relation $\succsim$ over X without burdening the reader with additional notation. Formal statements of all the axioms can be found in Appendix C.

The axioms broadly fall into three different categories: static, consumption stream, and recursive. Axioms 1 and 3–5 are *static* in the sense that they do not rely on the recursive structure of our domain; they simply restrict preferences on $\mathscr{K}\big(\mathscr{F}(\Delta(C \times X))\big)$, 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. It is only Axiom 6 that exploits the *recursive* structure of X in a novel way.

Axiom 1 summarizes a number of relatively standard properties in the menu-choice literature: the preference relation $\succsim$ is a weak order that satisfies Continuity, Lipschitz continuity (see Dekel, Lipman, and Rustichini 2001 and Dekel et al. 2007), Monotonicity with respect to set inclusion (bigger menus are better), and Aversion to Randomization (lower contour sets are convex). As pointed out above, information plays no role when there is no consumption choice in the future. Following Krishna and Sadowski (2014), Axiom 2 requires $\succsim|_L$ to satisfy standard versions of (vN-M) Independence, History Independence, and Stationarity.

The motivation for our more novel axioms (Axioms 3–6) is based on the type of information choice 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. We now discuss the extent to which the familiar properties of Temporal Separability, Strategic Rationality, Independence, and Stationarity are satisfied when the analyst is not able observe information choice and hence cannot condition behavior on it.[22]

4.1. Versions of Familiar Axioms

The axioms discussed in this section are non-recursive, in the sense that they do not rely on the recursive structure of X.

Temporal Separability. DM must choose what to learn before picking an act, and this choice may affect the continuation constraint and hence the value of possible continuation problems. We assume that the effect of information choice on the continuation value of an act depends only the marginal distribution over continuation problems the act induces in each state. That is, for any given finite menu x, DM's optimal learning will not change when substituting act $f \in x$ with g as long as both induce the same marginal distributions over C and X in each state s. Our axiom *State-Contingent Indifference to Correlation* (Axiom 3) posits that the value of the menu is unchanged under such a substitution.[23]

Strategic Rationality. Suppose that DM is offered the chance to replace a certain continuation problem with another. Clearly, any continuation problem y should leave DM no

(22) We remark that all our axioms except our continuity requirement can be falsified with finite data.

(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* choice problem x, rather than just singletons, because different information may be optimal for different choice problems.

worse off than receiving the worst consumption stream, ℓ_*. This is captured by part (a) of *Indifference to Incentivized Contingent Commitment* (Axiom 4). In general, DM's attitude towards such replacements may depend on his initial information choice, which is subjective, unobserved, and menu-dependent. Part (b) investigates the conditions under which DM is actually *indifferent* to replacing continuation lotteries with the worst consumption stream. The natural inference is that DM does not expect to reach these continuation lotteries.

Recall that the ICP requires DM to choose a partition of S. 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. That is, there should be a contingent plan that specifies which act DM will choose for each state, such that he is indifferent between the original menu and one where he is penalized (by receiving the worst consumption stream) whenever his choice does not coincide with that plan.[24]

Independence. We say that x and y are *concordant* if the same initial information choices are optimal for both x and y. In general, DM can potentially tailor his information choice to the choice problems x and y separately, while at the mixed menu $\frac{1}{2}x + \frac{1}{2}y$ he may need to compromise. However, if x and y are concordant — so that the optimal partitions for x and y are the same — then the same information choices are also optimal for the mixed menu. That is, $\frac{1}{2}x + \frac{1}{2}y$ should be concordant with x, if x is concordant with y. More generally, $\succsim$ restricted to a set of concordant choice problems should satisfy Independence. These requirements are the content of *Concordant Independence* (Axiom 5).

The key is to verify concordance of the choice problems x and y from behavior. We do this by perturbing x and y with the menu $x_1(P)$ — which requires DM to bet on the cells of the partition P in the first period and requires no choice after that, see [♠] — and then verifying that for each $P \in \mathscr{P}$, the marginal 'impact' of the perturbation is the same for both x and y.[25]

4.2. Self-Generation

We shall call a preference relation 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. Our axiom *Self-Generation* (Axiom 6) captures this insight by taking the axioms for choice discussed so far, none of which depend on the recursive structure of X, and turning them into a *recursive system of axioms* on X that apply at each date and state

(24) This is conceptually related to the Indifference to State Contingent Commitment Axiom introduced in Dillenberger et al. (2014). Both axioms relate partitional learning to a state contingent notion of strategic rationality which is a willingness to commit, as in Kreps (1979).

(25) Put differently, if x and y were not concordant, then there would be some partition Q for which the marginal impact of perturbing the menus by $x_1(Q)$ is different.

and after every history of choice. Note that Self-Generation is satisfied in any recursive model because in such a model, continuation preferences have the same form as ex ante preferences.

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 $\succsim_{(P,s)}$. Because P is unobservable, $\succsim_{(P,s)}$ must be inferred from behavior (ie, from $\succsim$). Towards this end, we rely only on the initial ranking of choice problems that gives rise to the same optimal choice of partition P.

To gain intuition, suppose P is the unique optimal choice for the choice problem 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 $y \succsim_{(P,s)} y'$.

Self-Generation (Axiom 6) then requires the following:

The binary relation $\succsim$ must be in

$$\Psi^* := \left\{\succsim' \text{ on } X : \text{(i) } \succsim' \text{ satisfies Axioms 1–5, and (ii) } \succsim'_{(P,s)} \in \Psi^*\right\}$$

The set Ψ^* can be viewed as 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 3, characterizes the largest such set Ψ^* via a well defined recursive value function, and establishes that it is non-empty.

It is important to note the self-referential character of Self-Generation, which requires induced preferences over the next period's continuation problems to be in Ψ^*, thereby requiring that preferences over continuation problems two periods ahead again have to be in Ψ^*, and so forth. One could, alternatively, write our axiom in extensive form, in which case it would simply require induced preferences after every history of state realizations and information choices to satisfy Axioms 1–5.

Finally, we observe that Axiom 6 is weaker than the Stationarity axiom in Gul and Pesendorfer (2004) or KS, in the sense that it only requires immediate and continuation preferences to be of the same type rather than identical.[26]

4.3. Representation Theorem

Theorem 3. *Let $\succsim$ be a binary relation on X. Then, the following are equivalent:*

(a) *$\succsim$ satisfies Axioms 1–6.*

(b) *There exists an ICP-representation of $\succsim$.*

(26) Note that we while place restrictions on contingent ex post preferences, Gul and Pesendorfer (2004) and KS restrict aggregated future preferences. In any event, both of their representations satisfy a suitable formulation of Self-Generation.

The proof of Theorem 3 is quite involved. To lay the groundwork for the proof, we establish the following representation of $\succsim$ (as Theorem 2 in the Supplementary Appendix), which is the starting point for our proof in Appendix D:

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

where $\mathbb{Q} \subset \mathscr{P}$ is a set of partitions of S, the measure $\pi(s \mid I)$ is the probability of s conditional on the event $I \subset S$, and utilities (v_s) over continuation problems depend only on the partition P. We say that V is *implemented* by $((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi)$. This representation already has all the features we need to establish, except that it is static; it does not exploit the recursive structure of X. Correspondingly, we do not rely on Axiom 6, but only on Axioms 1–5 to derive it. We also use Axiom 2 to establish a recursive representation of $\succsim|_L$, which is the *Recursive Anscombe-Aumann* representation in Krishna and Sadowski (2014), parametrized by $((u_s), \delta, \Pi)$, and discussed in Appendix D.1.

Adapting the terminology of Abreu, Pearce, and Stacchetti (1990), we define the set Φ^* of *self-generating* value functions, where each $v \in \Phi^*$ is implemented by some tuple $((u_s), \mathbb{Q}, (w_s(\cdot, P)), \pi)$ in a way that each $w_s(\cdot, P)$ is itself in Φ^* (see Appendix D.2). In Appendix D.3, we rely on Self-Generation (Axiom 6) to show that the representation V of $\succsim$ can be made self-generating. Clearly, a self-generating value function may not be recursive.

The remainder of our construction in Appendix D.4 has two main components. First, we construct an RIC ω_0 from a self-generating representation and argue that any other self-generating representation of $\succsim$ would yield the same ω_0 up to recursive Blackwell dominance. The intuition for this construction is precisely the one in our proof of Theorem 1, where we elicit ω_0 from $\succsim$ without having to elicit beliefs.

Second, we note the the self-generating representation above and the Recursive Anscombe-Aumann representations must agree on L. This lets us conclude that we can pair the parameters $((u_s), \delta, \Pi)$ with ω_0 to find the ICP-representation, $((u_s), \delta, \Pi, \omega_0)$, which is recursive on all of X, and where ω_0 is a canonical ICP. 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 ω in our representation, so that the representation becomes recursive when conditioning on both s and ω.

In Appendix C.6 we discuss how to further strengthen our notions of Stationarity and Separability to characterize the special case of the ICP representation where 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. Independently, we also confirm that imposing full Independence implies that information is not determined by a choice process, but instead exogenously arrives over time.

5. Related Literature

We now comment on the menu choice literature that shares some of our basic assumptions.

Kreps (1979) studies choice between menus of prizes. He rationalizes *monotonic preferences* — those that exhibit 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) replace Independence with *Aversion to Randomization* to model subjective uncertainty that is not fixed, but a choice variable.[27] 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, we consider menus of acts over lotteries, which are appropriate for a dynamic extension of Dillenberger et al. (2014). Our model also extends De Oliveira et al. (2016), in the sense that the choice of information in a period now affects the feasible choices of information in the future.[28]

DM controls his information over time. Thus, 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 to be made in the future. We then use the ranking of consumption streams to 'calibrate' preferences over all dynamic choice problems, 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. Table 1

(27) 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.

(28) 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, as we do in Section 6.

summarizes the position of our model with respect to the papers discussed thus far.

Information	***Uncertainty about vN-M taste***	***Uncertainty about state of world***
Static, fixed	Dekel, Lipman, and Rustichini (2001)	Dillenberger et al. (2014)
Static, choice	Ergin and Sarver (2010)	De Oliveira et al. (2016)
Dynamic, fixed	Krishna and Sadowski (2014)	*This paper*
Dynamic, choice process	★	*This paper*

Table 1: *Summary of Literature*

There are additional works that address unobservable information acquisition in different formal contexts. They include Piermont, Takeoka, and Teper (2016), who 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 menu-choice approach. Ellis (2016) identifies a partitional information constraint from ex post choice data. Caplin and Dean (2015) 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 Matejka and McKay (2014). To our knowledge, there is no counterpart to our recursive analysis of information choice in the random-choice literature.

6. Direct information Costs

ICPS can generate opportunity costs of information acquisition via tighter future constraints. Our main model allows us to focus *entirely* on the behavioral implications of this new type of dynamic costs. That said, 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 (2015), De Oliveira et al. (2016), and Hébert and Woodford (2016)).[29] In this section we (i) provide a recursive model based on *intertemporal* costs rather than ICPS, (ii) argue that the heart of our identification strategy is robust to this change,[30] and (iii) discuss the benefits of dealing with constraints rather than costs for identification of (or inference about) the relevant parameters.

(29) 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 ICPS 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.

(30) A formal treatment can be found in Section 1 of the Supplementary Appendix.

A *cost structure* $\mathscr{C}$ is a tuple $(\Theta, \theta_0, \tau, \rho)$, where Θ, θ_0, and τ are as before, and $\rho : \mathscr{P} \times \Theta \to \mathbb{R}_+$ is a *cost function* that determines the cost of learning a partition as a function of the cost state θ.[31] As with ICPs, after the choice of P and the realization of s, the transition operator τ determines the continuation control state $\theta' = \tau(P, s, \theta)$. We assume that $\rho(\{S\}, \theta) = 0$ for all θ. With timing unchanged from before, we can show (see Section 1.1 of the Supplementary Appendix) that $\mathscr{C}$ induces a unique value function

$$V(x, \theta, s) = \max_{P \in \mathscr{P}} \left[\sum_{J \in P} \left(\max_{f \in x} \sum_{s' \in J} \pi(s' \mid J)\, \mathbf{E}^{f(s')} \left[u_{s'}(c) + \delta V(y, \tau(P, s, \theta), s') \right] \right) - \rho(P, \theta) \right]$$

We refer to $\big((u_s), \delta, \Pi, \mathscr{C}\big)$ as an *Information Cost* representation if $V(\cdot, \theta_0, s_0) : X \to \mathbb{R}$ represents preferences on X.

In order to compare two cost structures, $\mathscr{C}$ and $\mathscr{C}'$, we must describe how costly they make it to follow a particular dynamic information plan. Recall that a dynamic information plan prescribes a choice of partition as a function of (x, θ, s). (See Section 2.2 for a definition of dynamic information plans.)

We say that the cost structure $\mathscr{C}$ *dominates* $\mathscr{C}'$ if for any information plan σ' there is another plan σ such that (i) σ is more informative than σ' at every date and state, and (ii) the expected cost of σ under the cost structure $\mathscr{C}$ is less than that of σ' under the cost structure $\mathscr{C}'$.

Theorem 4. *Let $\big((u_s), \delta, \Pi, \mathscr{C}\big)$ be an Information Cost representation for $\succsim$. Then*

- *The functions $(u_s)_{s \in S}$ are unique up to addition of constants and common scaling*
- *δ and Π are unique*
- *If the functions $(u_s)_{s \in S}$ are unbounded,[32] then $\mathscr{C}$ is unique up to dominance.*

The intuition for the identification result is similar to that described in Section 3.1 for the case of ICPs with the following qualifiers. First, as is apparent from the theorem and immediately intuitive, to identify potentially arbitrarily high information costs, arbitrarily high stakes are needed. This requires utilities to be unbounded, and the analyst to elicit those utilities before eliciting the cost structure. In contrast, identification of the ICP relies only on two outcomes that are strictly ranked in some state (recall that we only require *some* u_s to be non-trivial), which is much less demanding.

Second, the expected cost of a particular information plan, and hence our notion of dominance between cost structures, depends on the discount factor δ and the evolution of the payoff relevant state in S according to Π. In contrast, recursive Blackwell dominance between ICPs is independent of other preference parameters.

(31) It is easy to see that an ICP is a cost structure where $\rho(\cdot) \in \{0, \infty\}$.

(32) Unbounded utilities need a non-compact domain. We show in Section 1.1 of the Supplementary Appendix that our approach extends to a domain of dynamic choice problems with a sigma-compact consumption space.

Finally, constructing a choice problem that incentivizes a particular information plan σ requires mixing over continuation problems in exactly the right proportions to ensure that said plan corresponds to the optimal information strategy. In contrast, the exact probabilities assigned to different continuations problems is irrelevant in the identification argument for the ICP (see Section 3.1.) As a consequence, while our general identification strategy is robust to the consideration of information costs instead of constraints, drawing inference about cost structures in applied contexts will be much more involved than the procedure for inference about constraints that we discuss next in Section 7.1.

7. Concluding Remarks

7.1. Inference about the ICP from Limited Data

In applications, inference about the ICP benefits from three features of the identification strategy outlined in Section 3.1. First, identification of the ICP is (almost) independent of the other preference parameters, as it only uses the best and worst outcomes in each state. Second, while identification of the ICP relies on randomization over continuation problems, the weights used in this randomization are not important, and so can be dictated by the application at hand. Finally, to verify whether DM can follow a particular information plan, the analyst only needs to observe *one* appropriate binary choice – that between the best consumption stream and a choice problem that is uniformly strongly aligned with the plan in question.

To illustrate, imagine an agent who must routinely make a particular type of decision, for example when to service various pieces of equipment. Further suppose the agent has the option to consult an expert who can commit to providing information (as might be found, more generally, in a "market for signals"), which would allow him to make perfect decisions, thereby generating the best possible stream of outcomes. If the agent is unwilling to bear any positive cost of consulting the expert, then an analyst can infer that the agent himself must expect to be able to process all the relevant information, and hence to generate the best possible stream of outcomes on his own. Similarly, if the agent consults the expert only on some decisions, then he must at least expect to be able to process the information that is relevant for the remaining ones. The inference in this stylized example relies only on a very small amount of choice data (related to the third point above), does not require the analyst to know the cost the agent incurs for the wrong decision (related to the first point), and does not depend on the exact probability with which a particular decision comes up at any given point in time (related to the second point).

7.2. Learning the Payoff-relevant State after each Period

As is apparent from Equation [**Val**] in Section 2.2, last period's state of nature $s \in S$ is a state-variable in our recursive model, that is, DM always learns the realized state of nature at the end of a period. This is natural given our environment.

To see this, notice that the realized pair (c, y) of consumption and continuation problem has information about the realized state of the world. If DM did not observe the realized state, he would be willing to pay a premium for acts that reveal the state. Because there are such 'fully revealing' acts that are arbitrarily close to 'non-revealing' acts, not learning the state would lead to violations of continuity that strike us as implausible. For instance, DM might be able to place side bets on fully revealing acts with arbitrarily small stakes, effectively allowing him to learn the state at an arbitrarily small cost. Therefore, we simply assume that the state becomes known for free at the end of the period.[33]

7.3. ICPs 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 ICPs in the context of single-person decision making, it will be interesting to think about the strategic interaction of agents thus-constrained. To suggest just one instance, consider a monopolistic competition setting where each firm faces an ICP as in Example 1.2; 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 every other period — or will we observe heterogeneous behavior with constant volatility?[34]

(33) An alternative model could assume that DM learns neither state nor realized continuation problem at the end of a period. This would require modeling 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).

(34) We describe the problem as a firm's decision of *when* to pay attention. For example, in Maćkowiak and Wiederholdt (2009), firms also need to decide *what* to pay attention to. Some other works, such as Myatt and Wallace (2012), study a related problem of information acquisition in coordination games.

For Online Publication

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 $\|\varphi\|_{\mathrm{L}} := \sup_{y \neq y'} |\varphi(y) - \varphi(y')| / d_Y(y, y')$. This allows us to define the *bounded Lipschitz* norm $\|\varphi\|_{\mathrm{BL}} := \|\varphi\|_{\mathrm{L}} + \|\varphi\|_\infty$. Then, $\mathrm{BL}(Y) := \{\varphi \in \mathbb{R}^Y : \|\varphi\|_{\mathrm{BL}} < \infty\}$ is the space of real-valued, bounded, and Lipschitz functions on Y.

Define d_D on $\Delta(Y)$ as $d_D(\alpha, \beta) := \frac{1}{2} \sup \left\{ \left| \int \varphi \, \mathrm{d}\alpha - \int \varphi \, \mathrm{d}\beta \right| : \|\varphi\|_{\mathrm{BL}} \leq 1 \right\}$. 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. Recursive Domain

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

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

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

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

Now define recursively, for $n > 1$, $X_n := \mathcal{K}\big(\mathcal{F}(\Delta(C \times X_{n-1}))\big)$. The metric on $C \times X_{n-1}$ is the product metric; that is, $d_{C \times X_{n-1}}\big((c, x_{n-1}), (c', x'_{n-1})\big) = \max[d_C(c, c'), d^{(n-1)}(x_{n-1}, x'_{n-1})]$. 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 $\bigtimes_{n=1}^{\infty} X_n$ with the product topology. The Tychonoff metric induces this topology and is given as follows: For $x = (x_1, x_2, \dots), y = (y_1, y_2, \dots) \in \bigtimes_{n=1}^{\infty} X_n$,

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

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

The space of choice problems X is all members of $\bigtimes_{n=1}^{\infty} X_n$ 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}.[35] The space of choice problems, X, is our domain for choice, and it follows from standard arguments that X is (linearly) homeomorphic to $\mathcal{K}\big(\mathcal{F}(\Delta(C \times X))\big)$. We denote this homeomorphism by $X \simeq \mathcal{K}\big(\mathcal{F}(\Delta(C \times X))\big)$. In what follows, we shall abuse notation and use d as a metric both on X as well as $\mathcal{K}\big(\mathcal{F}(\Delta(C \times X))\big)$. 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 choice problems: For $x, y \in X$, $y \subset x$ if $y_n \subset x_n$ for all $n \geq 1$.

A.3. Recursive Information Constraints

Recall that $\mathcal{P}$ is the space of all partitions of S, where a typical partition is P. The partition P is *finer* than the partition Q if every cell in Q is the union of cells in P. For a partition P, define its *entropy* $H(P)$ as $H(P) := -\sum_{J \in P} \mu(J) \log \mu(J)$. Then, we can define a metric d on $\mathcal{P}$ as $\mathsf{d}(P, Q) := 2H(P \wedge Q) - H(P) - H(Q)$, where $P \wedge Q$ is the coarsest refinement of P and Q. In Section 5 of the Supplementary Appendix, we show that d is indeed a metric. Thus, $(\mathcal{P}, \mathsf{d})$ is a metric space.

Let $\Omega_1 := \mathcal{K}(\mathcal{P})$, and define recursively for $n > 1$, $\Omega_n := \mathcal{K}_\flat\left(\mathcal{P} \times \Omega_{n-1}^S\right)$ (see Section 3.2 for a definition of $\mathcal{K}_\flat$). Then, we can set $\Omega' := \bigtimes_{n=1}^{\infty} \Omega_n$. A typical member of Ω_n is ω_n, while $\boldsymbol{\omega}_n = (\omega_{n,s})_{s \in S}$ denotes a typical member of Ω_n^S.

Let $\psi_1 : \mathcal{P} \times \Omega_1^S \to \mathcal{P}$ be given by $\psi_1(P, \boldsymbol{\omega}_1) = P$, and define $\Psi_1 : \Omega_2 \to \Omega_1$ as $\Psi_1(\omega_2) := \{\psi_1(P, \boldsymbol{\omega}_1) : (P, \boldsymbol{\omega}_1) \in \omega_2\}$. Now define recursively, for $n > 1$, $\psi_n : \mathcal{P} \times \Omega_n^S \to \mathcal{P} \times \Omega_{n-1}^S$ as $\psi_n(P, \boldsymbol{\omega}_n) := \big(P, (\Psi_{n-1}(\omega_{n,s}))_s\big)$, and $\Psi_n : \Omega_{n+1} \to \Omega_n$ by $\Psi_n(\omega_{n+1}) := \{\psi_n(P, \boldsymbol{\omega}_n) : (P, \boldsymbol{\omega}_n) \in \omega_{n+1}\}$.

An $\omega \in \Omega'$ is *consistent* if $\omega_{n-1} = \Psi_{n-1}(\omega_n)$ for all $n > 1$. The set of *Recursive Information Constraint*s (RICs) is

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

that is, the set of RICs is the space of all consistent elements of Ω'.

Notice that Ω_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 Ω_n with

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

the Hausdorff metric in turn makes it a compact metric space. Thus, Ω endowed with the product metric is a compact metric space. (Moreover, Ω 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 5. *The set Ω is homeomorphic to $\mathcal{K}_\flat(\mathcal{P} \times \Omega^S)$.*

We write the homeomorphism as $\Omega \simeq \mathcal{K}_\flat(\mathcal{P} \times \Omega^S)$. 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. Value Function

We now prove Proposition 2.2 for the case of canonical ICPs, ie, RICs. The extension to the case of general ICPs 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.

Proof of Proposition 2.2. Define the operator $T : \mathbf{C}(X \times \Omega \times (S \cup \{s_0\})) \to \mathbf{C}(X \times \Omega \times (S \cup \{s_0\}))$ as follows:

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

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 $\big((u_s)_{s \in S}, \Pi, \delta, \omega\big)$, 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 $(x, \omega, s') \mapsto (P, \boldsymbol{\omega}') \in \omega$. Because the set of such $(P, \boldsymbol{\omega}')$ is finite, it follows that there is a *conserving plan*.[36] 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). □

A.5. Recursive Blackwell Order

In this section, we construct the recursive Blackwell order for RICs. Appendix A.6 exhibits an isomorphism between RICs and ICPs. The isomorphism now induces the recursive Blackwell order on ICPs.

Let $\hat{\omega} \in \Omega$ denote the RIC that delivers the coarsest partition in each period in every state. Define $\hat{\Omega}_0 := \mathcal{K}_\flat(\mathcal{P} \times \{\hat{\omega}\})$, and inductively define $\hat{\Omega}_{n+1} := \mathcal{K}_\flat(\mathcal{P} \times \hat{\Omega}_n)$ 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 $\gtrsim_0$ on $\hat{\Omega}_0$ as follows: $\omega_0 \gtrsim_0 \omega'_0$ if for all

(36) 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).

$(P', \hat{\boldsymbol{\omega}}) \in \omega'_0$, there exists $(P, \hat{\boldsymbol{\omega}}) \in \omega_0$ such that P is finer than P'. This allows us to define inductively, for all $n \geq 1$, an order $\gtrsim_n$ on $\hat{\Omega}_n$. For all $\omega_n, \omega'_n \in \hat{\Omega}_n$, $\omega_n \gtrsim_n \omega'_n$ if for all $(P', \boldsymbol{\omega}'_{n-1}) \in \omega'_n$, there exists $(P, \boldsymbol{\omega}_{n-1}) \in \omega_n$ such that (i) P is finer than P', and (ii) $\omega_{n-1,s} \gtrsim_{n-1} \omega_{n-1,s}$ for all $s \in S$.

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

Lemma A.1. For all $n \geq 0$, $\gtrsim_{n+1}$ extends $\gtrsim_n$, ie, $\gtrsim_{n+1} |_{\hat{\Omega}_n} = \gtrsim_n$.

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 $\omega_0 \gtrsim_0 \omega'_0$. Then, for $(P', \hat{\boldsymbol{\omega}}) \in \omega'_0$, there exists $(P, \hat{\boldsymbol{\omega}}) \in \omega_0$ such that P is finer than P'. Moreover, because $\gtrsim_0$ is reflexive, $\hat{\boldsymbol{\omega}} \gtrsim_0 \hat{\boldsymbol{\omega}}$. But this implies $\omega_0 \gtrsim_1 \omega'_0$. Conversely, let $\omega_0 \gtrsim_1 \omega'_0$. Then, for all $(P', \hat{\boldsymbol{\omega}}) \in \omega'_0$, there exists $(P, \hat{\boldsymbol{\omega}}) \in \omega_0$ such that (i) P is finer than P', and (ii) $\hat{\omega} \gtrsim_0 \hat{\omega}$ for all $s \in S$. But this implies $\omega_0 \gtrsim_0 \omega'_0$, which proves that $\gtrsim_{n+1} |_{\hat{\Omega}_n} = \gtrsim_n$ when $n = 0$.

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

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

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

By definition of $\hat{\Omega}$, there is some n such that $\omega, \omega' \in \hat{\Omega}_n$, and by Lemma A.1, the precise choice of this n is irrelevant. This implies $\gtrsim$ is well defined. We now show that $\gtrsim$ has a recursive definition as well.

Proposition A.2. For any $\omega, \omega' \in \hat{\Omega}$, the following are equivalent.

(a) $\omega \gtrsim \omega'$.

(b) for all $(P', \tilde{\boldsymbol{\omega}}') \in \omega'$, there exists $(P, \tilde{\boldsymbol{\omega}}) \in \omega$ such that (i) P is finer than P', and (ii) $\tilde{\omega}_s \gtrsim \tilde{\omega}'_s$ for all $s \in S$.

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

Proof. (a) implies (b). Suppose $\omega \gtrsim \omega'$. Then, by definition, there exists n such that $\omega, \omega' \in \hat{\Omega}_n$ and $\omega \gtrsim_n \omega'$. This implies that for all $(P', \tilde{\boldsymbol{\omega}}'_{n-1}) \in \omega'_n$, there exists $(P, \tilde{\boldsymbol{\omega}}_{n-1}) \in \omega_n$ such that (i) P is finer than P', and (ii) $\tilde{\omega}_{n-1,s} \gtrsim_{n-1} \tilde{\omega}'_{n-1,s}$ for all $s \in S$. But the latter property implies $\tilde{\omega}_s \gtrsim \tilde{\omega}'_s$ for all $s \in S$, which establishes (b). The proof that (b) implies (a) is similar and is therefore omitted.

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

We can now prove the existence of a recursive order on Ω. (Notice that $\mathrm{cl}(\hat{\Omega}) = \Omega$.) In particular, for all $\omega, \omega' \in \Omega$, we say that ω *recursively Blackwell dominates* ω' if for all $(P', \tilde{\boldsymbol{\omega}}') \in \omega'$, there exists $(P, \tilde{\boldsymbol{\omega}}) \in \omega$ 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.3. The order $\gtrsim$ on $\hat{\Omega}$ has a unique continuous extension to Ω, also denoted by $\gtrsim$. Moreover, on Ω, $\gtrsim$ is the unique non-trivial and continuous recursive Blackwell order.

Proof. Because $\Omega = \mathrm{cl}(\hat{\Omega})$, we simply extend $\gtrsim$ to Ω by re-defining it to be $\mathrm{cl}(\gtrsim)$. It is easy to see that $\gtrsim$ so defined is continuous and non-trivial. That $\gtrsim$ is a unique recursive Blackwell order follows immediately from the facts that $\hat{\Omega}$ is dense in Ω, the continuity of $\gtrsim$, and Proposition A.2. □

Let $\mathrm{proj}_n : \Omega \to \hat{\Omega}_n$ be the natural map associating with each ω, the 'truncated and concatenated' version ω_n which offers the same choices of partition as ω 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 (ω_n) is Cauchy, and converges to ω. The next corollary gives us an easy way to establish dominance.

Corollary A.4. For $\omega, \omega' \in \Omega$, $\omega \gtrsim \omega'$ if, and only if, for all $n \in \mathbb{N}$, $\omega_n \gtrsim \omega'_n$.

Proof. The 'only if' part is straightforward. The 'if' part follows from the continuity of $\gtrsim$. □

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

Corollary A.5. For all $\omega, \omega' \in \Omega$ and $m \geq 1$, $\omega_m \gtrsim \omega'_m$ implies $\omega_n \gtrsim \omega'_n$ for all $1 \leq n \leq m$.

Proof. Notice that $\omega_m, \omega'_m \in \Omega$. Therefore, by Corollary A.4, it follows that for all $n \geq 1$, $\mathrm{proj}_n\, \omega_m \gtrsim \mathrm{proj}_n\, \omega'_m$. For $n \geq m$, $\mathrm{proj}_n\, \omega_m = \omega_m$, but for $n \leq m$, $\mathrm{proj}_n\, \omega_m = \omega_n$, which implies that for all $n \leq m$, $\omega_n \gtrsim \omega'_n$. □

A.6. Isomorphisms of ICPs

Proof of Proposition 3.3. We first show that (a) implies (b). Towards this end, let $\mathcal{M} = (\Theta, \theta_0, \mathcal{P}, \Gamma, \tau)$ be an ICP. Recall the definition of the space Ω_n from Appendix A.3 and define the maps $\Phi_n : \Theta \to \Omega_n$ as follows. Let

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

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

Now, given θ_0, set $\Phi_n(\theta_0) =: \omega_n \in \Omega_n$. It is easy to see that the sequence

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

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 ICP $\mathcal{M}$ corresponds to an RIC ω.

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

Thus, Ω is the space of canonical ICPs in that every ICP can be embedded in Ω. Let $\mathscr{M} = (\Theta, \Gamma, \tau, \theta_0)$ and $\mathscr{M}' = (\Theta', \Gamma', \tau', \theta_0')$ be two ICPs in $\mathbf{M}$. Define the function $\mathsf{D} : \mathbf{M} \times \mathbf{M} \to \mathbb{R}$ as follows:

[**A.1**]
$$\mathsf{D}\left(\mathscr{M}(\theta_0), \mathscr{M}'(\theta_0')\right) := \max\left[d_H\left(\Gamma(\theta_0), \Gamma'(\theta_0')\right) \wedge 1, \tfrac{1}{2} \max_{P \in \Gamma(\theta_0), s \in S} \mathsf{D}\left(\mathscr{M}(\tau(\theta_0, P, s)), \mathscr{M}'(\tau'(\theta_0', P, s))\right)\right]$$

where $\mathscr{M}(\theta)$ denotes the ICP $\mathscr{M}$ with initial state θ. The function D captures the *discrepancy* between the ICPs $\mathscr{M}$ and $\mathscr{M}'$. 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.6. There is a unique function $\mathsf{D} \in \mathbf{B}(\mathbf{M} \times \mathbf{M})$ that satisfies equation [A.6].

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

$$T\mathsf{D}'\left(\mathscr{M}(\theta_0), \mathscr{M}'(\theta_0')\right) := \max\left[d_H\left(\Gamma(\theta_0), \Gamma'(\theta_0')\right) \wedge 1, \tfrac{1}{2} \max_{P \in \Gamma(\theta_0), s \in S} \mathsf{D}'\left(\mathscr{M}(\tau(\theta_0, P, s)), \mathscr{M}'(\tau'(\theta_0', P, s))\right)\right]$$

for all $\mathsf{D}' \in \mathbf{B}(\mathbf{M} \times \mathbf{M})$. It is easy to see that T is monotone in the sense that $\mathsf{D}_1 \leq \mathsf{D}_2$ implies $T\mathsf{D}_1 \leq T\mathsf{D}_2$. It also satisfies discounting, ie, $T(\mathsf{D} + a) \leq T\mathsf{D} + \frac{1}{2}a$ 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.6]. □

We can now define an isomorphism between ICPs. Two ICPs $\mathscr{M}$ and $\mathscr{M}'$ are *indistinguishable* if $\mathsf{D}\left(\mathscr{M}(\theta_0), \mathscr{M}'(\theta_0')\right) = 0$. Intuitively, indistinguishable ICPs 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.7. Let $\mathscr{M}, \mathscr{M}' \in \mathbf{M}$. Then, $\mathscr{M}$ is indistinguishable from $\mathscr{M}'$ 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 ICP $(\Theta, \Gamma, \tau, \tau(\theta_0, P, s))$ is indistinguishable from the ICP $\left(\Theta', \Gamma', \tau', \tau'(\theta_0', P, s)\right)$.

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

Corollary A.8. Let $\omega, \omega' \in \Omega$. Then, $\omega \neq \omega'$ implies $\mathsf{D}(\omega, \omega') > 0$.

Proof. It is easy to see that if $\mathsf{D}(\omega, \omega') = 0$, then $\omega_n = \omega_n'$ for all $n \geq 1$, which implies $\omega = \omega'$, as required. □

Corollary A.9. Let $\omega, \omega' \in \Omega$ be such that $\text{proj}_n(\omega) \not\gtrsim \text{proj}_n(\omega')$ for some $n \geq 1$, but for all $m < n$, $\text{proj}_m(\omega) \gtrsim \text{proj}_m(\omega')$. Then, there exists finite sequences $(P_k)_1^{n-1}$ and $(s_k)_1^{n-1}$ which induce RICS $\omega^i_{(n-k)} := \tau^*(\omega^i_{(n-k+1)}, P_k, s_k) \in \hat{\Omega}_{n-k}$ where $P_k \in \Gamma^*(\omega^i_{n-k+1})$, such that $\Gamma^*(\omega^1_1)$ does not setwise-Blackwell dominate $\Gamma^*(\omega^2_1)$.

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

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

[✠] $\mathcal{M} \trianglerighteq \mathcal{M}'$ 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) $(\Theta, \Gamma, \tau, \tau(\theta_0, P, s)) \trianglerighteq (\Theta', \Gamma', \tau', \tau'(\theta_0', P', s))$ for all $s \in S$.

It is easy to see that such a recursive order exists. Indeed, for any ICP $\mathcal{M}$, let $\omega_{\mathcal{M}}$ denote the canonical ICP that is indistinguishable from it. (By Corollary A.8 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) $\omega_{\mathcal{M}} \gtrsim \omega_{\mathcal{M}'}$. The recursive Blackwell order on $\mathbf{M}$ clearly satisfies the condition [✠]. We now demonstrate that it is the *largest* order that satisfies [✠].

Proposition A.10. Let $\trianglerighteq$ be a recursive order on $\mathbf{M}$ that satisfies [✠]. If $\mathcal{M} \trianglerighteq \mathcal{M}'$, 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 $\omega_{\mathcal{M}} \not\gtrsim \omega_{\mathcal{M}'}$. Corollaries A.4 and A.5 imply that there is a smallest n such that $\omega_{\mathcal{M},n} \not\gtrsim \omega_{\mathcal{M}',n}$ but that for all $m < n$, $\omega_{\mathcal{M},m} \gtrsim \omega_{\mathcal{M}',m}$ (where $\omega_{\mathcal{M},n} = \operatorname{proj}_n \omega_{\mathcal{M}}$ as defined in Appendix A.5). From Corollary A.9 it follows that there exists a finite sequence of partitions (P_k) and states (s_k) such that $\Gamma^*\big(\tau^{*(n)}(\theta_0, (P_k), (s_k))\big)$ does not setwise Blackwell dominate $\Gamma^*\big(\tau^{*(n)}(\theta_0', (P_k), (s_k))\big)$, 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 $\Gamma\big(\tau^{(n)}(\theta_0, (P_k), (s_k))\big)$ does not setwise Blackwell dominate $\Gamma'\big(\tau'^{(n)}(\theta_0', (P_k), (s_k))\big)$. Thus, it must necessarily be that $\mathcal{M} \not\trianglerighteq \mathcal{M}'$. $\square$

B. Identification and Behavioral Comparison: Proofs from Section 3

Based on the results established in Appendices A.5 and A.6, we now establish Theorems 1 and 2.

Recall that x is *strongly aligned* with ω if (i) $V(x, \omega, \pi_0) \geq V(x, \omega', \pi_0)$ for all $\omega' \in \Omega$, and (ii) ω' does not recursively Blackwell dominate ω implies $V(x, \omega, \pi_0) > V(x, \omega', \pi_0)$. We say that P is *supported* by ω if there exists $\boldsymbol{\omega}' \in \Omega^S$ such that $(P, \boldsymbol{\omega}') \in \omega$.

Lemma B.1. Let $(P, \boldsymbol{\omega}') \in \omega$. Then, there exists a choice problem $x(P, \boldsymbol{\omega}')$ recursively defined as

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

where $\operatorname{Unif}(\cdot)$ 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 ω, define $x_1(P) := \{f_{1,J} : J \in P\}$.

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

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

Then, given $(P, \boldsymbol{\omega}') \in \omega$, we have the menu $x_n(P, \boldsymbol{\omega}') := \left\{f_{n,J} : J \in P\right\}$.

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

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

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

as claimed. □

It is straightforward to verify that

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

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

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

Proof. Fix $(P, \boldsymbol{\omega}) \in \Omega$ and consider the menu $x(P, \boldsymbol{\omega})$ defined in [★]. As noted above, for all $\boldsymbol{\omega}'$, we have $V\left(x(P, \boldsymbol{\omega}), (P, \boldsymbol{\omega}), \pi_0\right) = V\left(x(P, \boldsymbol{\omega}'), (P, \boldsymbol{\omega}'), \pi_0\right)$. Moreover, it must be that for all $(Q, \boldsymbol{\omega}')$ (even for $Q = P$), we have $V\left(x(P, \boldsymbol{\omega}), (P, \boldsymbol{\omega}), \pi_0\right) \geq V\left(x(P, \boldsymbol{\omega}'), (Q, \boldsymbol{\omega}'), \pi_0\right)$ and in the case where Q is not finer than P and $Q \neq P$, $V\left(x(P, \boldsymbol{\omega}'), (P, \boldsymbol{\omega}'), \pi_0\right) > V\left(x(P, \boldsymbol{\omega}'), (Q, \boldsymbol{\omega}'), \pi_0\right)$ by construction of the menu $x(P, \boldsymbol{\omega}')$. (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 ω' does not recursively Blackwell dominate ω. Then, for some $(P, \tilde{\boldsymbol{\omega}}) \in \omega$, $x(P, \tilde{\boldsymbol{\omega}})$ defined in [★] is such that $V\left(x(P, \tilde{\boldsymbol{\omega}}), \omega, \pi_0\right) = V\left(x(P, \tilde{\boldsymbol{\omega}}), (P, \tilde{\boldsymbol{\omega}}), \pi_0\right) > V\left(x(P, \tilde{\boldsymbol{\omega}}), \omega', \pi_0\right)$.

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

From Corollary A.9 it follows that there exist finite sequences of partitions (P_k) and (P_k'), and states (s_k) such that $\Gamma^*\left(\tau^{*(n)}(\omega', (P_k'), (s_k))\right)$ does not setwise Blackwell dominate the set

$\Gamma^*\big(\tau^{*(n)}(\omega, (P_k), (s_k))\big)$, 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), $\omega^i_{n-k} = \tau^*(\omega^i_{n-k+1}, P_k, s_k)$ where $P_k \in \Gamma^*(\omega^i_{n-k+1})$, and $\Gamma^*(\omega^1_1)$.

Let $(P_1, \tilde{\boldsymbol{\omega}}) \in \omega$ be the unique first period choice under ω that makes the sequence (P_k) feasible. Then $x(P_1, \tilde{\boldsymbol{\omega}})$ defined in [★] is aligned with $(P_1, \tilde{\boldsymbol{\omega}})$. That is, after n stages of choice and a certain path of states we can appeal to Lemma B.2, which completes the proof. □

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 $\big((u_s), \Pi, \delta\big)$ is unique in the sense of the Theorem. Now, define $F_\omega := \{x(P, \tilde{\boldsymbol{\omega}}) : (P, \tilde{\boldsymbol{\omega}}) \in \omega\}$. It follows immediately from Lemma B.3 that F_ω is uniformly strongly aligned with ω. □

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

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

We are now in a position to prove Theorem 2.

Proof of Theorem 2. We first show the 'only if' part. On L, we have $\ell \succsim^\dagger \ell'$ implies $\ell \succsim \ell'$. This implies, by Lemma 34 of Krishna and Sadowski (2014), that $\succsim^\dagger|_L = \succsim|_L$. This, and the uniqueness of the RAA representation (Proposition 4.5) together imply that $\big((u_s), \delta, \Pi\big) = \big((u^\dagger_s), \delta^\dagger, \Pi^\dagger\big)$ 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. □

C. Axioms

In this section we formally present axioms on the preference $\succsim$ over X; by Theorem 3 these axioms are necessary and sufficient for an ICP representation as discussed in the text. In addition, Appendix C.6 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.

C.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 $\ell^\sharp, \ell_\sharp \in L$ and $N > 0$ such that for all $x, y \in X$ and $t \in (0, 1)$ with $t \geq Nd(x, y)$, we have $(1-t)x + t\ell^\sharp \succ (1-t)y + t\ell_\sharp$.

(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.[37] Item (e) is familiar from Ergin and Sarver (2010) and De Oliveira et al. (2016) and relaxes Independence in order to accomodate unobserved information choice.

The next axiom captures the special role played by consumption streams, which leave no consumption choice to be made in the future and therefore require no information (that is, all information alternatives perform equally well). The axiom thus requires $\succsim$ to satisfy additional standard assumptions when consumption streams are involved. In what follows, for any $c \in C$ and $\ell \in L$, let (c, ℓ) be the constant act that yields consumption c and continuation stream ℓ 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 3.1, we denote these by ℓ^* and ℓ_*, respectively. For each $I \subset S$, $f \in \mathscr{F}(\Delta(C \times X))$, $(c, y) \in C \times X$, and $\varepsilon \in [0, 1]$, define $f \oplus_{\varepsilon,I} y \in \mathscr{F}(\Delta(C \times X))$ by

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

That is, for any state $s \in I$, the act $f \oplus_{\varepsilon,I} y$ perturbs the continuation lottery with y.[38]

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).

(37) For a discussion of (c) see Dekel et al. (2007) and for (d) see Kreps (1979).

(38) 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.

C.2. State-Contingent Indifference to Correlation

As discussed in Section 4.1, we assume that DM's value for a menu does not change when substituting act f with g as long as they induce, on each state s, the same marginal distributions over C and X. 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 $[(x \setminus \{f\}) \cup \{g\}] \sim x$.[39]

C.3. Indifference to Incentivized Contingent Commitment

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 choice problem gives rise to a consumption stream. Therefore, any continuation problem y should leave DM no worse off than receiving the worst consumption stream, ℓ_*. In particular, because the best consumption stream, ℓ^*, leaves DM strictly better off than ℓ_* 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 ℓ_*.

Formally, let $X^* := \{(1-t)\,x + t\ell^* : x \in X \text{ is finite}, t \in (0,1)\}$. For a mapping $e : x \to (0,1]$, let $\left(x \oplus_{(e,s)} y\right) := \left\{f \oplus_{(e(f),s)} y : f \in x\right\}$, which perturbs the continuation lottery in state s for any act f in x by giving weight $e(f)$ to $\left(c_s^-, y\right)$. For $x \in X^*$ we then require $x \succ \left[x \oplus_{(e,s)} \ell_*\right]$ and $\left[x \oplus_{(e,s)} y\right] \succsim \left[x \oplus_{(e,s)} \ell_*\right]$ 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. In Section 4.1 we suggest a state contingent notion of strategic rationality, according to which 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 Ξ_x to be the collection of all functions $\xi : S \to x$. An *Incentivized Contingent Commitment* to $\xi \in \Xi_x$, is then the set

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

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

(39) 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* choice problem x, rather than just singletons, because different information may be optimal for different choice problems.

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

(a) If $x \in X^*$ and $e : x \to (0,1]$, then $x \succ \left[x \oplus_{(e,s)} \ell_*\right]$ and $\left[x \oplus_{(e,s)} y\right] \succsim \left[x \oplus_{(e,s)} \ell_*\right]$ for all $s \in S$ and $y \in X$.

(b) For all $x \in X$, there is $\xi \in \varXi_x$ such that $x \sim \mathcal{I}(\xi)$.

C.4. Concordant 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. As we argue in Section 4.1, 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 choice problems that lead to different optimal initial information choices, $\succsim|_{X'}$ should satisfy Independence if $X' \subset X$ consists only of concordant choice problems. We now introduce our behavioral notion of concordance (Definition C.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 choice problem 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 $x_1(P) := \left\{\ell^* \oplus_{(1,I^c)} \ell_* : I \in P\right\} \in \mathcal{K}(L)$, 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 3.1.)

Given these two observations, consider two choice problems 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).[40] To extend this notion to x and y with $x \nsim y$, note that no choice of act is required for any $\ell \in L$, and thus P_x must also be optimal for $(1-t)x + t\ell$. Therefore, if y is concordant with $(1-t)x + t\ell$, we will say that it is also concordant with x.

Definition C.1. For $\lambda \in (0,1)$, choice problems x and y are λ-*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 choice problems x and y are *concordant* if $(1-t)x + t\ell$ and y are λ-concordant for some $t \in [0,1)$, $\ell \in L$, and $\lambda \in (0,1)$.

Axiom 5 (Concordant Independence)**.** If x and y are λ-concordant, so are x and $\frac{1}{2}x + \frac{1}{2}y$. Furthermore, if $X' \subset X$ consists of concordant choice problems, then $\succsim|_{X'}$ satisfies Independence.[41]

(40) 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.

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

C.5. Self-Generation

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.[42] *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:

$$\succsim_{(P,s)} \text{ satisfies Axioms 1–5 and Self-Generation.}$$

It is important to note the self-referential character of Self-Generation, which is the only axiom that relies on the recursive structure of X (apart from L-Stationarity (Axiom 2(c)), which relies on the recursive structure of L); it requires 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.

Clearly, $\succsim_{(P,s)}$ must be inferred from the initial ranking of choice problems, 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 choice problem 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 C.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 $\left[x \oplus_{(\varepsilon,s)} y\right] \succsim \left[x \oplus_{(\varepsilon,s)} y'\right]$.[43]

We verify in Appendix D.3 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.

(42) 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.

(43) 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 $f \in x$.

Axiom 6 (Self-Generation)**.** If $\succsim$ is such that $\succsim_{(x,s)}$ (induced by $\succsim$ as in Definition C.2) is complete on $X' \subset X$, then $\succsim_{(x,s)}$ satisfies Axioms 1–6 on X'.[44]

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.

C.6. Invariant Per-Period Constraint and Fixed Arrival of Information

By Theorem 3, $\succsim$ satisfies Axioms 1–6 if, and only if, it has an ICP representation $\big((u_s), \delta, \Pi, \mathcal{M}\big)$. We now discuss two special cases of the ICP 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 $y \in X$.

For a finite $x \in X$ and for a mapping $e : x \to (0, 1]$, let $\big(x \oplus_{(e,s)} y\big) := \big\{f \oplus_{(e(f),s)} y : f \in x\big\}$, which perturbs the continuation lottery in state s for any act f in x by giving weight $e(f)$ to $\big(c_s^-, y\big)$.

Axiom 7 (Stationary Maximal Choice Problem)**.** $x \in X$ 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 choice problems 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 C.3. The ICP $\mathcal{M} = (\Theta, \theta, \Gamma, \tau)$ is an *invariant per-period constraint* if $\Gamma(\theta)$ is constant on Θ (or, equivalently, if Θ is a singleton).

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

Proposition C.4. If $\succsim$ has ICP-representation $\big((u_s), \delta, \Pi, \mathcal{M}\big)$, 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 ℓ^* 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 3.1 that the ICP $\mathcal{M} = (\Theta, \theta_0, \Gamma, \tau)$ is indistinguishable from the ICP $(\Theta, \tau(\theta_0, P, s), \Gamma, \tau)$ 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.[45]

(44) As noted in Section 4.2, an equivalent formulation of Axiom 6 is that $\succsim$ must belong to the recursively defined set Ψ^*, where $\Psi^* := \{\succsim \text{ on } X : \text{(i) } \succsim \text{ satisfies Axioms 1–5, and (ii) } \succsim_{(P,s)} \in \Psi^*\}$.

(45) 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.

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 C.5. The ICP $\mathscr{M} = (\Theta, \theta, \Gamma, \tau)$ captures *fixed arrival of information* if $\Gamma(\theta)$ is a singleton for all $\theta \in \Theta$.

Proposition C.6. If $\succsim$ has ICP-representation $\big((u_s), \delta, \Pi, \mathscr{M}\big)$, then it satisfies Axiom 8 if, and only if, $\mathscr{M}$ 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 θ. 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 C.7. At the end of Section 3.1 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 ICP, 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.

D. Existence

Theorem 2 in the Supplementary Appendix establishes that $\succsim$ satisfies Axioms 1–5 if, and only if, it has a representation of the form

[**D.1**]
$$V(x) = \max_{P \in \mathfrak{M}_p^\sharp} \sum_{J \in P} \left[\max_{f \in x} \sum_s \pi_0(s \mid J) \big[u_s(f_1(s)) + v_s(f_2(s), P)\big] \pi_0(J) \right]$$

where $\mathfrak{M}_p^\sharp$ 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 $P \in \mathfrak{M}_p^\sharp$, with the property that for all $P, P' \in \mathfrak{M}_p^\sharp$, $s \in S$, $v_s(\cdot, P)|_L = v_s(\cdot, P')|_L$.

For a fixed P in the representation in [**D**], let X'_P be defined as follows:

$$X'_P := \big\{x : V(x) = V\big(x, P, (v_s(\cdot, P))\big) \text{ for some } P \in \mathfrak{M}_p^\sharp \text{ and}$$
$$V(x) > V\big(x, Q, (v'_s(\cdot, Q))\big) \text{ for all } Q \in \mathfrak{M}_p^\sharp \text{ such that } P \neq Q\big\}$$

The following corollary records a useful property of the set X'_P.

Corollary D.1. Let $\succsim$ have a representation as in [**D**], let $x \in X'_P$ and let Y_x denote the set of choice problems that are (i) concordant with x, and (ii) have a unique optimal partition. Then, $Y_x = X'_P$.

The corollary follows immediately from the representation in [**D**]. Lemma 3.33 in the Supplementary Appendix derives the same conclusion from Axioms 1–5.

In the rest of this section, we impose Self-Generation (Axiom 6) and exhibit the existence of an ICP-representation. We begin with an analysis of preferences on L in Appendix D.1, the space of consumption streams. We then discuss self-generating representations and the concomitant dynamic plans in Appendix D.2.

D.1. 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 $s \in S$, $\delta \in (0,1)$, Π represent the transition operator for a fully connected Markov process on S, and π_0 be the unique invariant distribution of Π. 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

$$W(\ell; s) = \sum_{s' \in S} \Pi(s, s')\big[u_{s'}(\ell_1(s')) + \delta W(\ell_2(s'); s')\big]$$

and where u_s non-trivial for some $s \in S$. Then, W_0 can also be written as

$$W_0(\ell) = \sum_{s' \in S} \pi_0(s)\big[u_s(\ell_1(s)) + \delta W(\ell_2(s); s)\big]$$

because π_0 is the unique invariant distribution of Π 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 4 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 ℓ^* and ℓ_*. 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, ℓ^* is the (unique) consumption stream that delivers f^+ at each date and ℓ_* 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 3.3 in the Supplementary Appendix), so $(c, \ell^*) \succ_s (c, \ell_*)$.

D.2. 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 $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ is a tuple where

- $u_s \in \mathbf{C}(C)$ for all $s \in S$,
- $\mathbb{Q} \subset \mathscr{P}(S)$,
- $v_s(\cdot, P) \in \mathrm{BL}(X)$ for all $s \in S$ and $P \in \mathbb{Q}$,[46]
- $\pi \in \Delta(S)$,
- $u_s(c_s^\dagger) = v_s(\ell^\dagger, P) = 0$ 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

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

In that case, we say that the tuple $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ 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 $\mathbf{A} : 2^{\mathbf{C}(X)} \to 2^{\mathbf{C}(X)}$ as follows:

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

Proposition D.2. The operator $\mathbf{A}$ is well defined and has a largest fixed point $\Phi^* \neq \{\mathbf{0}\}$. Moreover, Φ^* 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 $\mathbf{0} \neq v_s(\cdot, P)\mathbf{0} \in \Phi$ for all $P \in \mathbb{Q}$, and any u_s, so that $\mathbf{A}\Phi \neq \varnothing$.) 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, Φ^* does not contain only the trivial function $\mathbf{0}$, fix $\mathbb{Q} = \big\{\{\{s\} : s \in S\}\big\}$ 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 π as the uniform distribution over S. Then V is implemented by $\big((u_s), \mathbb{Q}, \delta V), \pi\big)$, while δV is implemented by $\big((\delta u_s), \mathbb{Q}, \delta^2 V), \pi\big)$, 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 Φ^* is nonempty.

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

Finally, to see that Φ^* is a cone, let $v \in \Phi^*$ and suppose $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ implements v. Then, for all $\lambda \geq 0$, $\big((\lambda u_s), \mathbb{Q}, (\lambda v_s(\cdot, P)), \pi\big)$ implements λv, ie, $\lambda \Phi^*$ is also a fixed point of $\mathbf{A}$. Because Φ^* is the largest fixed point, it must be a cone. □

Notice that each $v \in \Phi^*$ is implemented by a tuple $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ with the property that each $v_s(\cdot, P) \in \Phi^*$. Adapting the terminology of Abreu, Pearce, and Stacchetti (1990), the set Φ^* 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 Φ^* follows relatively easily, as noted in the proof of Proposition D.2.) 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 D.3. There is a unique map $\varphi^* : \Phi^* \to \Omega$ that satisfies for some implementation $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ of v, that

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

and is independent of the implementation chosen.

Proof. Let $v^{(1)} \in \Phi_1$, and suppose $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$ 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 π will typically not be unique.) Then, define $\varphi_1 : \Phi_1 \to \Omega_1$ as

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

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

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

Notice that the same argument that established the uniqueness of φ_1 also applies here, to provide the uniqueness of φ_n.

Now, suppose $v \in \Phi^*$. This implies v has a partitional and separable implementation $\big((u_s), \mathbb{Q}, (v_s(\cdot, P)), \pi\big)$, 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

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

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

To recapitulate, we can now extract an RIC from a self-generating representation. In other words, the identification of the RIC ω_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 $\boldsymbol{\pi}$ over S^∞. 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

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

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 $\sigma_i^{(n)} : \mathfrak{H}_n \to \mathscr{P} \times \Omega^S$. Similarly, a *dynamic consumption plan* is a sequence $\sigma_c = (\sigma_c^{(1)}, \sigma_c^{(2)}, \dots)$ of mappings where $\sigma_c^{(n)} : \mathfrak{H}_n \to \mathscr{F}(\Delta(C \times X))$. A *dynamic plan* σ 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) $\sigma_i^{(n)}(\mathfrak{h}_n) \in \omega^{(n-1)}$, (ii) $\sigma_c^{(n)}(\mathfrak{h}_n) \in x^{(n-1)}$, and (iii) given the information plan $\sigma_i^{(n)}(\mathfrak{h}_n) = (P, \boldsymbol{\omega}') \in \omega^{(n-1)}$, $\sigma_c^{(n-1)}(\mathfrak{h}_n)$ is P-measurable, ie, for all $I \in P$ and for all $s, s' \in I$, $\sigma_c^{(n)}(\mathfrak{h}_n)(s) = \sigma_c^{(n)}(\mathfrak{h})(s')$.

Each dynamic plan along with initial states (x, ω_0, π_0) induces a probability measure over $\big(X \times \Omega \times S\big)^\infty$ 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 choice problems, RICs, and objective states induced by a dynamic plan, where $x^{(n)} \in X$ is the choice problem 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 $\sigma^{(n)}(\mathfrak{h}_n)$ only depends on $(x^{(n-1)}, \omega^{(n-1)}, s^{(n-1)})$.[47]

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 σ, where $\omega^{(n)} = \varphi_n(V)$ as in Proposition D.3 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 Φ^*, 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 D.4. 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

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

Proof. Consider $V \in \Phi^*$ with Lipschitz rank λ. Recall that for any $x \in X$, $\ell^\dagger \diamond_n x \in X$ denotes the choice problem that delivers $\ell^\dagger$ in every period until period $n-1$ and then, in period n, in

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

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 $\left|V(\ell^\dagger \diamond_n x) - V(\ell^\dagger \diamond_n y)\right| < \varepsilon$.

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

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

for all $n \geq N$. Recall that

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

Moreover, we have

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

which completes the proof. □

Adapting the terminology of Dubins and Savage (1976), we shall say that a function $V \in \Phi^*$ is *equalizing* if [**D.4**] holds. (To be precise, if [**D.4**] holds, then every dynamic plan is *equalizing* in sense of Dubins and Savage (1976).)

Given an initial $(x, \omega) \in X \times \Omega$, each σ induces a probability measure over $\bigtimes_n \mathfrak{H}_n$, 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 σ given the choice problem 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 choice problem x.

Let Σ 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 choice problem x and the RIC ω. In what follows, $V(x, \sigma)$ is the expected utility from following the dynamic plan σ given the choice problem x.

Lemma D.5. 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 σ', an agent with the utility function V is indifferent between following σ' 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 [**D.4**] in Proposition D.4).[48]

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 36). We can now adapt Theorem 9.2 in Stokey, Lucas, and Prescott (1989) where, as above, their equation 7 is replaced by [**D.4**], to show that σ 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). □

D.3. Existence of a Self-Generating Representation

Recall that a representation $V : X \to \mathbb{R}$ of $\succsim$ is a *self-generating representation* if $V \in \Phi^*$ (see section D.2 for the definition of Φ^*). Starting from the representation in [**D**], we show in this section that imposing Self-Generation (Axiom 6) on $\succsim$ implies it has a self-generating representation.

Proposition D.6. 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 D.3.2. We first show that $\succsim_{(x,s)}$ from Definition C.2 is well defined. We begin with a preliminary lemma.

Lemma D.7. 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. □

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

Lemma D.8. Let $\succsim$ have a representation as in [**D**]. For all $P \in \mathfrak{M}^\sharp_p$, there exists a finite $x \in X'_P \cap X^*$.

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

Lemma D.9. Let $\succsim$ have a representation as in [**D**]. For all $P \in \mathfrak{M}^\sharp_p$, $v_s(y, P) \geq v_s(\ell_*, P)$.

(48) 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.

Proof. Suppose instead that $v_s(y, P) < v_s(\ell_*, P)$. Consider $x \in X'_P \cap X^*$ which exists by Lemma D.8. Then, for $\varepsilon > 0$ small enough such that P remains optimal, $\left[x \oplus_{\varepsilon,s} \ell_*\right] \succ \left[x \oplus_{\varepsilon,s} y\right]$. 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

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

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

Lemma D.10. Let $\succsim$ have a representation as in [**D**]. 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 C.2 applies. In particular, $\succsim_{(x,s)}$ is represented by $v_s(\cdot, P)$ on 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 [**D**] 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 Corollary D.1). By Lemma D.7, $[x \oplus_{(\varepsilon,s)} y], [x \oplus_{(\varepsilon,s)} y'] \in B(x; \varepsilon)$. Then, $\left[x \oplus_{(\varepsilon,s)} y\right] \succsim \left[x \oplus_{(\varepsilon,s)} y'\right]$ if, and only if, $V\left(x \oplus_{(\varepsilon,s)} y\right) \geq V\left(x \oplus_{(\varepsilon,s)} y'\right)$. Suppose $f \oplus_{\varepsilon,s} y$ is optimally chosen from $x \oplus_{\varepsilon,s} y$ in the state s. Then, it must be that

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

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 $\left[x \oplus_{(\varepsilon,s)} y\right] \succsim \left[x \oplus_{(\varepsilon,s)} y'\right]$. 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 Corollary D.1 $x' \in X'_P$. 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. □

Lemma D.11. The binary relation $\succsim_{(P,s)}$ on X which is represented by $v_s(\cdot, P)$ satisfies Axioms 1–6.

Proof. By Lemma D.10, $\succsim_{(P,s)} = \succsim_{(x,s)}$ for some $x \in X'_P$. By Self-Generation (Axiom 6), $\succsim_{(x,s)}$ satisfies Axioms 1–6 on X. □

Before we prove Proposition D.6, an interlude.

D.3.1. *Some Properties of Consumption Streams*

We now relate preferences on L to those on X.

Let $\tilde{X}_1 := \mathcal{K}\left(\mathcal{F}(\Delta(C \times \{\ell_*\}))\right)$ be the space of one-period problems that always give ℓ_* at the beginning the second period. Inductively define $\tilde{X}_{n+1} := \mathcal{K}\left(\mathcal{F}(\Delta(C \times \tilde{X}_n))\right)$ 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 D.12. The set $\tilde{X} \subset X$ is dense in X.

Proof. Recall that X is the space of all *consistent* sequences in $\bigtimes_{n=1}^{\infty} X_n$, where $X_1 := \mathcal{K}\big(\mathcal{F}(\Delta(C))\big)$ and $X_{n+1} := \mathcal{K}\big(\mathcal{F}(\Delta(C \times X_n))\big)$. As noted in the construction in Appendix A.2, every $x \in X$ is a sequence of the form $x = (x_1, x_2, \ldots, x_n, \ldots)$ where $x_n \in X_n$, and the metric on X is the product metric.

For any $x = (x_1, x_2, \ldots) \in X$ and $n \geq 1$ set $\tilde{x}_n \in \tilde{X}_n$ to be x_n concatenated with ℓ_*. 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. □

Lemma D.13. Let $\succsim$ satisfy Axioms 1–5. Then, for any $s \in S$ and $P \in \mathfrak{M}_P^\sharp$, $\ell^* \succsim_{(P,s)} \ell \succsim_{(P,s)} \ell_*$ for all $\ell \in L$.

Proof. The preference $\succsim$ has a separable and partitional representation as in [**D**]. 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 D.1, $\succsim_s$ 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 $Q \in \mathfrak{M}_P^\sharp$ in the representation [**D**]. But $v_s(\cdot, P)$ represents $\succsim_{(P,s)}$ which implies that $\ell^* \succsim_{(P,s)} \ell \succsim_{(P,s)} \ell_*$ for all $\ell \in L$, $s \in S$. □

Proposition D.14. Let $\succsim$ satisfy Axioms 1–6. Then, for all $x \in X$, $\ell^* \succsim x$.

Proof. By the continuity of $\succsim$ and by Lemma D.12, 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 [**D**] that $V(\tilde{x}_1) \leq V(\ell^*)$ for all $\tilde{x}_1 \in \tilde{X}_1$. Notice that the representation in [**D**] 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 $P \in \mathfrak{M}_P^\sharp$, 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 D.13, 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 D.13 to establish that ℓ^* 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,

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

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

D.3.2. *Proof of Proposition D.6*

Proof. To see that (b) implies (a), suppose $\succsim$ has the representation [**D**]. By Proposition 3.35 of the Supplementary Appendix $\succsim$ satisfies Axioms 1–5. All that remains to establish is that $\succsim$ also satisfies Axiom 6.

Given a representation as in [**D**] that is *also* self-generating, let $x \in X$ be finite and $P \in \mathfrak{M}_p^\sharp$ be an optimal partition for x. Observe first that if $x \in X'_P$, then by Lemma D.10 $\succsim_{(x,s)}$ 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 $X' \subset X$ 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.

We assume, without loss of generality,[49] that $[x \oplus_{(\varepsilon,s)} y] \sim [x \oplus_{(\varepsilon,s)} y']$. To build intuition, suppose the only optimal partitions for x are $P, Q \in \mathfrak{M}_p^\sharp$ 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

$$[(1-t)\, x \oplus_{(\varepsilon,s)} y + t x_1\,(P)] \succ [(1-t)\, x \oplus_{(\varepsilon,s)} y' + t x_1(P)]$$

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 for x. Thus, if $\succsim_{(x,s)}$ is complete on X', then there is P in $\mathfrak{M}_p^\sharp$ 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 $P \in \mathfrak{M}_p^\sharp$. 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 D.11 has two implications. First, $\succsim_{(P,s)}$ has a separable and partitional representation $v'_s(\cdot, P)$ as in [**D**]. 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 D.14. Together with Lemma D.13 and IICC (Axiom 4), this implies that $\ell^* \succsim_{(P,s)} y \succsim_{(P,s)} \ell_*$ 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 [**D**], that is, it can be written as

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

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 [**D**], and so on, *ad infinitum*, which demonstrates that $V \in \Phi^*$. □

(49) For a detailed justification of this assumption, see the proof of Lemma 3.33 in the Supplementary Appendix.

D.4. Recursive Representation

We now establish a recursive representation for $\succsim$, thereby proving Theorem 3.

Recall from Appendix D.1 that $\succsim|_L$ has a standard RAA representation $\big((u_s), \delta, \Pi\big)$. That is, there exist functions $V_L^*(\cdot, s) : L \to \mathbb{R}$ such that $V_L^*(\ell, \pi_0) := \sum_s \pi_0(s) V_L^*(\ell, s)$ represents $\succsim|_L$, and

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

where $u_s\big(c_s^\dagger\big) = 0$ 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 $\big((u_s)_{s\in S}, \delta, \Pi\big)$.

By Proposition D.6 $\succsim$ 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 ω_0 and observe that by Proposition 2.2, the tuple $\big((u_s)_{s\in S}, \Pi, \delta, \omega_0\big)$ 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 3.

Proposition D.15. 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.[50]

Proof. In this proof, we frequently refer to objects defined in Appendix D.2. 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 D.5, 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 D.5 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. □

Suppose V represents $\succsim$ and $V \in \Phi^*$. Then, there exists an implementation of V, given by $\big((u_s), \mathbb{Q}, (v_s^{(1)}(\cdot, P)), \pi\big)$. 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

(50) It follows immediately from Proposition D.15 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).

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.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1990. “Toward a Theory of Discounted Repeated Games with Imperfect Monitoring”. *Econometrica* 58 (5): 1041–1063.

Bertsekas, Dimitri P., and Steven Shreve. 2000. *Stochastic Optimal Control: The Discrete-Time Case*. Belmont, MA: Athena Scientific.

Blackwell, David. 1953. “Equivalent Comparison of Experiments”. *Annals of Mathematical Statistics* 24:265–272.

Blanchard, Olivier Jean, and Stanley Fischer. 1989. *Lectures on Macroeconomics*. Cambridge, Ma: MIT Press.

Caplin, Andrew, and Mark Dean. 2015. “Revealed Preference, Rational Inattention, and Costly Information Acquisition”. *American Economic Review* 105 (7): 2183–2203.

De Oliveira, Henrique. 2016. *Blackwell’s Informativeness Theorem using Category Theory*. Tech. rep. Pennsylvania State University.

De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and M. Kemal Ozbek. 2016. “Rationally Inattentive Preferences and Hidden Information Costs”. *Theoretical Economics*: forthcoming.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 2001. “Representing Preferences with a Unique Subjective State Space”. *Econometrica* 69 (4): 891–934.

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.

Dillenberger, David, R. Vijay Krishna, and Philipp Sadowski. 2017. *Supplement to ‘Subjective Information Choice Processes’*. Tech. rep. Duke University.

Dillenberger, David, Juan S. Lleras, Philipp Sadowski, and Norio Takeoka. 2014. “A Theory of Subjective Learning”. *Journal of Economic Theory* 153:287–312.

Dubins, Lester E., and Leonard J. Savage. 1976. *Inequalities for Stochastic Processes: How to Gamble if You Must*. Mineola, New York: Dover Publications Inc.

Dudley, Richard Mansfield. 2002. *Real Analysis and Probability*. 2nd edition. Cambridge University Press.

Ellis, Andrew. 2016. *Foundations for Optimal Inattention*. Tech. rep. London School of Economics.

Ergin, Haluk, and Todd Sarver. 2010. “A Unique Costly Contemplation Representation”. *Econometrica* 78 (4): 1285–1339.

Gilboa, Itzhak, and David Schmeidler. 1989. “Maxmin Expected Utility with a Non-Unique Prior”. *Journal of Mathematical Economics* 18:141–153.

Gul, Faruk, and Wolfgang Pesendorfer. 2004. “Self-Control and the Theory of Consumption”. *Econometrica* 72 (1): 119–158.

Hébert, Benjamin, and Michael Woodford. 2016. *Rational Inattention with Sequential Information Sampling*. Tech. rep. Stanford University GSB.

Herstein, Israel Nathan, and John Milnor. 1953. “An Axiomatic Approach to Measurable Utility”. *Econometrica* 21 (2): 291–297.

Higashi, Youichiro, Kazuya Hyogo, and Norio Takeoka. 2009. “Subjective random discounting and intertemporal choice”. *Journal of Economic Theory* 144 (3): 1015 –1053.

Karatzas, Ioannis, and William Sudderth. 2010. “Two Characterizations of Optimality in Dynamic Programming”. *Applied Mathematics and Optimization* 61:421–434.

Kreps, David M. 1979. “A Representation Theorem for ‘Preference for Flexibility’”. *Econometrica* 47 (3): 565–577.

– . 2012. *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton, NJ: Princeton University Press.

Kreps, David M., and Evan L Porteus. 1978. “Temporal Resolution of Uncertainty and Dynamic Choice Theory”. *Econometrica* 46 (1): 185–200.

Krishna, R. Vijay, and Philipp Sadowski. 2014. “Dynamic Preference for Flexibility”. *Econometrica* 82 (2): 655–703.

Laffont, Jean-Jacques. 1989. *The Economics of Uncertainty and Information*. The MIT Press.

Luenberger, David G. 1969. *Optimization by Vector Space Methods*. John Wiley and Sons.

Maćkowiak, Bartosz, and Mirko Wiederholdt. 2009. “Optimal Sticky Prices under Rational Inattention”. *American Economic Review* 99 (3): 769–803.

Mariotti, Thomas, Martin Meier, and Michele Piccione. 2005. “Hierarchies of Beliefs for Compact Possibility Models”. *Journal of Mathematical Economics* 41 (3): 303–324.

Markowitz, Harry M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York, NY: Wiley.

Massa, Massimo, and Andrei Simonov. 2006. “Hedging, Familiarity and Portfolio Choice”. *Review of Financial Studies* 19 (2): 633–685.

Matejka, Filip, and Alisdair McKay. 2014. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model". *American Economic Review* forthcoming.

Mullainathan, Sendhil, and Eldar Shafir. 2013. *Scarcity*. Times Books.

Myatt, David P., and Christopher Wallace. 2012. "Endogenous Information Acquisition in Coordination Games". *Review of Economic Studies* 79:340–374.

Orkin, Michael. 1974. "On Stationary Policies—The General Case". *The Annals of Statistics* 2 (1): 219–222.

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.

Persico, Nicola. 2000. "Information Acquisition in Auctions". *Econometrica* 68 (1): 135–148.

Piermont, Evan, Norio Takeoka, and Roee Teper. 2016. "Learning the Krepsian State: Exploration Through Consumption". *Games and Economic Behavior* 100 (69–94).

Rust, John. 1994. "Structural Estimation of Markov Decision Processes". In *Handbook of Econometrics*, ed. by Robert F. Engle and Daniel F. McFadden, vol. IV. Elsevier.

Sims, Christopher A. 2003. "Implications of Rational Inattention". *Journal of Monetary Economics* 50 (3): 665–690.

– . 1998. "Stickiness". *Carnegie-Rochester Conference Series on Public Policy* 49:317–356.

Steiner, Jakub, Colin Stewart, and Filip Matějka. 2015. *Rational Inattention Dynamics: Inertia and Delay in Decision-Making*. Tech. rep. CERGE-EI.

Stigler, George. 1961. "The Economics of Information". *Journal of Political Economy* 69 (3): 213–225.

Stokey, Nancy L., Robert E. Lucas Jr, and Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics*. Cambridge, Ma: Harvard University Press.

Woodford, Michael. 2012. *Inattentive Valuation and Reference-Dependent Choice*. Tech. rep. Columbia University.