Dynamic Rationally Inattentive Discrete Choice: A Posterior-Based Approach*

Jianjun Miao† Hao Xing‡

June 7, 2019

Abstract

We adopt the posterior-based approach to study dynamic discrete choice problems with rational inattention. We show that the optimal solution for the Shannon entropy case is characterized by a system of equations that resembles the dynamic logit rule. We propose an efficient algorithm to solve this system and apply our model to explain phenomena such as status quo bias, confirmation bias, and belief polarization. We also study the dynamics of consideration sets. Unlike the choice-based approach, our approach applies to general uniformly posterior-separable information cost functions. A key condition for our approach to work in dynamic models is the convexity of the difference between the discounted generalized entropy of the prior beliefs about the future states and the generalized entropy of the current posterior.

Keywords: Rational Inattention, Endogenous Information Acquisition, Entropy, Dynamic Discrete Choice, Dynamic Programming

JEL Classifications: D11, D81, D83.

*We thank John Leahy, Jakub Steiner, Colin Stewart, and Mike Woodford for helpful comments.

†Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. Email: miaoj@bu.edu. Tel: (617) 3536675

‡Department of Finance, Questrom School of Business, Boston University, 595 Commonwealth Ave, Boston MA 02215. Email: haoxing@bu.edu.

1 Introduction

Economic agents often make dynamic discrete choices, such as whether to stay at home or take a job and which job to take, when to replace a car and which new car to buy, when to invest in a project and which project to invest, and so on. When making these decisions people often face imperfect information about payoffs. People must choose what information to acquire and when to acquire it given their limited attention to the available information.

We adopt the rational inattention (RI) framework introduced by Sims (1998, 2003) to study the optimal information acquisition and choice behavior in a dynamic discrete choice model. In the model a decision maker (DM) can choose a signal about a payoff-relevant state of the world before taking an action in each period. The state follows a finite Markov chain with a transition kernel depending on the current states and actions. The DM receives flow utilities, that depend on the current states and chosen actions, and pays a utility cost to acquire information, that is proportional to the reduction in the uncertainty measured by the entropy of his beliefs. The DM's objective is to maximize the expected discounted utility less the cost of the information he acquires. We call this problem the dynamic RI problem. While we focus our analysis on the case with the Shannon entropy (Shannon (1948)), our approach applies to general uniformly posterior-separable (UPS) information cost functions introduced by Caplin and Dean (2013) (henceforth CD) and Caplin, Dean, and Leahy (2018b) (henceforth CDL)).

We make three contributions to the literature. First, we characterize the solution to the dynamic RI problem using the posterior-based approach. We find that the optimal choice rule in the case of the Shannon entropy cost is consistent with the dynamic logic behavior (Rust (1987)) with respect to payoffs that differ from the DM's true payoffs by an endogenous additive term. Following Steiner, Stewart, and Matějka (2017) (henceforth SSM), we call this term a default rule or a predisposition, which depends on the history of actions but does not depend on the history of states. Relative to the dynamic logit behavior with the DM's true payoffs, the default rule increases the relative payoffs associated with actions that are chosen with a high probability on average across all states at a given history.

Solving the dynamic RI problem is difficult because the current information acquisition affects future beliefs, which in turn influence the continuation value in a nonlinear way. We prove that the continuation value is actually convex in the revised prior beliefs following any history (reached with positive probabilities). By dynamic programming, the current choice and the continuation value are linked by the Bellman equation. It is unclear whether this dynamic programming problem is concave. SSM (2017) argue that the solution can be obtained by ignoring the effect of information acquisition on future beliefs. They also argue that one can treat the continuation value as fixed when optimizing at each history and apply the static RI solution of CD (2013) and Matějka and

McKay (2015) (henceforth MM) sequentially. The key step behind their approach is to transform the dynamic RI problem into a control problem by choosing a default rule and a state-dependent choice rule separately. Applying first-order conditions with respect to these variables separately yields the dynamic logit solution.

There is an implicit concavity assumption behind the choice-based approach of SSM (2017). Their optimality conditions are derived from the coordinate-wise first-order conditions, which are necessary, but may not be sufficient if the optimization problem is not jointly concave. SSM (2017) suggest to use the sufficient conditions in CD (2013) and CDL (2018a). However, these conditions are invalid if the optimization problem is not jointly concave once the impact of the current choice on the future beliefs is taken into account. We give a counterexample to illustrate the issue of the nonconcavity and show that the first-order conditions together with the CD and CDL type conditions can lead to nonoptimal solutions.

Our posterior-based approach is built on the insights of CD (2013) in a static model and takes into account the issue of joint concavity in a dynamic setting. We derive the posterior-based Bellman equation using the predictive distribution as the state variable. This distribution given any history can be viewed as the prior belief about the future states at that history. It is revised from the current posterior through the state transition kernel. For the posterior-based approach to work, we need the net utility function to be concave in the current posterior. In our dynamic model the net utility function consists of the current net utility and the continuation value. As discussed earlier the continuation value is convex in the predictive distribution (prior beliefs) and hence the current posterior. On the other hand, the current net utility is concave in the current posterior. We prove that the overall net utility is concave if and only if the difference between the discounted entropy of the prior belief about the future states and the entropy of the current posterior given the same history is convex. This assumption is satisfied if the discount factor is between zero and one for the Shannon entropy cost.

The same intuition applies to the choice-based approach discussed earlier. The current choice rule affects the current posterior and hence the predictive distribution of the future states. Thus the convexity of the continuation value may dominate the concavity of the current payoffs in the choice rule. Without checking concavity, the first-order Kuhn-Tucker conditions can lead to a nonoptimal solution.

Our second contribution is to propose a Markovian characterization of the optimal RI solution and an efficient algorithm to solve a Markovian solution. For a Markovian solution, the predictive distribution of the next-period states depends only on the current action, the default rule depends only on the last period action, and the choice rule depends only on the current state and the last period action. This Markovian property allows us to characterize the dynamic logit solution using a computable system of nonlinear difference equations.

Our algorithm extends the forward-backward Arimoto-Blahut algorithm of Tanaka, Sandberg, and Skoglund (2018) (henceforth TSS) to infinite-horizon models with discounting. This algorithm is based on the Arimoto-Blahut algorithm for solving static channel capacity and rate distortion problems in information theory in the engineering literature (Arimoto (1972) and Blahut (1972)). It is a block coordinate descent algorithm applied to a special class of objective functions (Bertsekas (2016)). A sufficient condition for convergence is that the objective function is concave. We show that this condition is satisfied in our model so that the forward-backward Arimoto-Blahut algorithm converges to the optimal solution to the dynamic RI problem.

Our third contribution is to apply our theoretical results and numerical methods to solve some economic examples based on a matching state problem often studied in the literature (e.g., CD (2013), SSM (2017), and CDL (2018a)). We show that RI can help explain some phenomena documented in the psychology literature, such as status quo bias, confirmation bias, and belief polarization. We find that the status quo bias discussed by SSM does not arise when the decision horizon is sufficiently long. The reason is that the probability of switching states in the future is getting larger if the horizon is longer. Thus the DM has incentives to acquire new information and take a different action. We also show that there is a positive feedback between beliefs and actions when the state transition kernel depends on actions. This property is useful to understand the preceding behavioral biases. We also apply our algorithm to solve a medium-scale dynamic model with 10 states and 10 actions and analyze the dynamics of consideration sets.

As discussed earlier, our paper is closely related to CD (2013), MM (2015), SSM (2017), and CDL (2018a,b). SSM (2017) is the first paper that extends the static model of MM (2015) to a dynamic setting and derives the dynamic logit rule.[1] SSM apply the locally invariant posterior (LIP) property in CD (2013) to show that interior solutions are Markovian. Our characterization allows for corner solutions and the proof is different from theirs. Our paper also fills a gap in their analysis that the optimization objective must be jointly concave. More importantly, we adopt the posterior-based approach, which delivers more structures for our dynamic analysis and applies to a large class of UPS information cost functions. By contrast, the choice-based approach of SSM typically fails for cost functions not based on the Shannon entropy. Complementing SSM (2017), we derive a more general system of optimality conditions that facilitates the efficient forward-backward Arimoto-Blahut algorithm to solve large-scale models numerically. We also extend the SSM model to allow the state transition kernel to depend on actions. This generalization permits us to study a wide range of economic and psychological behavior.

Our paper is also related to Hébert and Woodford (2018) and Zhong (2019), who adopt the posterior-based approach to study optimal stopping problems under RI with general information

[1] See Mattsson and Weibull (2002) and Fudenberg and Strzalecki (2015) for related models.

cost functions in the continuous-time setup.[2] Unlike their papers, ours is the first to study optimal control problems, for which the concavity of the objective function in dynamic models is important for the optimality of the first-order conditions. We show that such concavity is determined by the convexity of the difference between the discounted generalized entropy of the prior beliefs about the future states and the generalized entropy of the current posterior.

Most existing work on RI has focused on models with a continuous choice set, which are typically set up in the linear-quadratic-Gaussian framework (e.g., Peng and Xiong (2006), Luo (2008), Maćkowiak and Wiederholt (2009), Mondria (2010), Van Nieuwerburgh and Veldkamp (2010), Miao (2019), and Miao, Wu, and Young (2019)). Woodford (2009) is the first paper that studies a dynamic binary choice problem under RI (the problem of a firm that decides each period whether to reconsider its price). Jung et al (2018) show that rationally inattentive agents can constrain themselves voluntarily to a discrete choice set even when the initial choice set is continuous. See Sims (2011) and Maćkowiak, Matějka and Wiederholt (2018) for surveys and references cited therein.

2 Model

2.1 Setup

Consider a T-period decision problem with $T \leq \infty$ and time is denoted by $t = 1, 2, ..., T$. Uncertainty is represented by a discrete finite state space $X \equiv \{1, 2, ..., m\}$ and a prior distribution $\mu_1 \in \Delta(X)$, where we use $\Delta(Z)$ to denote the set of distributions on any finite set Z. The decision maker (DM) makes choices from a finite action set denoted by A satisfying $|X| \geq 2$ and $|A| \geq 2$. We can allow the action set A to depend on the current state as in the literature on Markov decision processes (Rust (1994) and Puterman (2005)), without affecting our key results but complicating notation. The state transition kernel is given by $\pi(x_{t+1}|x_t, a_t) \in \Delta(X|X \times A)$, where $\Delta(X|X \times A)$ denotes the set of all conditional distributions for the state $x_{t+1} \in X$ given the state $x_t \in X$ and the action $a_t \in A$ for $t \geq 1$.[3] SSM (2017) show that one can redefine the state space so that the state transition kernel is independent of the action. We allow such dependence so that our model is more flexible in applications and is also consistent with the literature on Markov decision processes (Rust (1994) and Puterman (2005)).

The DM receives flow utilities that depend on the current states and actions only. The period utility function is given by a bounded function $u : X \times A \rightarrow \mathbb{R}$. For the finite-horizon case with $T < \infty$, we allow u to be time dependent and include a terminal utility function $u_{T+1} : X \rightarrow \mathbb{R}$. SSM (2017) allow u to depend on the entire history of states and actions, which may generate

[2] The posterior-based approach is often applied in the Bayesian persuasion literature. See Kamenica (2018) for a survey and the references cited therein.

[3] As convention we define a conditional probability $P(C|B) = P(C \cap B)/P(B)$ whenever $P(B) > 0$; otherwise, set $P(C|B) = 0$ until further notice.

history-dependent solutions.

Prior to choosing an action in any period t, the DM can acquire costly information about the history of the state x^t, where we use x^t to denote the history $\{x_1, x_2, ..., x_t\}$. More accurate information will lead to better choices, but are more costly, with entropy-based costs to be discussed later. As SSM (2017) show, we do not need to model the endogenous choice of the information structure separately. Instead we can reformulate the problem in which the DM makes stochastic choices and signals correspond to actions directly. As CD (2013) argue, we can also identify signals with the corresponding posteriors. Thus we will focus on the model with stochastic choices (see Lemma 6 in Appendix E for the UPS information cost).

Define a (state-dependent) choice rule $\mathbf{p}$ as a sequence of conditional probability distributions

$$\mathbf{p} = \left\{p_t\left(a_t|x^t, a^{t-1}\right) \in \Delta\left(A|X^t \times A^{t-1}\right) : \text{all } \left(x^t, a^{t-1}\right),\ 1 \le t \le T\right\}.$$

The joint distribution of the state and action trajectories is denoted by $\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}$, which is uniquely determined by the initial state distribution $\mu_1 \in \Delta(X)$, the state transition kernel $\pi(x_{t+1}|x_t, a_t)$, and the choice rule $\mathbf{p}$ by a recursive formula

$$\mu_{t+1}\left(x^{t+1}, a^t\right) = \pi(x_{t+1}|x_t, a_t)\, p_t\left(a_t|x^t, a^{t-1}\right) \mu_t\left(x^t, a^{t-1}\right) \tag{1}$$

for any $t \ge 1$. Set $a_0 = \varnothing$ so that $p_1\left(a_1|x^1, a^0\right) = p_1(a_1|x_1)$ and $\mu_1\left(x^1, a^0\right) = \mu_1(x_1)$.

The choice rule $\mathbf{p}$ generates expected discounted utility

$$J\left(x^{T+1}, \mathbf{p}\right) = \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} u(x_t, a_t) + \beta^T u_{T+1}(x_{T+1})\right],$$

where $\beta \in (0,1)$ denotes the discount factor and the expectation is taken with respect to the joint distribution μ_{T+1} for $\left(x^{T+1}, a^T\right)$.

2.2 Information Cost

We model the information cost using the Shannon entropy-based directed information in information theory (Massey (1990)). To handle the infinite-horizon case, we incorporate discounting.

Definition 1 *For a discount factor $\beta \in (0,1)$, the discounted directed information from the stochastic process X^T to the stochastic process Y^T is defined as*

$$I_\beta\left(X^T \to Y^T\right) = \sum_{t=1}^{T} \beta^{t-1} I\left(X^t; Y_t|Y^{t-1}\right),$$

where $I\left(X^t; Y_t|Y^{t-1}\right)$ denotes the conditional mutual information defined as

$$I\left(X^t; Y_t|Y^{t-1}\right) = \mathbb{E}\left[\ln \frac{Q_t\left(X^t, Y_t|Y^{t-1}\right)}{Q_t\left(X^t|Y^{t-1}\right) Q_t\left(Y_t|Y^{t-1}\right)}\right]. \tag{2}$$

Here the expectation is taken with respect to the joint distribution of X^t *and* Y^t *denoted by* Q_t*, and* $Q_t\left(X^t, Y_t|Y^{t-1}\right)$, $Q_t\left(X^t|Y^{t-1}\right)$, *and* $Q_t\left(Y_t|Y^{t-1}\right)$ *denote the induced conditional distributions for any* $1 \leq t \leq T$.

The directed information measures directional information flow from X^T to Y^T and is different from the mutual information, which quantifies the amount of information that can be obtained about a random variable by observing another. The mutual information says little about causal relationships, because it is symmetric. By contrast, the directed information is asymmetric and is a measure of the predictive information transfer given observations. If $\beta = 1$, the discounted directed information reduces to the standard directed information $I\left(X^T \to Y^T\right)$ in information theory (Massey (1990)).

There are several equivalent definitions of the conditional mutual information. The definition in (2) is equivalent to

$$I\left(X^t; Y_t|Y^{t-1}\right) = H\left(X^t|Y^{t-1}\right) - H\left(X^t|Y^t\right) = H\left(Y_t|Y^{t-1}\right) - H\left(Y_t|X^t, Y^{t-1}\right),$$

where $H(\cdot|\cdot)$ denotes the conditional Shannon entropy which measures the amount of information about one variable given another variable.[4] The first equality in the above equation states that the conditional mutual information $I\left(X^t; Y_t|Y^{t-1}\right)$ measures the reduction of uncertainty about X^t after observing an additional data Y_t given the history of data Y^{t-1}. This expression is critical for the posterior-based approach. The interpretation for the second equality is similar, which is critical for the choice-based approach.

In our model we adopt the discounted directed information for the stochastic processes of states and actions, x^T and a^T :

$$I_\beta\left(x^T \to a^T; \mathbf{p}\right) = \sum_{t=1}^{T} \beta^{t-1} I\left(x^t; a_t|a^{t-1}\right), \tag{3}$$

where the joint distribution for x^t and a^t satisfies $\mu_t\left(x^t, a^t\right) = p_t\left(a_t|x^t, a^{t-1}\right)\mu_t\left(x^t, a^{t-1}\right)$. Since this distribution depends on the choice rule $\mathbf{p}$, we introduce an argument $\mathbf{p}$ to the discounted directed information in (3). SSM (2017) essentially adopt the same cost function.

Our approach also applies to a general class of information cost functions that satisfy the UPS property introduced by CD (2013) and CDL (2018b). By contrast, the choice-based approach of SSM (2017) typically does not work for cost functions other than the Shannon entropy. We will illustrate this point in Section 3.4 and Appendix E. In the main text we shall focus on the Shannon entropy-based cost function, which is consistent with much of the literature and allows us to compare with the literature more easily.

[4] See Cover and Thomas (2006) for a standard textbook reference for entropy, mutual information, and related notions.

2.3 Decision Problem

We now formulate the dynamic discrete choice problem under RI as follows:

Problem 1 *(dynamic RI problem)*

$$\max_{\mathbf{p}\in\mathbf{\Pi}} \quad J\left(x^{T+1},\mathbf{p}\right)-\lambda I_{\beta}\left(x^{T}\rightarrow a^{T};\mathbf{p}\right), \tag{4}$$

where $\Pi\equiv\prod_{t=1}^{T}\Delta\left(A|X^{t}\times A^{t-1}\right)$ *and* $\lambda>0$.

The second term in (4) measures the information cost corresponding to the information transfer from the state process x^T to the action process a^T. The parameter λ measures the marginal cost of information in utility units. When $\lambda = 0$, the problem is reduced to the standard Markov decision process formulation described in Puterman (2005) and Rust (1994). When $\lambda > 0$, there is a tradeoff between information acquisition and utility maximization. Acquiring more precise information about the state of the system helps the DM make a better choice. But this causes the control actions to be statistically more dependent on the state, which generates a larger information cost.

It is straightforward to show that there exists a solution to Problem 1 because the space Π is compact and the objective function is continuous on Π (see Proposition 1 in SSM (2017)).

3 Preliminaries and Basic Intuition

In this section we first present the solution in the static case analyzed by CD (2013), MM (2015), and CDL (2018a). We then study the two-period case and illustrate the difficulty of the dynamic model and our solution approach. Finally we discuss the extension to the general UPS cost functions.

3.1 Static Case

When $T = 1$ and $u_{T+1} = 0$, we obtain the following static problem analyzed by MM (2015), CD (2013), and CDL (2018a).

Problem 2 *(choice-based static RI problem) Choose* $p\in\Delta\left(A|X\right)$ *to solve*

$$V\left(\mu\right)\equiv\max \quad \mathbb{E}\left[u\left(x,a\right)\right]-\lambda I\left(x;a\right)=\sum_{x,a}p\left(a|x\right)\mu\left(x\right)\left[u\left(x,a\right)-\lambda\ln\frac{p\left(a|x\right)}{q\left(a\right)}\right], \tag{5}$$

subject to

$$q\left(a\right)=\sum_{x}p\left(a|x\right)\mu\left(x\right),\ \ a\in A, \tag{6}$$

where the prior distribution $\mu\equiv\{\mu\left(x\right)\}$ *is given.*

Following SSM (2017) we call q a default rule. CD (2013), MM (2015), and CDL (2018a) establish the following result:

Proposition 1 *The choice rule* $\{p(a|x)\}$ *and the default rule* $\{q(a)\}$ *are an optimal solution to Problem 2 if and only if they satisfy conditions (6),*

$$p(a|x) = \frac{q(a)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)}, \tag{7}$$

if $\mu(x) > 0$, *and*

$$\sum_x \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \leq 1, \tag{8}$$

with equality if $q(a) > 0$. *The optimal posterior satisfies*

$$\mu(x|a) = \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \text{ if } q(a) > 0.$$

This result appears in Gallager (1968, Theorem 9.4.1) and Berger (1971, Theorem 2.5.2).[5] For a better understanding of our analysis of the dynamic model, we present the main arguments of the proof according to the choice-based approach. We first define a function

$$F(p,q) \equiv \sum_{x,a} \mu(x) p(a|x) \left(u(x,a) - \lambda \ln \frac{p(a|x)}{q(a)} \right),$$

and verify that $F(p,q)$ is jointly concave in (p,q). Blahut (1972, Theorem 4) establishes the following result:

Lemma 1 *Let* $p \in \Delta(A|X)$ *be fixed. Then* $\max_{q\in\Delta(A)} F(p,q)$ *is a concave optimization problem and the optimal solution is given by* $q(a) = \sum_x \mu(x) p(a|x)$.

This lemma implies that the static RI problem is equivalent to the following optimization problem:

$$\max_{p\in\Delta(A|X),\ q\in\Delta(A)} F(p,q). \tag{9}$$

For fixed q and for any x with $\mu(x) > 0$, optimization over $p(a|x)$ gives the expression in (7), which is a function of q. We write it as $p^*(q)$ and then obtain

$$F(p^*(q), q) = \sum_x \mu(x) \lambda \ln \sum_a q(a) \exp(u(x,a)/\lambda), \tag{10}$$

which is a concave function of q. Now use the necessary and sufficient Kuhn-Tucker condition for $q \in \Delta(A)$ to derive (8).

From the derivation above we deduce that equations (6) and (7) are coordinate-wise first-order necessary conditions for the problem in (9). These two equations together with (8) for the problem

[5] See Denti, Marinacci, and Montrucchio (2019) for the general continuous states and actions case.

in (10) are necessary and sufficient conditions for optimality. Notice that condition (6) is implied by conditions (7) and (8) and all these first-order conditions are sufficient for optimality if the function $F(p,q)$ is jointly concave. Such concavity is ensured in the static RI problem.

When the sets X and A are large, the computation of the optimal stochastic choice is complicated. Arimoto (1972) and Blahut (1972) propose the following efficient algorithm:

1. Start with a guess $q^{(0)}(a) > 0$ for all a or $p^{(0)}(a|x) > 0$ for all (x,a).

2. Compute
$$q^{(k)}(a) = \sum_x \mu(x) p^{(k-1)}(a|x),$$
$$p^{(k)}(a|x) = \frac{q^{(k)}(a) \exp(u(x,a)/\lambda)}{\sum_{a'} q^{(k)}(a') \exp(u(x,a')/\lambda)}.$$

3. Iterate on $k \geq 1$ until convergence.

Since $F(p,q)$ is jointly concave in p and q, every limit point of the sequence generated by the Arimoto-Blahut algorithm is a global maximizer of F. To understand the intuition for this algorithm and (8), we combine the above two equations to derive

$$q^{(k+1)}(a) = \left[\sum_x \frac{\mu(x) \exp(u(x,a)/\lambda)}{\sum_{a'} q^{(k)}(a') \exp(u(x,a')/\lambda)}\right] q^{(k)}(a).$$

The term in brackets is the left side of (8) and determines if $q(a)$ rises or falls. The algorithm can converge to two limit points. One limit is $q(a) > 0$ and the term in brackets is equal to one, a case that includes $q(a) = 1$, or the term in brackets is less than one and the other limit is $q(a) = 0$. Condition (7) represents a twist in state-dependent choice in the direction of the high payoff states. Condition (8) ensures that these twists average out to one. If they do not then the probability of an action needs to be raised or lowered accordingly.

Next we present the posterior-based approach of CD (2013). Observe that one can equivalently rewrite the joint distribution as $\mu(x,a) = \mu(x) p(a|x) = q(a) \mu(x|a)$, and decompose the prior into

$$\mu(x) = \sum_a \mu(x|a) q(a), \tag{11}$$

where $\mu(x|a) \in \Delta(X|A)$ denotes the posterior distribution if $q(a) > 0$. If $q(a) = 0$, we simply set $\mu(x|a) = 0$. Now Problem 2 becomes:

Problem 3 *(posterior-based RI problem) Choose $\{\mu(x|a)\} \in \Delta(X|A)$ and $q \in \Delta(A)$ to solve*

$$V(\mu) = \max \ \mathbb{E}[u(x,a)] - \lambda I(x;a) = \sum_{x,a} q(a) \mu(x|a) \left(u(x,a) - \lambda \ln \frac{\mu(x|a)}{\mu(x)}\right)$$

subject to (11).

Following CD (2013) and CDL (2018a), we rewrite this problem as

$$\overline{V}(\mu) = \max_{q,\mu(\cdot|\cdot)} \sum_a q(a) N^a(\mu(\cdot|a)), \quad V(\mu) = \overline{V}(\mu) - \lambda H(\mu) \tag{12}$$

subject to (11), where $N^a(\mu(\cdot|a))$ denotes the net utility of action a defined as

$$N^a(\mu(\cdot|a)) \equiv \sum_x \mu(x|a) u(x,a) + \lambda H(\mu(\cdot|a)),$$

and $H(\mu) = -\sum_x \mu(x) \ln \mu(x)$ is the entropy of μ, which does not affect the optimization.

Notice that $N^a(\mu(\cdot|a))$ is concave in $\mu(\cdot|a)$, but the problem in (12) is not jointly concave in q and $\mu(\cdot|\cdot)$ due to the cross product term as pointed out by CD (2013). Thus one cannot simply use the Kuhn-Tucker conditions to solve this problem. CD (2013) instead propose a geometric approach from the convex analysis and derive conditions that are equivalent to those in Proposition 1.

The following result is critical for the dynamic model and its proof together with all other proofs are given in Appendix A.

Proposition 2 *(i) The optimal posteriors $\mu(\cdot|a)$ for all chosen actions a such that $q(a) \in (0,1)$ are independent of the prior $\mu \in \Delta(X)$ in the convex hull of these posteriors. (ii) The optimal payoff for the static RI problem is given by*

$$V(\mu) = \sum_x \mu(x) \widetilde{V}(x) = \sum_x \mu(x) \widehat{V}(x) - \lambda H(\mu), \tag{13}$$

where for $\mu(x) > 0$

$$\widetilde{V}(x) = \lambda \ln \left[\sum_a q(a) \exp(u(x,a)/\lambda) \right], \ \widehat{V}(x) = \widetilde{V}(x) - \lambda \ln \mu(x),$$

and $\widehat{V}(x)$ is independent of the prior $\mu \in \Delta(X)$ in the convex hull of the optimal posteriors $\mu(\cdot|a)$ for all chosen actions a such that $q(a) \in (0,1)$. (iii) $\overline{V}(\mu)$ is concave, $V(\mu)$ is convex, and for $\mu(i) \in (0,1)$, $i = 1, ..., m-1$,

$$\frac{\partial \overline{V}(\mu)}{\partial \mu(i)} = \widehat{V}(i) - \widehat{V}(m), \ \frac{\partial V(\mu)}{\partial \mu(i)} = \widetilde{V}(i) - \widetilde{V}(m).$$

Part (i) is the LIP property discovered by CD (2013). The first equality in (13) can be derived using either the choice-based or posterior-based approach. The second equality and the local invariance property for $\widehat{V}(x)$ are our new finding and can be best understood using the geometric approach of CD (2013). Specifically, $\mu(\cdot|a)$ is the tangent point of the net utility associated with the chosen action a and $\widehat{V}(x)$ satisfies

$$\overline{V}(\mu) = \sum_a q(a) N^a(\mu(\cdot|a)) = \sum_x \widehat{V}(x) \mu(x),$$

at the optimum. The value $\overline{V}(\mu)$ is the height above $\mu(x)$ of the convex hull connecting $N^a(\mu(\cdot|a))$ for all chosen actions a. The value $\widehat{V}(x)$ is the height of the hyperplane containing this convex hull at the point with $\mu(x)=1$ and $\mu(x')=0$ for all $x' \neq x$. This value is independent of the prior μ in that convex hull. This result does not appear in the literature and is critical for the analysis of the dynamic model. Notice that we need at least two chosen actions to form a convex hull. If there is only one chosen action a, then $q(a)=1$ and the posterior is the same as the prior. In this case the convex hull is a degenerate singleton. Part (iii) follows from part (ii) and the convexity of $V(\mu)$ poses difficulty of the dynamic RI problem.

[Insert Figure 1 Here.]

Figure 1 is similar to Figure 5 of CDL (2018a) in the case with two states $\{x, x'\}$ and two actions $\{a, b\}$. Net utilities are represented by the two solid curves. The concavification $\overline{V}(\mu)$ is the concave envelope of these two curves. The optimal posteriors $\mu(\cdot|a)$ and $\mu(\cdot|b)$ are given by the tangent points at which the hyperplane supports the two net utility functions. The value $\widehat{V}(x)$ is given by the height of the hyperplane at the point with $\mu(x)=1$. Both the optimal posteriors, $\widehat{V}(x)$, and $\widehat{V}(x')$ are invariant to changes of $\mu(x')$ within the interval $(\mu(x'|a), \mu(x'|b))$. If $\mu(x') \in (0, \mu(x'|a)]$, then $q(a)=1$ and $\mu(x'|a)=\mu(x')$. If $\mu(x') \in [\mu(x'|b), 1]$, then $q(b)=1$ and $\mu(x'|b)=\mu(x')$.

3.2 Two-Period Case

As a prelude for our dynamic analysis we study the two-period case with $T=2$ and $u_{T+1}=0$. The decision problem is given by

$$\max_{p_1,\ p_2} \mathbb{E}\left[u(x_1, a_1) + \beta u(x_2, a_2)\right] - \lambda I(x_1; a_1) - \lambda\beta I\left(x^2; a_2|a_1\right), \tag{14}$$

where $p_1(a_1|x_1) \in \Delta(A|X)$ and $p_2\left(a_2|x^2, a_1\right) \in \Delta\left(A|X^2 \times A\right)$ and the joint distribution $\mu_2\left(x^2, a^2\right)$ satisfies (1). In Section 4.1 we will show that we only need to focus on choice rules that depend on the current state only without loss of performance. We thus assume that p_2 takes the form $p_2(a_2|x_2, a_1)$ and hence $I\left(x^2; a_2|a_1\right) = I(x_2; a_2|a_1)$.

To apply the posterior-based approach, we rewrite the objective function in (14) as

$$\begin{aligned} & J\left(q_1, \mu_1(\cdot|\cdot), q_2, \mu_2(\cdot|\cdot)\right) \equiv \mathbb{E}\left[u(x_1, a_1) + \beta u(x_2, a_2)\right] - \lambda I(x_1; a_1) - \lambda\beta I(x_2; a_2|a_1) \\ = & \sum_{a_1, x_1} q_1(a_1)\, \mu_1(x_1|a_1) \left[u(x_1, a_1) - \lambda \ln \frac{\mu_1(x_1|a_1)}{\mu_1(x_1)}\right] \\ & + \beta \sum_{a_1, a_2, x_2} q_1(a_1)\, q_2(a_2|a_1)\, \mu_2\left(x_2|a^2\right) \left[u(x_2, a_2) - \lambda \ln \frac{\mu_2\left(x_2|a^2\right)}{\mu_2(x_2|a_1)}\right], \end{aligned}$$

where $q_1 \in \Delta(A)$, $q_2 \in \Delta(A|A)$, $\mu_1(\cdot|\cdot) \in \Delta(X|A)$, $\mu_2(\cdot|\cdot) \in \Delta\left(X|A^2\right)$, and for all $x_2 \in X$, $a_1 \in A$,

$$\mu_2(x_2|a_1) \equiv \sum_{a_2} \mu_2\left(x_2|a^2\right) q_2(a_2|a_1) = \sum_{x_1} \pi(x_2|x_1, a_1)\, \mu_1(x_1|a_1). \tag{15}$$

We consider interior solutions with $q_1(a_1) > 0$ and $q_2(a_2|a_1) > 0$ for simplicity. Our analysis also applies to corner solutions as discussed in Section 4. We call $\mu_2(x_2|a_1)$ the *predictive distribution* of x_2 given a_1 if $q_1(a_1) > 0$. Equation (15) shows that this predictive distribution can be written in two ways: (1) marginalizing the posterior distribution $\mu_2(x_2|a^2)$ of x_2 given a^2 over the conditional distribution $q_2(a_2|a_1)$ of a_2 given a_1; (2) marginalizing the transition kernel $\pi(x_2|x_1, a_1)$ over the distribution $\mu_1(x_1|a_1)$ of x_1 given a_1.

Now we formulate the two-period RI problem as follows:

Problem 4 *Choose* $q_1 \in \Delta(A), q_2 \in \Delta(A|A), \mu_1(\cdot|\cdot) \in \Delta(X|A),$ *and* $\mu_2(\cdot|\cdot) \in \Delta\left(X|A^2\right)$ *to solve*

$$\max J(q_1, \mu_1(\cdot|\cdot), q_2, \mu_2(\cdot|\cdot))$$

subject to (15) and $\mu_1(x_1) = \sum_{a_1} q_1(a_1)\mu_1(x_1|a_1)$ *for all* $x_1 \in X$.

We solve Problem 4 by dynamic programming using the predictive distribution as the state variable. First consider the RI problem in period 2 for $q_1(a_1) > 0:$

$$V_2(\mu_2(\cdot|a_1)) = \max \sum_{x_2,a_2} q_2(a_2|a_1)\mu_2\left(x_2|a^2\right)\left[u(x_2,a_2) - \lambda \ln \frac{\mu_2\left(x_2|a^2\right)}{\mu_2(x_2|a_1)}\right] \tag{16}$$

subject to

$$\mu_2(x_2|a_1) = \sum_{a_2} \mu_2\left(x_2|a^2\right) q_2(a_2|a_1)$$

for all $x_2 \in X$ such that $\mu_2(x_2|a_1) > 0$. The choice variables are $\mu_2\left(\cdot|a^2\right)$ and $q_2(\cdot|a_1)$. Taking the predictive distribution $\mu_2(\cdot|a_1)$ as the prior at history a_1, we view this problem as a static RI problem studied in the previous subsection.

By Proposition 2, we have

$$V_2(\mu_2(\cdot|a_1)) = \sum_{x_2} \mu_2(x_2|a_1)\widetilde{V}_2(x_2,a_1), \tag{17}$$

where

$$\widetilde{V}_2(x_2,a_1) = \lambda \ln\left[\sum_{a_2} q_2(a_2|a_1)\exp(u(x_2,a_2)/\lambda)\right], \tag{18}$$

and $q_2(a_2|a_1)$ is an optimal solution. Moreover, the function

$$\widehat{V}_2(x_2,a_1) \equiv \widetilde{V}_2(x_2,a_1) - \lambda \ln \mu_2(x_2|a_1), \tag{19}$$

and the optimal posterior $\mu_2\left(\cdot|a^2\right) \in (0,1)$ are independent of the prior $\mu_2(\cdot|a_1)$ in the convex hull of the optimal posteriors $\mu_2\left(\cdot|a^2\right)$ for all a_2 such that $q_2(a_2|a_1) > 0$. Since the history a_1 enters the problem in (16) through $\mu_2(\cdot|a_1)$ only, $\mu_2\left(\cdot|a^2\right)$ is independent of a_1 and can be written as $\mu_2(\cdot|a_2)$. Similarly, $\widehat{V}_2(x_2,a_1)$ is also independent of a_1 and can be written as $\widehat{V}_2(x_2)$.

By dynamic programming, the problem in period 1 is to choose $\{\mu_1(x_1|a_1)\}$ and $\{q_1(a_1)\}$ to solve:

$$\begin{aligned} V_1(\mu_1) &= \max \sum_{a_1,x_1} q_1(a_1)\mu_1(x_1|a_1)\left[u(x_1,a_1) - \lambda \ln \frac{\mu_1(x_1|a_1)}{\mu_1(x_1)}\right] \\ &\quad +\beta \sum_{a_1} q_1(a_1) V_2(\mu_2(\cdot|a_1)) \end{aligned} \tag{20}$$

subject to $\mu_1(x_1) = \sum_{a_1} q_1(a_1)\mu_1(x_1|a_1)$ for all $x_1 \in X$ and

$$\mu_2(x_2|a_1) = \sum_{x_1} \pi(x_2|x_1,a_1)\mu_1(x_1|a_1) \text{ for all } x_2 \in X. \tag{21}$$

The link between the problems in the two periods is through the predictive distribution $\mu_2(x_2|a_1)$ in (21).

We transform the problem in (20) into a posterior-based form similar to (12). Substituting (17), (19), and (21) into (20) yields

$$V_1(\mu_1) = \max_{q_1,\mu_1(\cdot|\cdot)} \sum_{a_1} q_1(a_1) N_G^{a_1}(\mu_1(\cdot|a_1)) - \lambda H(\mu_1), \tag{22}$$

where we define the net utility associated with action a_1 as

$$N_G^{a_1}(\mu_1(\cdot|a_1)) \equiv \sum_{x_1} \mu_1(x_1|a_1)\hat{u}(x_1,a_1) - \lambda G(\mu_1(\cdot|a_1)).$$

Here the new utility function is given by

$$\hat{u}(x_1,a_1) = u(x_1,a_1) + \beta \sum_{x_2} \pi(x_2|x_1,a_1)\widehat{V}_2(x_2),$$

and the entropy cost is given by

$$G(\mu_1(\cdot|a_1)) = \beta H(\mu_2(\cdot|a_1)) - H(\mu_1(\cdot|a_1)), \tag{23}$$

where $\mu_2(\cdot|a_1)$ satisfies (21).

The problem in (22) is similar to, but different from that in (12) due to additional terms introduced by the problem in period 2. We arrange terms suitably in (22) in order to apply the general method of CD (2013) that works for general information cost functions. We need the net utility $N_G^{a_1}(\mu_1(\cdot|a_1))$ to be concave in $\mu_1(\cdot|a_1)$ so that the supporting hyperplane theorem can be applied. Since we have shown that $\widehat{V}_2(x_2)$ is independent of the prior $\mu_2(\cdot|a_1)$ in the convex hull of posteriors $\mu_2(\cdot|a_2)$ for all chosen actions a_2 by Proposition 2 and $\mu_1(\cdot|a_1)$ affects the objective in the second period only through $\mu_2(\cdot|a_1)$, we deduce that $\widehat{V}_2(x_2)$ is independent of $\mu_1(\cdot|a_1)$. Thus $\hat{u}(x_1,a_1)$ is independent of $\mu_1(\cdot|a_1)$.

Now the concavity of the net utility function $N_G^{a_1}(\mu_1(\cdot|a_1))$ is equivalent to the convexity of G. This property holds for general UPS information cost functions as shown in Appendix E. Since the Shannon entropy is a concave function, it is not obvious that G is a convex function.

Lemma 2 *For any $\beta \in (0,1]$ and $a_1 \in A$ with $q_1(a_1) > 0$ the function $G(\mu_1(\cdot|a_1))$ is convex in $\mu_1(\cdot|a_1)$ and the net utility $N_G^{a_1}(\mu_1(\cdot|a_1))$ is concave in $\mu_1(\cdot|a_1)$.*

We can then apply Lemmas 1 through 3 in CD (2013) to derive the necessary and sufficient conditions for optimality for the two-period problem. These conditions can be equivalently written in the choice-based form as in Proposition 1. In particular, the choice rule $\{p_1(a_1|x_1)\}$ in period 1 is with respect to the payoff $v_1(x_1,a_1) = u(x_1,a_1) + \beta \sum_{x_2} \pi(x_2|x_1,a_1) \widetilde{V}_2(x_2,a_1)$. We will present these conditions in Propositions 6, 7, and 10 for the general dynamic RI problem. We prefer the choice-based form because it is easier to solve numerically using our algorithm described in Section 4.3. SSM (2017) apply the choice-based approach to derive similar conditions. In the next subsection we will illustrate the pitfall of this approach in dynamic models.

3.3 Nonconcavity

In this subsection we provide an example to illustrate that concavity is important for the choice-based approach of SSM (2017) to work. Without concavity, the Kuhn-Tucker first-order conditions are not sufficient for optimality. We still consider the two-period RI problem in (14) and focus on interior solutions.

SSM (2017) propose a method by first breaking down the dynamic RI problem into a sequence of static RI problems and then applying Proposition 1 to each static problem. Adapting their method and using Lemma 1, we find that the RI problem in (14) is equivalent to the following control problem

$$V_1(\mu_1) \equiv \max_{p_1,p_2,q_1,q_2} F(p_1,p_2,q_1,q_2), \tag{24}$$

where $p_1 \in \Delta(A|X)$, $p_2 \in \Delta(A|X \times A)$, $q_1 \in \Delta(A)$, $q_2 \in \Delta(A|A)$, and

$$\begin{aligned} F(p_1,p_2,q_1,q_2) &= \sum_{a_1,x_1} \mu_1(x_1)\, p_1(a_1|x_1) \left[u(x_1,a_1) - \lambda \ln \frac{p_1(a_1|x_1)}{q_1(a_1)} \right] \\ &\quad + \beta \sum_{a_1,a_2,x_2} \mu_2(x_2,a_1)\, p_2(a_2|x_2,a_1) \left[u(x_2,a_2) - \lambda \ln \frac{p_2(a_2|x_2,a_1)}{q_2(a_2|a_1)} \right], \end{aligned} \tag{25}$$

where the joint distribution $\mu_2(x_2,a_1)$ satisfies

$$\mu_2(x_2,a_1) = \sum_{x_1} \pi(x_2|x_1,a_1)\, p_1(a_1|x_1)\, \mu_1(x_1) = \mu_2(x_2|a_1)\, q_1(a_1). \tag{26}$$

Notice that F may not be jointly concave, but is coordinate-wise concave. If p_1, p_2, q_1, and q_2 are optimal solutions to (24), then they are also coordinate-wise optimal. We can then use the coordinate-wise first-order conditions to characterize the optimal solution.

By dynamic programming, we write the problem in period 2 as

$$W_2(\{\mu_2(x_2,a_1)\}) \equiv \sum_{a_1} q_1(a_1)\, \widetilde{W}_2(a_1), \tag{27}$$

where for $q_1(a_1) > 0$

$$\widetilde{W}_2(a_1) \equiv \max_{p_2,q_2} \sum_{a_2,x_2} \mu_2(x_2|a_1)\, p_2(a_2|x_2,a_1) \left[u(x_2,a_2) - \lambda \ln \frac{p_2(a_2|x_2,a_1)}{q_2(a_2|a_1)} \right]. \tag{28}$$

This is the same as the static RI problem in (5) with prior $\mu_2(\cdot|a_1)$. We can use Proposition 1 to derive $\{p_2(a_2|x_2,a_1)\}$ and $\{q_2(a_2|a_1)\}$.

By Proposition 2, the value function satisfies

$$\widetilde{W}_2(a_1) = \sum_{x2} \mu_2(x_2|a_1)\, \widetilde{V}_2(x_2,a_1) = \sum_{x_2} \mu(x_2|a_1)\, \widehat{V}(x_2) - \lambda H(\mu_2(\cdot|a_1)), \tag{29}$$

where

$$\widetilde{V}_2(x_2,a_1) = \lambda \ln \sum_{a_2} q_2(a_2|a_1) \exp[u(x_2,a_2)/\lambda], \quad \widehat{V}(x_2) = \widetilde{V}_2(x_2,a_1) - \ln \mu_2(x_2|a_1), \tag{30}$$

and $\widehat{V}(x_2)$ is independent of $\mu_2(\cdot|a_1)$ in the convex hull of optimal posteriors for chosen actions. It follows from (26), (27), and (29) that

$$W_2(\{\mu_2(x_2,a_1)\}) = \sum_{x_2,a_1} \mu_2(x_2,a_1)\, \widetilde{V}_2(x_2,a_1) = \sum_{x_2,x_1,a_1} \pi(x_2|x_1,a_1)\, p_1(a_1|x_1)\, \mu_1(x_1)\, \widetilde{V}_2(x_2,a_1).$$

Notice that $\widetilde{W}_2(a_1) = V_2(\mu_2(\cdot|a_1))$.

By dynamic programming, we rewrite the period 1 objective function in (24) as

$$\sum_{a_1,x_1} \mu_1(x_1)\, p_1(a_1|x_1) \left[u(x_1,a_1) + \sum_{x_2} \pi(x_2|x_1,a_1)\, \widetilde{V}_2(x_2,a_1) - \lambda \ln \frac{p_1(a_1|x_1)}{q_1(a_1)} \right]. \tag{31}$$

SSM (2017) suggest to treat the problem in period 1 as a static RI problem with exogenous payoff

$$v_1(x_1,a_1) \equiv u(x_1,a_1) + \sum_{x_2} \pi_2(x_2|x_1,a_1)\, \widetilde{V}_2(x_2,a_1),$$

and derive coordinate-wise first-order conditions with respect to p_1 and q_1 for fixed p_2 and q_2.

We find that the first-order conditions for p_1, q_1, p_2, and q_2 are the same as those derived using the posterior-based approach in the previous subsection. The CD and CDL type sufficient conditions are also the same. This result holds for any horizon $T \geq 2$ as shown in Propositions 6 and 7 and Appendix C. There is an implicit concavity assumption behind the derivations above. We need the objective function in (31) to be jointly concave in p_1 and q_1. SSM (2017) argue that one can treat the continuation value $\widetilde{V}_2(x_2,a_1)$ as fixed by ignoring the effect of p_1 and q_1 on future beliefs $\mu_2(x_2|a_1)$.

Since $\mu_2(x_2|a_1)$ satisfies

$$\mu_2(x_2|a_1) = \sum_{x_1} \pi(x_2|x_1,a_1)\, \mu_1(x_1|a_1) = \frac{\sum_{x_1} \pi(x_2|x_1,a_1)\, p_1(a_1|x_1)\, \mu_1(x_1)}{q_1(a_1)}, \tag{32}$$

and $\widetilde{V}_2(x_2, a_1)$ satisfies (30), we know that p_1 and q_1 affect $\widetilde{V}_2(x_2, a_1)$ in a nonlinear way. Taking this impact into account, we use (30) and (32) to rewrite (31) as

$$\sum_{x_1} \mu_1(x_1) \left[\sum_{a_1} p_1(a_1|x_1)\, \hat{u}(x_1, a_1) + \lambda \Phi(p_1, q_1; x_1) \right], \tag{33}$$

where we define

$$\hat{u}(x_1, a_1) \equiv u(x_1, a_1) + \beta \sum_{x_2} \pi(x_2|x_1, a_1) \left[\widehat{V}_2(x_2) + \lambda \ln \mu_1(x_1) \right],$$

and

$$\Phi(p_1, q_1; x_1) \equiv \sum_{x_2, a_1} \pi(x_2|x_1, a_1)\, p_1(a_1|x_1) \ln \frac{\left[\sum_{x_1'} \pi(x_2|x_1', a_1)\, p_1(a_1|x_1')\right]^{\beta}}{p_1(a_1|x_1)\, [q_1(a_1)]^{\beta-1}}.$$

Now the concavity of the optimization problem is equivalent to the concavity of $\Phi(p_1, q_1; x_1)$. In Lemma 4 of Appendix C we show that $\Phi(p_1, q_1; x_1)$ is indeed concave in p_1 and q_1 if $\beta \in (0, 1]$. Taking first-order conditions for (33) give the same result as in Proposition 6 and the sufficient conditions in Proposition 7 guarantee the optimality of these conditions.

We provide a numerical example to illustrate that $\Phi(p_1, q_1; x_1)$ is not concave if $\beta > 1$ is sufficiently large. The intuition is that the convexity of the continuation value $\widetilde{V}_2(x_2, a_1)$ in $\mu_2(x_2|a_1)$ may dominate the concavity of the payoff in period 1. We introduce discounting to the example in Section 4.1 of SSM (2017). Let $X = A = \{1, 2\}$, $u(x_t, a_t) = 1$ if $x_t = a_t$ and $u(x_t, a_t) = 0$ if $x_t \neq a_t$ for $t = 1, 2$. Let $\mu_1(1) = 0.5$ and $\lambda = 1$. The state transition kernel $\pi_{t+1}(x_{t+1}|x_t, a_t)$ is independent of a_t and satisfies $\pi_{t+1}(x_{t+1}|x_t, a_t) = 1 - \gamma$ if $x_{t+1} = x_t$, and $\pi_{t+1}(x_{t+1}|x_t, a_t) = \gamma$, otherwise, for $t = 1$.

Let $\gamma = 0.2$ and $\beta = 10$.[6] We numerically find that the following interior solution satisfies all first-order conditions including the CD and CDL type sufficient conditions: $q_1(1) = q_1(2) = 1/2$, $q_2(2|2) = q_2(1|1) = 0.376$, $p_1(1|1) = p_1(2|2) = 0.406$, $p_2(1|1, 1) = 0.621$, $p_2(1|1, 2) = 0.819$, $p_2(1|2, 1) = 0.181$, $p_2(1|2, 2) = 0.379$, and $\mu_2(x_2|a_1) = 0.4435$ if $x_2 = a_1$; $= 0.5565$, otherwise.

Figure 2 plots the continuation value $\widetilde{W}_2(a_1 = 1)$ as a function of the prior belief $\mu_2(x_2 = 1|a_1 = 1)$ at history $a_1 = 1$, the contour of the period 1 objective in (31) as a function of $p_1(1|1)$ and $p_1(1|2)$, and the contour of $\Phi(p_1, q_1; x_1 = 1)$ as a function $p_1(1|1)$ and $p_1(1|2)$. We have used $\sum_{x_1} p_1(a_1|x_1)\, \mu_1(x_1)$ to replace $q_1(a_1)$ in the last two functions. Figure 2 shows that the first two functions are convex and the last function is not concave. Thus the solution given above is not optimal, even though it satisfies all Kuhn-Tucker first-order conditions.

[Insert Figure 2 Here.]

[6] One can reinterpret the discount factor β as a scaling factor for the utility function and the marginal information cost in period 2.

3.4 UPS Cost Functions

In this subsection we show that our approach applies to the UPS information cost functions introduced by CD (2013) and CDL (2018b), but the choice-based approach of SSM (2017) may not. Replace the mutual information $I(x,a)$ in Problem 2 by the UPS information cost function defined as

$$C_H(\mu, \mu(\cdot|\cdot), q) = H(\mu) - \sum_a q(a) H(\mu(\cdot|a)),$$

where $H : \Delta(X) \to \mathbb{R}_+$ is a concave function and $\mu(x) = \sum_a q(a)\mu(x|a)$. Clearly, observing information reduces uncertainty so that $C_H(\mu, \mu(\cdot|\cdot), q) \geq 0$. We may view H as a generalized entropy and the Shannon entropy is a special case with $H(\nu) = -\sum_\nu \nu(x)\ln\nu(x)$. For a UPS cost function, the net utility function is concave in the static case. We can then apply Lemmas 1 through 3 of CD (2013) to characterize the solution.

In Appendix E we study a two-period RI problem and show that the concavity of the total net utility function is equivalent to the convexity of the function G defined in (23), where H is the generalized entropy defined here. As long as G is convex, we can still apply Lemmas 1 through 3 of CD (2013) in the static case to characterize the solution in the dynamic case. The UPS cost functions also satisfy the LIP property (CD (2013) and CDL (2018b)), which is important for dynamic models. This property allows us to derive Markovian solutions that facilitate efficient numerical methods.

To illustrate why the choice-based approach may not work for general UPS cost functions, we let H be the weighted entropy (Belis and Guiasu (1968)) defined as $H(\nu) = -\sum_x w(x)\nu(x)\ln\nu(x)$, where the weighting function satisfies $w(x) \geq 0$ and $\sum_x w(x) = 1$. In this case the cost function becomes

$$C_H(\mu, \mu(\cdot|\cdot), q) = \sum_{x,a} w(x) q(a)\mu(x|a)\ln\frac{\mu(x|a)}{\mu(x)} = \sum_{x,a} w(x)\mu(x)p(a|x)\ln\frac{p(a|x)}{q(a)}.$$

Following the choice-based approach described in Section 3.1, we define

$$F(p,q) = \sum_{a,x}\mu(x)p(a|x)\left[u(x,a) - \lambda w(x)\ln\frac{p(a|x)}{q(a)}\right].$$

One can check that Lemma 1 does not hold in general so that the static RI problem is not equivalent to the problem in (9) for general UPS cost functions. Similarly, Lemma 2 in SSM (2017) also fails for general UPS cost functions in dynamic models. Thus the coordinate-wise first-order conditions for p and q cannot be used to characterize the solutions to dynamic RI problems.

4 Main Results

In this section we characterize solutions by dynamic programming, provide necessary and sufficient first-order conditions for optimality, and describe a numerical algorithm to solve these conditions.

4.1 Dynamic Programming

We start by the finite-horizon case $T < \infty$ in Problem 1. We adopt the joint distribution $\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}$ as the state variable and define the value function as

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \max_{\{p_k\}_{k=t}^T} \mathbb{E}\left[\sum_{k=t}^{T} \beta^{k-t}\left(u\left(x_k, a_k\right) - \lambda I\left(x^k; a_k | a^{k-1}\right)\right)\right] + \beta^{T+1-k}\mathbb{E}u_{T+1}\left(x_{T+1}\right) \tag{34}$$

for $t \geq 1$, where each $p_k \in \Delta\left(A | X^k \times A^{k-1}\right)$ and the expectation is taken with respect to the marginal distribution of x^T and a^T induced by the recursive equation (1) starting at $\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}$.

By the principle of optimality (Stokey, Lucas with Prescott (1989)), the value function satisfies the Bellman equation

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \max_{p_t} \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t | a^{t-1}\right) + \beta W_{t+1}\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right) \tag{35}$$

subject to (1), for $t = 1, 2, ..., T$, with the terminal condition

$$W_{T+1}\left(\left\{\mu_{T+1}\left(x^{T+1}, a^T\right)\right\}\right) = \mathbb{E}u_{T+1}\left(x_{T+1}\right). \tag{36}$$

The dynamic programming problem (35) is complicated to solve in general because of the history dependence. Following Proposition 1 of TSS (2018), we can show that this problem can be simplified. Specifically the search for the optimal stochastic choice rules can be restricted to the class of rules of the form $p_t\left(a_t | x_t, a^{t-1}\right)$ without loss of performance. For this class of rules, the mutual information satisfies $I\left(x^t; a_t | a^{t-1}\right) = I\left(x_t; a_t | a^{t-1}\right)$ and the value function depends only on the marginal distribution $\mu_t\left(x_t, a^{t-1}\right)$.[7]

We now generalize this result to the infinite-horizon case as $T \to \infty$.

Proposition 3 *Consider the infinite-horizon RI problem in (4) with $T = \infty$.*

(i) There exists an optimal choice rule of the form $\hat{\mathbf{p}} = \left\{\hat{p}_t\left(a_t | x_t, a^{t-1}\right)\right\}_{t=1}^{\infty}$.

(ii) Under the choice rule $\hat{\mathbf{p}}$, we have

$$I\left(x^t; a_t | a^{t-1}\right) = I\left(x_t; a_t | a^{t-1}\right) \text{ for all } t \geq 1.$$

(iii) As $T \to \infty$, the T-horizon value function $W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right)$ converges to the infinite-horizon value function which depends only on the marginal distribution $\left\{\mu_t\left(x_t, a^{t-1}\right)\right\}$.

By this proposition we can write the one-period payoff as

$$\begin{aligned}
\mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x_t; a_t | a^{t-1}\right) &= \sum_{x_t, a^t} \mu_t\left(x_t, a^t\right)\left[u\left(x_t, a_t\right) - \lambda \ln \frac{\mu_t\left(x_t | a^t\right)}{\mu_t\left(x_t | a^{t-1}\right)}\right] \\
&= \sum_{x_t, a^t} \mu_t\left(a^{t-1}\right) q_t\left(a_t | a^{t-1}\right) \mu_t\left(x_t | a^t\right)\left[u\left(x_t, a_t\right) - \lambda \ln \frac{\mu_t\left(x_t | a^t\right)}{\mu_t\left(x_t | a^{t-1}\right)}\right],
\end{aligned}$$

[7] The key to the proof is to apply the additivity property of the Shannon mutual information.

where $\mu_t\left(a^{t-1}\right)$, $q_t\left(a_t|a^{t-1}\right)$, $\mu_t\left(x_t|a^t\right)$, and $\mu_t\left(x_t|a^{t-1}\right)$ denote the marginal distribution of a^{t-1}, the conditional distribution of a_t given history a^{t-1}, the posterior distribution of x_t given a^t, and the predictive distribution of x_t given a^{t-1}, respectively.

Noticing that $\mu_t\left(a^t\right) = \mu_t\left(a^{t-1}\right) q_t\left(a_t|a^{t-1}\right)$, we can write the value function as

$$W_t\left(\left\{\mu_t\left(x_t, a^{t-1}\right)\right\}\right) = \sum_{a^{t-1}} \mu_t\left(a^{t-1}\right) V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right),$$

whenever $\mu_t\left(a^{t-1}\right) > 0$, where $V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right)$ satisfies the Bellman equation

$$\begin{aligned} V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right) &= \max \sum_{x_t, a_t} q_t\left(a_t|a^{t-1}\right) \mu_t\left(x_t|a^t\right)\left[u\left(x_t, a_t\right) - \lambda \ln \frac{\mu_t\left(x_t|a^t\right)}{\mu_t\left(x_t|a^{t-1}\right)}\right] \\ &\quad + \beta \sum_{a_t} q_t\left(a_t|a^{t-1}\right) V_{t+1}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) \end{aligned} \tag{37}$$

subject to

$$\mu_t\left(x_t|a^{t-1}\right) = \sum_{a_t} q_t\left(a_t|a^{t-1}\right) \mu_t\left(x_t|a^t\right) \text{ all } x_t \in X, \tag{38}$$

$$\mu_{t+1}\left(x_{t+1}|a^t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t, a_t\right) \mu_t\left(x_t|a^t\right) \text{ all } x_{t+1} \in X. \tag{39}$$

The choice variables are $q_t\left(\cdot|a^{t-1}\right) \in \Delta(A)$ and $\mu_t\left(\cdot|a_t, a^{t-1}\right) \in \Delta(X)$ for all $a_t \in A$. Equations (38) and (39) give two ways to decompose the predictive distribution. In the finite-horizon case with $T < \infty$, there is a terminal condition

$$V_{T+1}\left(\mu_{T+1}\left(\cdot|a^T\right)\right) = \sum_{x_{T+1}} \mu_{T+1}\left(x_{T+1}|a^T\right) u_{T+1}\left(x_{T+1}\right).$$

Notice that $q_t\left(a_t|a^{t-1}\right) = \mu_t\left(x_t|a^t\right) = 0$ whenever $\mu_t\left(x_t|a^{t-1}\right) = 0$. At the initial date we set $\mu_1\left(x_1, a^0\right) = \mu_1\left(x_1|a^0\right) \equiv \mu_1\left(x_1\right)$, $\mu_1\left(a^0\right) \equiv 1$, $q_1\left(a_1|a^0\right) \equiv q_1\left(a_1\right)$, and $p_1\left(a_1|x_1, a_0\right) \equiv p_1\left(a_1|x_1\right)$. We also have $V_1\left(\mu_1\left(\cdot|a^0\right)\right) = W_1\left(\left\{\mu_1\left(x_1, a^0\right)\right\}\right)$.

We will focus on the analysis of the Bellman equation in (37), in which the predictive distribution $\mu_t\left(\cdot|a^{t-1}\right) > 0$ is used as the state variable and the choice variables are the posterior $\left\{\mu_t\left(x_t|a^t\right)\right\} \in \Delta(X|A)$ and the default rule $\left\{q_t\left(a_t|a^{t-1}\right)\right\} \in \Delta(A)$ given a^{t-1}. After obtaining the optimal solutions for $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$, we can derive the stochastic choice rule using the Bayes rule:

$$p_t\left(a_t|x_t, a^{t-1}\right) = \frac{\mu_t\left(x_t|a^t\right) q_t\left(a_t|a^{t-1}\right)}{\mu_t\left(x_t|a^{t-1}\right)}. \tag{40}$$

4.2 Markovian Logit Rule

The solution to the dynamic RI problem may be history dependent, making both analytical characterizations and numerical methods complicated. Unlike SSM (2017), we will adopt the posterior-based approach to provide a dynamic logit rule characterization in Appendix B. To simplify the analysis, here we focus on Markovian solutions.

Definition 2 *An optimal solution to the dynamic RI problem in (4) is Markovian if, for any two histories a^{t-1} and $\left\{b^{t-2}, a_{t-1}\right\}$ reached with positive probabilities and any $t \leq T$, the implied predictive distributions satisfy $\mu_t(x_t|a^{t-1}) = \mu_t(x_t|a_{t-1}, b^{t-2})$.*

Intuitively, the predictive distribution is the state variable in the posterior-based dynamic programming problem. If this state variable is history independent, then the optimal solution will also be history independent. We thus have the following result.

Proposition 4 *For a Markovian solution to the dynamic RI problem in (4), the choice rule $p_t(a_t|x_t, a^{t-1})$ and the default rule $q_t(a_t|a^{t-1})$ take the form of $p_t(a_t|x_t, a_{t-1})$ and $q_t(a_t|a_{t-1})$, respectively, for any $t \leq T$.*

In Appendix F we provide two numerical examples to illustrate Definition 2 and Proposition 4. We say that a solution to the dynamic RI problem in (4) is interior if $q_t(a_t|a^{t-1}) > 0$ for any action $a_t \in A$ and all $t \geq 1$.

Proposition 5 *An interior solution is Markovian.*

Now we provide necessary and sufficient conditions for Markovian solutions, which may not be interior, using the posterior-based approach.

Proposition 6 *(necessary conditions) Let $\beta \in (0, 1)$. Then the Markovian solution to the dynamic RI problem in (4), $\{p_t\left(a_t|x_t, a_{t-1}\right)\}_{t=1}^T$ and $\{q_t\left(a_t|a_{t-1}\right)\}_{t=1}^T$, is characterized by the following system of difference equations for $t = 1, 2, ..., T$:*

$$p_t\left(a_t|x_t, a_{t-1}\right) = \frac{q_t\left(a_t|a_{t-1}\right) \exp\left(v_t\left(x_t, a_t\right)/\lambda\right)}{\sum_{a_t'} q_t\left(a_t'|a_{t-1}\right) \exp\left(v_t\left(x_t, a_t'\right)/\lambda\right)} \text{ for } \mu_t\left(x_t, a_{t-1}\right) > 0, \tag{41}$$

$$q_t\left(a_t|a_{t-1}\right) = \sum_{x_t} p_t\left(a_t|x_t, a_{t-1}\right) \mu_t\left(x_t|a_{t-1}\right) \text{ for } \mu_t\left(a_{t-1}\right) > 0, \tag{42}$$

where

$$v_t\left(x_t, a_t\right) = u\left(x_t, a_t\right) + \beta \sum_{x_{t+1}} \pi\left(x_{t+1}|x_t, a_t\right) \widetilde{V}_{t+1}\left(x_{t+1}, a_t\right), \tag{43}$$

$$\widetilde{V}_t\left(x_t, a_{t-1}\right) = \lambda \ln \sum_{a_t} q_t\left(a_t|a_{t-1}\right) \exp\left(v_t\left(x_t, a_t\right)/\lambda\right), \tag{44}$$

$$\mu_{t+1}\left(x_{t+1}, a_t\right) = \sum_{x_t, a_{t-1}} \pi\left(x_{t+1}|x_t, a_t\right) p_t\left(a_t|x_t, a_{t-1}\right) \mu_t\left(x_t, a_{t-1}\right), \tag{45}$$

$$\mu_t\left(x_t|a_{t-1}\right) = \frac{\mu_t\left(x_t, a_{t-1}\right)}{\mu_t\left(a_{t-1}\right)} \text{ for } \mu_t\left(a_{t-1}\right) = \sum_{x_t} \mu_t\left(x_t, a_{t-1}\right) > 0. \tag{46}$$

The value functions satisfy

$$W_t\left(\{\mu_t\left(x_t, a_{t-1}\right)\}\right) = \sum_{x_t, a_{t-1}} \mu_t\left(x_t, a_{t-1}\right) \widetilde{V}_t\left(x_t, a_{t-1}\right), \tag{47}$$

$$V_t\left(\mu_t\left(\cdot|a_{t-1}\right)\right) = \sum_{x_t} \mu_t\left(x_t|a_{t-1}\right) \widetilde{V}_t\left(x_t, a_{t-1}\right). \tag{48}$$

In the finite-horizon case with $T < \infty$, there is a terminal condition $\widetilde{V}_{T+1}\left(x_{T+1}, a_T\right) = u_{T+1}\left(x_{T+1}\right)$.

Proposition 6 shows that if the Markovian solution for $\{p_t\left(a_t|x_t, a_{t-1}\right)\}$ and $\{q_t\left(a_t|a_{t-1}\right)\}$ solves the dynamic RI problem, then it also solves the static RI problem with the prior belief at history a_{t-1}, $\mu_t\left(x_t|a_{t-1}\right)$, and payoff function $v_t\left(x_t, a_t\right)$.

The realized continuation value $\widetilde{V}_t\left(x_t, a_{t-1}\right)$ satisfies (44) and the expected optimal continuation value at time t is given by (47). The value function for the posterior-based dynamic programming satisfies (48). As in SSM (2017), the optimal choice rule is consistent with the dynamic logit rule with payoffs $v_t\left(x_t, a_t\right)/\lambda + \ln q_t\left(a_t|a_{t-1}\right)$ (Rust (1987)):

$$p_t\left(a_t|x_t, a_{t-1}\right) = \frac{\exp\left[v_t\left(x_t, a_t\right)/\lambda + \ln q_t\left(a_t|a_{t-1}\right)\right]}{\sum_{a_t'} \exp\left[v_t\left(x_t, a_t'\right)/\lambda + \ln q_t\left(a_t'|a_{t-1}\right)\right]}.$$

The payoff differs from the DM's true payoff $v_t\left(x_t, a_t\right)/\lambda$ by a predisposition term, $\ln q_t\left(a_t|a_{t-1}\right)$. The predisposition increases the relative payoff associated with actions that are chosen with high probability on average across all states given the DM's last acquired information.

Proposition 6 is related to Theorem 1, Proposition 3, and Lemma 6 of SSM (2017), who show that the dynamic RI problem can be characterized by solving a sequence of static RI problems. Our characterization complements theirs and facilitates numerical computations presented in the next subsection. In Appendix C we modify the choice-based approach of SSM (2017) and establish the concavity of the objective function. Then we derive the same conditions as in Proposition 6 for interior solutions only. The next result presents sufficient conditions.

Proposition 7 *(sufficient conditions) Suppose that the dynamic RI problem in (4) with $\beta \in (0,1)$ has an optimal Markovian solution. If conditions (41) through (46) and the following condition are satisfied*

$$\sum_{x_t} \frac{\mu_t\left(x_t|a_{t-1}\right)\exp\left(v_t\left(x_t, a_t\right)/\lambda\right)}{\sum_{a_t'} q_t\left(a_t'|a_{t-1}\right)\exp\left(v_t\left(x_t, a_t'\right)/\lambda\right)} \leq 1, \quad t = 1, 2, ..., T,$$

with equality if $q_t\left(a_t|a_{t-1}\right) > 0$, then $\{p_t\left(a_t|x_t, a_{t-1}\right)\}_{t=1}^T$ and $\{q_t\left(a_t|a_{t-1}\right)\}_{t=1}^T$ are an optimal Markovian solution.

We stress that the additional sufficient condition is valid only for jointly concave problems. Such concavity for dynamic RI models is difficult to establish using the choice-based approach. We resolve this issue using the posterior-based approach of CD (2013).

4.3 Numerical Methods

The conditions presented in Proposition 6 are a system of nonlinear difference equations, which is nontrivial to solve both analytically and numerically. To solve this system numerically, we extend the forward-backward Arimoto-Blahut algorithm proposed by TSS (2018) to our infinite-horizon model. We present the algorithm in Appendix D. Here we sketch the key idea.

We classify the difference equations in Proposition 6 into two groups. Equations (42), (45), and (46) form the first group, which characterizes $\{\mu_t(x_t, a_{t-1})\}$, $\{\mu_t(x_t|a_{t-1})\}$, and $\{q_t(a_t|a_{t-1})\}$, and equations (41), (43), and (44) form the second group, which characterizes $\{p_t(a_t|x_t, a_{t-1})\}$, $\{v_t(x_t, a_t)\}$ and $\{\widetilde{V}_t(x_t, a_{t-1})\}$. If the solution for $\{p_t(a_t|x_t, a_{t-1})\}$, $\{v_t(x_t, a_t)\}$ and $\{\widetilde{V}_t(x_t, a_{t-1})\}$ is known, the equations in the first group can be solved forward in time to obtain $\{\mu_t(x_t, a_{t-1})\}$, $\{\mu_t(x_t|a_{t-1})\}$, and $\{q_t(a_t|a_{t-1})\}$. On the other hand, if the solution for $\{\mu_t(x_t, a_{t-1})\}$, $\{\mu_t(x_t|a_{t-1})\}$, and $\{q_t(a_t|a_{t-1})\}$ is known, the equations in the second group, which can be viewed as Bellman equations, can be solved backward in time to compute $\{p_t(a_t|x_t, a_{t-1})\}$, $\{v_t(x_t, a_t)\}$ and $\{\widetilde{V}_t(x_t, a_{t-1})\}$. Thus, to solve these sequences simultaneously, we use the following iterative method: First, fixing the horizon $T < \infty$, perform the forward computation using the current best guess of the second group of unknowns, and then perform the backward computation using the updated guess of the first group of unknowns. Repeat this forward-backward iteration until convergence to obtain the solution for the T-horizon problem. Increase T until the value function $W_t(\{\mu_t(x_t, a_{t-1})\})$ converges to obtain the solution for the infinite-horizon problem. Discounting by $\beta \in (0, 1)$ is important for the last convergence.

The algorithm above generalizes the Arimoto-Blahut algorithm, which can be viewed as a block coordinate descent algorithm in multivariate convex optimization problems (Bertsekas (2016)). A sufficient condition for the convergence is that the objection function is jointly concave. We have shown that this condition is satisfied in our model so that the forward-backward Arimoto-Blahut algorithm converges to the optimal solution to the dynamic RI problem. We will use this algorithm to solve some numerical examples in the next section. Whenever a Markovian solution exists, our algorithm will find such a solution. But if a Markovian solution does not exist, our algorithm will find a Markovian solution that approximates the true history-dependent solution. We can design a similar algorithm for the history-dependent solution in Proposition 10 of Appendix B. This algorithm becomes complicated for long-horizon problems as the history increases with the horizon and becomes infeasible under infinite-horizon.

5 Applications

In this section we apply our results to a matching state problem often studied in the literature (SSM (2017), CD (2013) and CDL (2018a)). This problem can be used to describe many economic decisions, e.g., consumer choices, project selection, and job search. Suppose that $X = A$ and the utility function satisfies $u(x_t, a_t) = 1$ if $x_t = a_t$; and $u(x_t, a_t) = 0$, otherwise. We assume that the transition kernel is independent of actions in Section 5.1 as in SSM (2017) and allow it to depend on actions in Section 5.2. In these two subsections we also assume that $|X| = |A| = 2$ and $\mu_1(x_1 = 1) = 0.5$. In Section 5.3 we allow $|X| = |A| > 2$ and study the dynamics of consideration sets.

5.1 Transition Kernel Independent of Actions

As in SSM (2017), we assume $\pi(x_{t+1}|x_t, a_t) = \gamma$ whenever $x_{t+1} \neq x_t$ for any $a_t \in A$. We use this example to illustrate that rationally inattentive behavior exhibits the status quo bias over a short horizon, but not over an infinite horizon. Moreover, the infinite-horizon behavior exhibits inertia. For comparison, the optimal solution for the case without information cost ($\lambda = 0$) is to choose an action to match the state in each period.

With information cost $\lambda > 0$, we first consider the infinite-horizon stationary case, in which $p_t(a_t|x_t, a_{t-1})$, $q_t(a_t|a_{t-1})$, and $\mu_t(x_t|a_t)$ do not depend on time. By equations (38) and (39), we have

$$q(1|1)\,\mu(2|1) + q(2|1)\,\mu(2|2) = (1-\gamma)\,\mu(2|1) + \gamma\mu(1|1).$$

By symmetry $\mu(1|1) = \mu(2|2)$ and $q(1|1) = q(2|2)$. Then we obtain

$$q(a_t = 1|a_{t-1} = 1) = q(a_t = 2|a_{t-1} = 2) = 1-\gamma,$$

as long as $\mu(2|1) \neq \mu(1|1)$. By symmetry the initial default rule satisfies $q_1(a_1 = 1) = 1/2$. Using equations (41) through (47), we can determine the optimal stochastic rule $\{p_t\}$ and the path of optimal payoffs $\{W_t\}$. Using the forward-backward Arimoto-Blahut algorithm, we numerically verify the above interior solution and find that there is no transition in this example. The solution immediately reaches the stationary case in period 2.

Our solution above verifies part 1 of Proposition 5 in SSM (2017), which considers more general payoff functions and transition probabilities. The DM's choices exhibit inertia. That is, when the exogenous state is more persistent, the DM's choice behavior is also more persistent. For our example, they have the same persistence.

It is interesting to compare with the finite-horizon solution. SSM (2017) study the two-period case and their Proposition 4 shows that when γ is sufficiently small, $q_2(1|1) = q_2(2|2) = 1$ and $\Pr(a_1 = a_2) = 1$. That is, if the probability of changing states is sufficiently small, the DM's behavior exhibits status quo bias in the sense that he acquires information only in the first period and relies on that information in both periods. Through extensive numerical experiments, we find that this result does not hold in the infinite-horizon case. In particular, we always have the interior solution described above for any $\gamma \in (0,1)$ given $\mu_1(x_1 = 1) = 0.5$.[8]

[Insert Figure 3 Here.]

The intuition behind the above result is the following. In the two-period case, when γ is sufficiently small, the DM believes that any state in period 1 is more likely to remain the same in period 2. Thus the DM does not want to acquire new information and just follows the first period choice. However, when the horizon becomes longer, future states are more likely to switch. In particular, the switching probability is given by $1-(1-\gamma)^T$, which increases to 1 for $\gamma \in (0,1)$ as $T \to \infty$. Thus it is more valuable to acquire new information when the decision horizon is longer. But when the decision horizon is sufficiently short, the DM will not acquire any information, e.g., in the terminal period.

Figure 3 illustrates the analysis above. The parameter values are $T = 10$, $\mu_1(x_1 = 1) = 0.5$, $\lambda = 1$, $\gamma = 0.03$, and $\beta = 0.8$. We find that $q_1(a_1 = 1) = 1/2$ by symmetry and

$$q_t(a_t = 1|a_{t-1} = 1) = \begin{cases} 0.97 & \text{for } t = 2,3,...,6, \\ 0.973 & \text{for } t = 7, \\ 1 & \text{for } t = 8,9,10. \end{cases}$$

The left panel of Figure 2 presents the paths of $p_t(a_t = 1|x_t = 1, a_{t-1} = 1)$ and $p_t(a_t = 1|x_t = 1, a_{t-1} = 2)$. At time 1, they are the same because a_1 is not present. Then $p_t(a_t = 1|x_t = 1, a_{t-1} = 1)$ increases to 1 and $p_t(a_t = 1|x_t = 1, a_{t-1} = 2)$ decreases to zero, consistent with the inertia behavior shown in Proposition 5 of SSM (2017). We also find that, when $T \to \infty$, there is no terminal time and $q_t(a_t = 1|a_{t-1} = 1) = 0.97 = 1-\gamma$ for all $t \geq 2$.

Our analysis indicates that it is rational inattention combined with the short horizon that generates status quo bias. This bias does not exist under infinite horizon. The inertia behavior exists in both finite- and infinite-horizon settings.

5.2 Transition Kernel Depends on Actions

We now show that the results are very different when the state transition kernel depends on actions. For simplicity, assume that $\pi(x_{t+1}|x_t, a_t) = \alpha \in [0,1]$ if $x_{t+1} = a_t$, for $t \geq 1$. That is, the probability that the state in the next period confirms the current action is equal to α and is independent of the current state. We use this example to show that status quo bias can persist in the long run and confirmation bias and belief polarization can also arise.

Notice that the optimal solution in the case without information cost ($\lambda = 0$) is always to choose an action to match the state in each period as in the previous subsection. Next consider the two-period case with costly information acquisition ($\lambda > 0$).

[8]We have a full characterization of interior solutions under both finite and infinite horizons for general initial priors and any number of states and actions. The result is available upon request. Also see Proposition 9.

Proposition 8 *Consider the two-period RI model with* $\beta = 1$ *and* $u_3(x_3) = 0$*. Let* $\pi(x_{t+1}|x_t, a_t) = \alpha \in [0,1]$ *whenever* $x_{t+1} = a_t$*. Let* $\alpha^* \equiv \frac{\exp(1/\lambda)}{\exp(1/\lambda)+1}$ *and* $\alpha^{**} \equiv \frac{1}{\exp(1/\lambda)+1}$*. Then the solution satisfies* $q_1(1) = 1/2$ *and* $\Pr(a_2 = a_1) = 1$ *for* $\alpha > \alpha^*$ *and* $q_1(1) = 1/2$ *and* $\Pr(a_2 \neq a_1) = 1$ *for* $\alpha < \alpha^{**}$*. For* $\alpha \in (\alpha^*, \alpha^{**})$*, the solution is interior with*

$$q_2(1|1) = q_2(2|2) = \frac{\alpha(\exp(1/\lambda) + 1) - 1}{\exp(1/\lambda) - 1}.$$

This proposition shows that the status quo bias can emerge. In particular, if α is sufficiently large, the DM believes that there is a high probability that x_2 confirms a_1 and thus he does not reverse his decision. But if α is sufficiently small, he reverses his decision.

Using numerical methods, we find that the solutions for any $T > 2$ are similar to those for $T = 2$. This result is different from the case in which the state transition kernel is independent of actions. In that case the status quo bias does not occur under infinite horizon because the probability that the state will eventually switch is equal to 1. By contrast, for the model in this subsection, the state transition probability is independent of the current state, but dependent on the current action. If the probability that the state in the next period matches the current action is sufficiently high, the DM will not reverse his initial decision in that $\Pr(a_t = a_1) = 1$ for all $t > 1$. On the other hand, if this probability is sufficiently low, the DM will reverse his initial decision forever in that $\Pr(a_t = a_1') = 1$ for all $t > 1$ and $a_1' \neq a_1$.

[**Insert Figure 4 Here**]

Figure 4 illustrates the transition dynamics for the parameter values $\lambda = 1, \beta = 0.8$, and $\alpha = 0.9$. We find that there is no transition and the solution becomes stationary from the second period on.

There is a positive feedback between beliefs and actions in the model of this subsection. When the DM believes that the state in the next period is sufficiently likely to be consistent with the DM's current action, he will choose the same action in the next period in order to match the state. In this case he acquires information only in period 1 and uses the same information in the future. Even though the realized state in the future is different from his initial action, he still mistakenly sticks to the initial chosen action because processing new information is costly.

The model here also has implications for the confirmation bias and the belief polarization in the psychology literature. Confirmation bias is the tendency to search for, interpret, favor, and recall information in a way that confirms one's preexisting beliefs or hypotheses. This behavior happens in our model because the DM will stick to his initial choice if he entertains a strong belief that the future state is likely to be consistent with his current action. If there are more individuals, belief polarization may occur. Suppose that two individuals with the same prior about the states have different beliefs about state transition probabilities. One believes the future state is more likely to be consistent with the current action, and the other believes the opposite. Then after the same state is realized over time, each one believes his own belief is correct: individual 1 will choose the same action as the initial one. The other will always choose the action different from the initial one.

5.3 Dynamics of Consideration Sets

In this subsection we allow any finite number of actions and states and study how consideration sets evolve over time. Let $|X| = |A| = m > 2$. CDL (2018a) define the consideration set as the set of actions which are chosen with positive probabilities in a static setup. In our dynamic model the DM's choices depend on past information or actions. Since we focus on Markovian solutions, we define a consideration set at date t conditional on a_{t-1} as

$$B_t(a_{t-1}) = \{a_t \in A : q_t(a_t|a_{t-1}) > 0 \text{ for } \mu_t(a_{t-1}) > 0\}.$$

Notice that we require that a_{t-1} be chosen with a positive probability unconditionally in the last period for this definition to be well defined.

We first consider the case in which the transition kernel does not depend on actions. The following result characterizes the steady-state behavior and shows that the status quo bias in the sense that $\Pr(a_t = a_{t-1}) = 1$ does not emerge under infinite horizon.

Proposition 9 *For the infinite-horizon matching state problem under RI, let $\pi(x_{t+1}|x_t, a_t) = 1-\gamma$ if $x_{t+1} = x_t$; and $\pi(x_{t+1}|x_t, a_t) = \gamma/(m-1)$, otherwise. For any symmetric interior solution in the long run, $q(a_t|a_{t-1}) = 1-\gamma$ if $a_t = a_{t-1}$ and $q(a_t|a_{t-1}) = \gamma/(m-1)$ if $a_t \neq a_{t-1}$.*

We are unable to provide an analytical characterization of the transition dynamics. Thus we use the forward-backward Arimoto-Blahut algorithm to solve numerical examples. Set $m = 10$, $\mu_1(k) = \frac{1-\delta}{1-\delta^{10}}\delta^{k-1}$ for $1 \leq k \leq 10$, $\delta = 0.8$, $\beta = 0.5$, $\lambda = 1$, and $\gamma = 0.2$. According to this prior distribution, the DM believes state i is more likely than state j if $1 \leq i < j \leq m$. For the static case we are able to replicate the numerical results in CDL (2018a). Figure 5 presents the solution for the infinite-horizon case. We find that the solution converges to the steady state characterized in Proposition 9 starting from period 6 on.

[Insert Figure 5 Here]

In period 1 the DM chooses the first 5 actions/alternatives as they have high prior probabilities. Thus the consideration set in period 1 is $B_1 = \{1, 2, 3, 4, 5\}$. The DM makes 'mistakes' in that for each of the chosen alternatives $a_1 \in B_1$, the probability of it in fact matching the state is about $\mu_1(x_1 = a_1|a_1) = 0.40 < 1$. The posterior is identical for all $a_1 \in B_1$ despite the fact that states have different prior probabilities as shown in CDL (2018a). Moreover, the posterior distribution $\mu_1(\cdot|a_1')$ is a permutation of $\mu_1(\cdot|a_1)$ for $a_1', a_1 \in B_1$. For example, $\mu_1(1|1) = \mu_1(2|2) = 0.40$,

$\mu_1(1|2) = \mu_1(2|1)$, and $\mu_1(i|1) = \mu_1(i|2)$ for all $i = 3, 4, ..., 10$. This property holds true for any period and thus we only consider the history $a_t = 1$ in Figure 4.

Moving to period 2, states have a high probability (0.8) to remain the same. The new prior at history $a_1 = 1$ is the predictive distribution $\mu_2(x_2|a_1 = 1)$, which still puts a large weight on state 1, but the weights on other states are spread out. The posterior $\mu_2(x_2|a_2 = 1)$ has a similar property. Moreover, the DM chooses action 1 with a high probability given that he already chose $a_1 = 1$, $q_2(a_2 = 1|a_1 = 1) = 0.88$.

Similar pattern persists as time goes by. From period 2 on, the consideration set $B_t(a_{t-1} = 1)$ strictly increases over time until all alternatives are chosen in period 6. In the meantime, $q_t(a_t|a_{t-1} = 1)$ slowly declines to its steady state value of 0.80. The posterior distribution converges to the steady state: $\mu(1|1) = 0.33$, $\mu(i|1) = 0.074$ for all $i \neq 1$.

Next we consider the case in which the transition kernel depends on actions. For simplicity, let $\pi(x_{t+1}|x_t, a_t) = \alpha$ if $x_{t+1} = a_t$; and $\pi(x_{t+1}|x_t, a_t) = (1-\alpha)/(m-1)$, otherwise. We find that the results are very different from those in the case in which the transition kernel is independent of actions discussed earlier. But they are similar to those in Section 5.2.

We are unable to establish a theoretical result like Proposition 8. Through extensive numerical experiments, we find the following results for the infinite-horizon case. Due to the positive feedback between belief and behavior, when the DM believes that the state in the next period is sufficiently likely to be consistent with his current action (i.e., α is sufficiently large), he will choose the same action in the next period in order to match the state, i.e., $q_t(a_t|a_{t-1}) = 1$ for $a_t = a_{t-1}$. In this case he acquires information only in period 1 and uses the same information in the future. His behavior exhibits a status quo bias.

In the other extreme, if α is sufficiently small, the DM believes that the state is unlikely to match the action. He will not choose the same action in the next period, i.e., $q_t(a_t|a_{t-1}) = 0$ for $a_t = a_{t-1}$. Moreover he will choose any other action with an equal probability, i.e., $q_t(a_t|a_{t-1}) = 1/(m-1)$ for all $a_t \neq a_{t-1}$, due to the symmetry of the model specification.

For intermediate values of α, the solution is interior and characterized by the logit rule as in Proposition 6. For space consideration we omit the discussion here.

The results for the finite- and infinite-horizon cases are similar. Figure 6 presents the solutions for the three-period model with $T = 3$. We still use the previous parameter values except for different transition kernels. The top three panels show the results for $\alpha = 0.6$, which illustrate the status quo bias. The bottom three panels show the results for $\alpha = 0.02$. In this case the DM will not choose action 1 in period 2 even though he chooses it in period 1 with the highest probability. The reason is that he believes the state in period 2 will move away from state 1 with probability 0.98, and thus his revised prior belief $\mu_2(x_2|a_1)$ at history $a_1 = 1$ puts a small weight on state 1. The consideration set in period 2 conditional on $a_1 = 1$ is $B_2(a_1 = 1) = \{2, ..., 10\}$. Despite the fact that $q_2(a_2 = 1|a_1 = 1) = 0$, the unconditional probability of $a_2 = 1$ is $\mu_2(a_2 = 1) = 0.6$ so that we can compute $q_3(a_3|a_2 = 1)$. We find that the consideration set in period 3 conditional on $a_2 = 1$ is $B_3(a_2 = 1) = \{2, 3, ..., 10\}$.

[**Insert Figure 6 Here**]

6 Conclusion

We adopt the posterior-based approach to study dynamic RI problems and provide a transparent characterization of optimal solutions for the case with the Shannon entropy similar to the dynamic logit rule. We provide an efficient algorithm to solve the optimal solutions and apply our model to explain some behavioral biases. We also study the connection with the choice-based approach and show that, without checking the concavity of the objective function, first-order Kuhn-Tucker conditions can lead to nonoptimal solutions for dynamic RI problems. Unlike the choice-based approach, our approach applies to general UPS cost functions that satisfy a convexity condition on the difference between the discounted generalized entropy of the prior beliefs about the future states and the generalized entropy of the current posterior. Because this large class of cost functions can help explain some behavior that violates the predictions of the RI models with the Shannon entropy cost (CDL (2018b)), our approach should find wide applications in dynamic settings.

Online Appendix

A Proofs

Proof of Proposition 1: See MM (2015) and CDL (2018a). Q.E.D.

Proof of Lemma 1: See Theorem 4 of Blahut (1972). Q.E.D.

Proof of Proposition 2: (i) It follows from Corollary 1 of CD (2013).

(ii) By MM (2015) and CD (2013), the optimal posterior satisfies

$$\mu(x|a) = \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \text{ if } q(a) > 0.$$

We can then compute that

$$\begin{aligned}
\overline{V}(\mu) &= \sum_a q(a) N^a(\mu(\cdot|a)) = \sum_a q(a) \sum_x \mu(x|a)\left[u(x,a) - \lambda \ln \mu(x|a)\right] \\
&= \sum_a q(a) \sum_x \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} u(x,a) \\
&\quad -\lambda \sum_a q(a) \sum_x \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \ln \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \\
&= -\lambda \sum_a q(a) \sum_x \frac{\mu(x)\exp(u(x,a)/\lambda)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \ln \frac{\mu(x)}{\sum_{a'} q(a')\exp(u(x,a')/\lambda)} \\
&= \lambda \sum_x \mu(x) \ln\left[\sum_{a'} q(a')\exp\left(u(x,a')/\lambda\right)\right] - \lambda \sum_x \mu(x)\ln\mu(x).
\end{aligned}$$

Define $\widetilde{V}(x)$ as in the proposition for $\mu(x) > 0$. Then the optimal expected payoff is

$$V(\mu) = \overline{V}(\mu) + \lambda \sum_x \mu(x)\ln\mu(x) = \sum_x \mu(x)\widetilde{V}(x).$$

Let $\widehat{V}(x) \equiv \widetilde{V}(x) - \lambda \ln \mu(x)$. Then it satisfies

$$\overline{V}(\mu) = \sum_x \widehat{V}(x)\mu(x).$$

By CD (2013), $\mu(\cdot|a)$ is independent of the prior $\mu \in \Delta(X)$ in the convex hull of $\mu(\cdot|a)$ for all a such that $q(a) \in (0,1)$. Since $\widehat{V}(x)$ is the height of the hyperplane containing that convex hull at the point with $\mu(x) = 1$ and $\mu(x') = 0$ for all $x' \neq x$. This value is independent of the prior μ in the convex hull spanned by the optimal posteriors. Notice that if the convex hull is a singleton (i.e., $q(a) = 1$ for some a), then $\mu(x|a) = \mu(x)$ and $\widehat{V}(x)$ depends on μ.

(iii) Since $\overline{V}(\mu)$ is the concavification of the net utilities, it is concave. Next consider the problem in (5). Let $\theta \in (0,1)$, μ, $\mu' \in \Delta(X)$, and

$$\mu^* = \theta\mu + (1-\theta)\mu'.$$

Let $p^*(a|x)$ and $q^*(a)$ be the associated optimal solution. Then

$$q^*(a) = \sum_x p^*(a|x)\,\mu^*(x) = \theta q_1^*(a) + (1-\theta)\,q_2^*(a),$$

where

$$q_1^*(a) = \sum_x p^*(a|x)\,\mu(x), \quad q_2^*(a) = \sum_x p^*(a|x)\,\mu'(x).$$

Since the Shannon entropy is a concave function, we deduce that

$$\begin{aligned}
& V\left(\theta\mu + (1-\theta)\,\mu'\right) \\
= & \sum_{x,a} p^*(a|x)\,\mu^*(x)\left[u(x,a) - \lambda \ln \frac{p^*(a|x)}{q^*(a)}\right] \\
= & \sum_{x,a} p^*(a|x)\,\mu^*(x)\left[u(x,a) - \lambda \ln p^*(a|x)\right] - \lambda H(q^*) \\
\leq & \ \theta \sum_{x,a} p^*(a|x)\,\mu(x)\left[u(x,a) - \lambda \ln \frac{p^*(a|x)}{q_1^*(a)}\right] \\
& + (1-\theta)\sum_{x,a} p^*(a|x)\,\mu'(x)\left[u(x,a) - \lambda \ln \frac{p^*(a|x)}{q_2^*(a)}\right] \\
\leq & \ \theta V(\mu) + (1-\theta)\,V\left(\mu'\right).
\end{aligned}$$

Thus $V(\mu)$ is convex. It follows from (13) that

$$\begin{aligned}
V(\mu) = & \sum_{i=1}^{m-1} \mu(i)\left[\widehat{V}(i) + \lambda \ln \mu(i)\right] + \left[1 - \sum_{i=1}^{m-1}\mu(i)\right]\widehat{V}(m) \\
& + \lambda \left[1 - \sum_{i=1}^{m-1}\mu(i)\right] \ln \left[1 - \sum_{i=1}^{m-1}\mu(i)\right].
\end{aligned}$$

If μ is in the convex hull of the optimal posteriors for at least two chosen actions, then $\widehat{V}$ is independent of μ in that convex hull. We obtain

$$\frac{\partial V(\mu)}{\partial \mu(i)} = \left[\widehat{V}(i) - \lambda \ln \mu(i)\right] - \left[\widehat{V}(m) + \lambda \ln \mu(m)\right] = \widetilde{V}(i) - \widetilde{V}(m), \ i = 1, \ldots, m-1.$$

If $q(a) = 1$ for some action a, then $\mu(x|a) = \mu(x)$. Thus $V(\mu) = \sum_x \mu(x)\,u(x,a)$ and $\widetilde{V}(x) = u(x,a)$. The above formula still holds. CDL (2018a) show that the set $\Delta(X)$ can be partitioned into sets of priors, each of which is associated with a given consideration set. The derivative formula above applies to each set of priors and crosses boundaries of neighboring sets continuously. The formula for $\partial \overline{V}(\mu)/\partial \mu(i)$ follows from (12). Q.E.D.

Proof of Lemma 2: Define $\tilde{G}(\mu_1(\cdot|a_1), \tilde{\mu}_2(\cdot|a_1))$ as

$$\tilde{G}(\mu_1(\cdot|a_1), \tilde{\mu}_2(\cdot|a_1)) = \sum_{x_1,x_2} \pi(x_2|x_1,a_1)\mu_1(x_1|a_1) \ln \frac{\mu_1(x_1|a_1)}{[\tilde{\mu}_2(x_2|a_1)]^\beta}.$$

Therefore

$$G(\mu_1(\cdot|a_1)) = \tilde{G}\Big(\mu_1(\cdot|a_1), \sum_{x_1} \pi(\cdot|x_1, a_1)\mu_1(x_1|a_1)\Big).$$

Notice that $\tilde{G}(\mu_1(\cdot|a_1), \tilde{\mu}_2(\cdot|a_1))$ is a convex combination of $\mu_1(x_1|a_1) \ln \frac{\mu_1(x_1|a_1)}{[\tilde{\mu}(x_2|a_1)]^\beta}$ for all x_1 and x_2. The expression $\mu_1(x_1|a_1) \ln \frac{\mu_1(x_1|a_1)}{[\tilde{\mu}(x_2|a_1)]^\beta}$ is a jointly convex function of $\mu_1(x_1|a_1)$ and $\tilde{\mu}_2(x_2|a_1)$ for any $\beta \in (0, 1]$. Therefore, $\tilde{G}$ is jointly convex in $\mu_1(\cdot|a_1)$ and $\tilde{\mu}_2(\cdot|a_1)$.

For any $\theta \in [0, 1]$ and $\mu_1(\cdot|a_1)$, $\mu_1'(\cdot|a_1)$,

$$\begin{aligned}
&G\big(\theta\mu_1(\cdot|a_1) + (1-\theta)\mu_1'(\cdot|a_1)\big) \\
&= \tilde{G}\Big(\theta\mu_1(\cdot|a_1) + (1-\theta)\mu_1'(\cdot|a_1), \theta\sum_{x_1} \pi(\cdot|x_1, a_1)\mu_1(x_1|a_1) + (1-\theta)\sum_{x_1} \pi(\cdot|x_1, a_1)\mu_1'(x_1, a_1)\Big) \\
&\leq \theta\tilde{G}\Big(\mu_1(\cdot|a_1), \sum_{x_1} \pi(\cdot|x_1, a_1)\mu_1(x_1|a_1)\Big) + (1-\theta)\tilde{G}\Big(\mu_1'(\cdot|a_1), \sum_{x_1} \pi(\cdot|x_1, a_1)\mu_1'(\cdot|a_1)\Big) \\
&= \theta G(\mu_1(\cdot|a_1)) + (1-\theta)G(\mu_1'(\cdot|a_1)),
\end{aligned}$$

where the inequality follows from the definition of a jointly convex function. Q.E.D.

Proof of Proposition 3: We study Problem 1 with $T = \infty$ by dynamic programming. To distinguish between the finite- and infinite-horizon cases, we use W_t^T to denote the value function when the horizon is $T < \infty$. For any $t \geq 1$, define the infinite-horizon value function as

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \max_{\{p_k\}_{k=t}^\infty} \mathbb{E}\left[\sum_{k=t}^\infty \beta^{k-t}\left(u\left(x_k, a_k\right) - \lambda I\left(x^k; a_k|a^{k-1}\right)\right)\right], \tag{A.1}$$

where $p_k \in \Delta\left(A|X^k \times A^{k-1}\right)$ and the expectation is taken with respect to the distribution induced by μ_t, p_k, and the state transition kernel π for $k \geq t$. By the principle of optimality it satisfies the Bellman equation

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \max_{p_t} \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t|a^{t-1}\right) + \beta W_{t+1}\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right)$$

subject to the law of motion (1). We now study the solution to this Bellman equation, which is a non-stationary dynamic programming problem.

Define $\Delta\left(X^\infty \times A^\infty\right)$ as the space of all probability distributions on the set $X^\infty \times A^\infty$. Let this space be endowed with the weak convergence topology. Under this topology $\Delta\left(X^\infty \times A^\infty\right)$ is compact. Define $\mathbb{V}$ as the space of all sequences of bounded and continuous functions $f \equiv \{f_t\}_{t=1}^\infty$ that map $\Delta\left(X^\infty \times A^\infty\right)$ into the real line. Define the sup-norm on this space:

$$\|f\| = \sup_{t\geq 1} \sup_{\mu\in\Delta(X^\infty\times A^\infty)} f_t\left(\mu\right) < \infty, \ f \in \mathbb{V}.$$

Then $\mathbb{V}$ is a Banach space under this norm. Each marginal distribution $\mu_t\left(x^t, a^{t-1}\right) \in \Delta\left(X^t \times A^{t-1}\right)$ can be embedded in the space $\Delta\left(X^\infty \times A^\infty\right)$ by identifying $\mu_t\left(x^t, a^{t-1}\right)$ as $\mu_t\left(\left(x^t, a^{t-1}\right) \times X^\infty \times A^\infty\right)$. Define an operator $\mathcal{T}$ on $\mathbb{V}$ as follows:

$$\begin{aligned} (\mathcal{T}f)_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) &= \max_{p_t} \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t | a^{t-1}\right) \\ &\quad + \beta f_{t+1}\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right) \end{aligned} \tag{A.2}$$

for $t \geq 1$, subject to (1), where $\{f_t\}_{t=1}^\infty \in \mathbb{V}$. Then we can verify that $\mathcal{T}f = \{(\mathcal{T}f)_t\}_{t=1}^\infty \in \mathbb{V}$. Moreover, $\mathcal{T}$ satisfies the Blackwell sufficient conditions for $\beta \in (0,1)$ and thus it is a contraction mapping. By the contraction mapping theorem, $\mathcal{T}$ has a unique fixed point $\{W_t\}_{t=1}^\infty$ satisfying the Bellman equation:

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \max_{p_t} \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t | a^{t-1}\right) + \beta W_{t+1}\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right),$$

subject to (1).

Now consider the limit of the finite-horizon problem. Define a sequence of functions $\left\{W_t^T\right\}_{t=1}^\infty \in \mathbb{V}$, where $\left\{W_t^T\right\}_{t=1}^{T+1}$ satisfies (35) and (36) and $W_t^T = 0$ for $t > T+1$. Since $\mathcal{T}$ is a contraction mapping, we know that

$$\lim_{k\to\infty} \mathcal{T}^k \bar{W} = W, \tag{A.3}$$

for any $\bar{W} \in \mathbb{V}$. Let $\bar{W}\left(\mu\left(x^\infty, a^\infty\right)\right) = \mathbb{E}u_{T+1}\left(x_{T+1}\right)$ for any bounded u_{T+1} and any $t \geq 1$ and let $k = T + 1 - t$. Then using (35), (36), and (A.2) we can verify that

$$\left(\mathcal{T}^k \bar{W}\right)_t = W_t^T. \tag{A.4}$$

As $T \to \infty$, we have $k \to \infty$. It follows from (A.3) that

$$\lim_{k\to\infty} \left(\mathcal{T}^k \bar{W}\right)_t = \lim_{T\to\infty} W_t^T = W_t.$$

Equation (A.4) is the same as the finite-horizon problem in (35) and (36).

It follows from TSS (2018) that there exists an optimal plan of the form $\mathbf{p}^T = \{p_t^T\left(a_t | x_t, a^{t-1}\right)\}_{t=1}^T$ for any finite T-horizon RI problem. Since the space $\prod_{t=1}^\infty \Delta\left(A | X \times A^{t-1}\right)$ is compact, there is a subsequence such that $\mathbf{p}^{T_k}$ converges to a limit $\hat{\mathbf{p}} = \{\hat{p}_t\left(a_t | x_t, a^{t-1}\right)\}_{t=1}^\infty$. Since $\mathbf{p}^T$ is optimal, we have

$$W_t^T\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t | a^{t-1}\right) + \beta W_{t+1}^T\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right),$$

where the expectation is taken with respect to the distribution $\mu_t\left(x^t, a^t\right)$ induced by $\mathbf{p}^T$ and the state transition kernel and

$$\mu_{t+1}\left(x^{t+1}, a^t\right) = \pi\left(x_{t+1} | x_t, a_t\right) p_t^T\left(a_t | x^t, a^{t-1}\right) \mu_t\left(x^t, a^{t-1}\right). \tag{A.5}$$

Taking limit as $T_k \to \infty$, we have

$$W_t\left(\left\{\mu_t\left(x^t, a^{t-1}\right)\right\}\right) = \mathbb{E}\left[u\left(x_t, a_t\right)\right] - \lambda I\left(x^t; a_t|a^{t-1}\right) + \beta W_{t+1}\left(\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}\right),$$

where the expectation is taken with respect to the distribution induced by $\hat{\mathbf{p}}$ and the state transition kernel and

$$\mu_{t+1}\left(x^{t+1}, a^t\right) = \pi\left(x_{t+1}|x_t, a_t\right) \hat{p}_t\left(a_t|x_t, a^{t-1}\right) \mu_t\left(x^t, a^{t-1}\right). \tag{A.6}$$

Thus the choice rule $\hat{\mathbf{p}} = \{\hat{p}_t\left(a_t|x_t, a^{t-1}\right)\}_{t=1}^{\infty}$ is optimal for the infinite-horizon RI problem.

The proof of part (i) is completed. The proofs of the other two parts follow from TSS (2018) closely and is omitted here. Q.E.D.

Proof of Propositions 4: By the definition of the Markovian solution, we can compute that

$$\begin{aligned}\mu_{t+1}(x_{t+1}|a_t) =& \frac{\mu_{t+1}(x_{t+1}, a_t)}{\mu_{t+1}(a_t)} = \frac{\sum_{a^{t-1}} \mu_{t+1}(x_{t+1}, a_t, a^{t-1})}{\sum_{a^{t-1}} \mu_{t+1}(a_t, a^{t-1})} \\ =& \frac{\sum_{a^{t-1}} \mu_{t+1}(x_{t+1}|a_t, a^{t-1})\mu_{t+1}(a_t, a^{t-1})}{\sum_{a^{t-1}} \mu_{t+1}(a_t, a^{t-1})} = \mu_{t+1}(x_{t+1}|a_t, a^{t-1}).\end{aligned} \tag{A.7}$$

Therefore, we can replace the history dependent predictive distribution $\mu_{t+1}(\cdot|a_t, a^{t-1})$ by the history independent predictive distribution $\mu_{t+1}(\cdot|a_t)$.

By LIP property of CD (2013), the optimal posterior $\mu_{t+1}\left(x_{t+1}|a^{t+1}\right)$ at history a^t is independent of the priors $\mu_{t+1}\left(x_{t+1}|a^t\right)$ in the convex hull of the optimal posteriors associated with at least two chosen actions a_{t+1}. Thus $\mu_{t+1}\left(x_{t+1}|a^{t+1}\right)$ is independent of a^t. By a similar computation to (A.7), we obtain $\mu_{t+1}\left(x_{t+1}|a^{t+1}\right) = \mu_{t+1}\left(x_{t+1}|a_{t+1}\right)$. If there is only one chosen action a_{t+1}, then the optimal posterior $\mu_{t+1}\left(x_{t+1}|a^{t+1}\right)$ is the same as the prior at a^t, $\mu_{t+1}\left(x_{t+1}|a^t\right)$, which is equal to $\mu_{t+1}\left(x_{t+1}|a_t\right)$ by (A.7). For both cases we have $\mu_{t+1}\left(x_{t+1}|a^{t+1}\right) = \mu_{t+1}\left(x_{t+1}|a_{t+1}, a_t\right)$ using a similar computation in (A.7).

The predictive distribution has the decomposition

$$\begin{aligned}\mu_{t+1}(x_{t+1}|a_t) =& \mu_{t+1}(x_{t+1}|a_t, a^{t-1}) = \sum_{a_{t+1}} q_{t+1}(a_{t+1}|a_t, a^{t-1})\mu_{t+1}(x_{t+1}|a_{t+1}, a^t) \\ =& \sum_{a_{t+1}} q_{t+1}(a_{t+1}|a_t, a^{t-1})\mu_{t+1}(x_{t+1}|a_{t+1}, a_t).\end{aligned}$$

Therefore $q(a_{t+1}|a_t, a^{t-1})$ is independent of a^{t-1} and can be replaced by $q_{t+1}(a_{t+1}|a_t)$. Finally,

$$\begin{aligned}p_{t+1}(a_{t+1}|x_{t+1}, a^t) &= \frac{\mu_{t+1}(x_{t+1}|a^{t+1})q_{t+1}(a_{t+1}|a^t)}{\mu_{t+1}(x_{t+1}|a^t)} \\ &= \frac{\mu_{t+1}(x_{t+1}|a_{t+1}, a_t)q_{t+1}(a_{t+1}|a_t)}{\mu_{t+1}(x_{t+1}|a_t)} = p_{t+1}(a_{t+1}|x_{t+1}, a_t),\end{aligned}$$

for any $\mu_{t+1}(x_{t+1}|a^t) > 0$. Replacing the history dependent choice rule and default rule by their history independent version does not change the value of the dynamic RI problem. Therefore the optimal choice rule and default rule are both history independent. Q.E.D.

Proof of Proposition 5: The first period predictive distribution is the prior μ_1. The second period predictive distribution is $\mu_2(\cdot|a_1)$. Because the solution is interior, $q_2(a_2|a_1) > 0$ for any $a_2 \in A$. Then all predictive distributions $\mu_2(\cdot|a_1)$ for different a_1 are in the interior of the convex hull spanned by optimal posteriors $\mu_2(\cdot|a^2)$. By the LIP property, $\mu_2(\cdot|a^2)$ takes the form $\mu_2(\cdot|a_2)$. The third period predictive distribution is determined by

$$\mu_3(x_3|a_2) = \sum_{x_2} \pi(x_3|x_2, a_2)\mu_2(x_2|a_2),$$

which does not depend on a_1. We can show that $\mu_{t+1}\left(x_{t+1}|a^t\right)$ takes the form of $\mu_{t+1}\left(x_{t+1}|a_t\right)$ using the same argument by induction. Thus an interior solution is Markovian. Q.E.D.

Proof of Propositions 6 and 7: When the solution is Markovian, we verify the optimality conditions for history-dependent solution in Proposition 10 of Appendix B are equivalent to their history-independent analogue stated in Propositions 6 and 7.

For a Markovian solution, it follows from (A.7) that $\mu_t(x_t|a^{t-1}) = \mu_t(x_t|a_{t-1})$ for any history a^{t-1} such that $\mu_t(a^{t-1}) > 0$. Proposition 4 implies that $p_t(a_t|x_t, a^{t-1}) = p_t(a_t|x_t, a_{t-1})$ and $q_t(a_t|a^{t-1}) = q_t(a_t|a_{t-1})$. Therefore $v_t(x_t, a^t)$ in (B.4) takes the form of $v_t(x_t, a_t)$ in (43) and $\widetilde{V}_t(x_t, a^{t-1})$ in (B.5) takes the form of $\widetilde{V}_t(x_t, a_{t-1})$ in (44). Therefore (B.1) and (B.2) are equivalent to (41) and (42). It remains to verify (45) and (46), we start from the history-dependent law of motion in (B.6). The left-hand side can be transformed into

$$\mu_{t+1}(x_{t+1}, a^t) = \mu_{t+1}(x_{t+1}|a^t)\, q_t(a_t|a^{t-1})\, \mu_t(a^{t-1}) = \mu_{t+1}(x_{t+1}|a_t)\, q_t(a_t|a_{t-1})\, \mu_t(a^{t-1}),$$

where the second equality follows from replacing the history-dependent predictive distribution and the default rule by their history-independent versions. On the right-hand side of (B.6),

$$p_t(a_t|x_t, a^{t-1})\, \mu_t(x_t, a^{t-1}) = p_t(a_t|x_t, a^{t-1})\, \mu_t(x_t|a^{t-1})\, \mu_t(a^{t-1}) = p_t(a_t|x_t, a_{t-1})\, \mu_t(x_t|a_{t-1})\, \mu_t(a^{t-1}).$$

Combining the preceeding two equations with (B.6), we obtain

$$\mu_{t+1}(x_{t+1}|a_t)q_t(a_t|a_{t-1})\mu_t(a^{t-1}) = \mu_t(a^{t-1}) \sum_{x_t} \pi(x_{t+1}|x_t, a_t)p_t(a_t|x_t, a_{t-1})\mu_t(x_t|a_{t-1}).$$

Because $\mu_t(a^{t-1}) > 0$ and $\mu_t(a_{t-1}) > 0$, multiplying both sides by $\frac{\mu_t(a_{t-1})}{\mu_t(a^{t-1})}$ yields

$$\mu_{t+1}(x_{t+1}|a_t)\mu_t(a_t, a_{t-1}) = \sum_{x_t} \pi(x_{t+1}|x_t, a_t)p_t(a_t|x_t, a_{t-1})\mu_t(x_t, a_{t-1}).$$

Summing over a_{t-1} on both sides, we confirm (45) and (46). Q.E.D.

Proof of Proposition 8: There are two types of solutions. By symmetry of the problem, we first solve for a symmetric interior solution satisfying $q_1(a_1 = 1) = 1/2$ and $q_2(1|1) = q_2(2|2) = z$. By Proposition 6, we compute

$$\begin{aligned}\widetilde{V}_2(1,1) &= \widetilde{V}_2(2,2) = \lambda \ln[z \exp(1/\lambda) + 1 - z], \\ \widetilde{V}_2(1,2) &= \widetilde{V}_2(2,1) = \lambda \ln[(1-z)\exp(1/\lambda) + z],\end{aligned}$$

$$v_1(1,1) = v_1(2,2) = 1 + \beta\alpha\widetilde{V}_2(1,1) + \beta(1-\alpha)\widetilde{V}_2(2,1),$$

$$v_1(1,2) = v_1(2,1) = \beta\alpha\widetilde{V}_2(2,2) + \beta(1-\alpha)\widetilde{V}_2(1,2).$$

It follows from $\mu_1(1) = 1/2$ that the DM's initial value is given by

$$\begin{aligned}V_1 &= \frac{1}{2}\lambda \ln \frac{1}{2}[\exp(v_1(1,1)/\lambda) + \exp(v_1(1,2)/\lambda)] \\ &\quad + \frac{1}{2}\lambda \ln \frac{1}{2}[\exp(v_1(2,1)/\lambda) + \exp(v_1(2,2)/\lambda)] \\ &= \lambda \ln \frac{1}{2}[\exp(v_1(1,1)/\lambda) + \exp(v_1(1,2)/\lambda)].\end{aligned}$$

Thus maximizing V_1 is equivalent to maximizing

$$\left(ze^{\frac{1}{\lambda}} + 1 - z\right)^{\beta\alpha}\left[(1-z)e^{\frac{1}{\lambda}} + z\right]^{\beta(1-\alpha)}.$$

This is a concave function of z. The first-order condition gives

$$z = \frac{\alpha(\exp(1/\lambda) + 1) - 1}{\exp(1/\lambda) - 1}.$$

Thus, if

$$\alpha^{**} \equiv \frac{1}{\exp(1/\lambda) + 1} < \alpha < \frac{\exp(1/\lambda)}{\exp(1/\lambda) + 1} \equiv \alpha^*,$$

then the optimal solution is interior $z \in (0,1)$. If $\alpha \geq \alpha^*$, the solution is at the corner $z = 1$. If $\alpha \in [0, \alpha^{**}]$, the solution is at the other corner $z = 0$. We then obtain the desired result.

It remains to show that the corner solution in which $q_1(1) = 1$ is not optimal. Then we use $q_1(1) = 0$ and Proposition 6 to derive

$$V_1 = \frac{1}{2}v_1(1,1) + \frac{1}{2}v_1(2,1),$$

where $v_1(1,1)$ and $v_1(2,1)$ are given above. Since $\exp(x/\lambda)$ is a convex function x, we obtain that

$$\frac{1}{2}v_1(1,1) + \frac{1}{2}v_1(2,1) < \lambda \ln \frac{1}{2}[\exp(v_1(1,1)/\lambda) + \exp(v_1(2,1)/\lambda)].$$

Since $v_1(2,1) = v_1(1,2)$ for the above symmetric interior solution, we deduce that the corner solution gives a smaller initial value than the above symmetric interior solution. Similarly the other corner solution in which $q_1(2) = 0$ is not optimal. Q.E.D.

Proof of Proposition 9: By (38) and (39), we have the two-way decomposition of the predictive distribution for the long run stationary solution:

$$\mu\left(x_t = 1|a_{t-1}\right) = \left(1-\gamma\right)\mu\left(1|a_{t-1}\right) + \sum_{i=2}^{m}\frac{\gamma}{m-1}\mu\left(i|a_{t-1}\right) = \sum_{i=1}^{m} q\left(i|a_{t-1}\right)\mu\left(1|a_t = i\right).$$

By symmetry $\mu\left(i|i\right)$ is the same for all i, and $\mu\left(i|j\right)$ is the same for all $i \neq j$. We thus obtain the desired result. Q.E.D.

B Dynamic Logit Rule

In this appendix we present the characterization of history-dependent solutions, which is used in the proof of Propositions 6 and 7.

Proposition 10 *Let $\beta \in (0,1)$. (i) The choice rule $\{p_t(a_t|x_t, a^{t-1})\}_{t=1}^{T}$ and the default rule $\{q_t(a_t|a^{t-1})\}_{t=1}^{T}$ are the optimal (history-dependent) solution to the dynamic RI problem in (4) if and only if they satisfy the following system of difference equations for $t = 1, 2, \ldots, T$:*

$$p_t\left(a_t|x_t, a^{t-1}\right) = \frac{q_t\left(a_t|a^{t-1}\right)\exp\left(v_t\left(x_t, a^t\right)/\lambda\right)}{\sum_{a_t'} q_t\left(a_t'|a^{t-1}\right)\exp\left(v_t\left(x_t, a_t', a^{t-1}\right)/\lambda\right)} \text{ for } \mu_t\left(x_t, a^{t-1}\right) > 0, \tag{B.1}$$

$$q_t\left(a_t|a^{t-1}\right) = \sum_{x_t} p_t\left(a_t|x_t, a^{t-1}\right)\mu_t\left(x_t|a^{t-1}\right) \text{ for } \mu_t\left(a^{t-1}\right) > 0, \tag{B.2}$$

and the following condition is satisfied

$$\sum_{x_t}\frac{\mu_t\left(x_t|a^{t-1}\right)\exp\left(v_t\left(x_t, a^t\right)/\lambda\right)}{\sum_{a_t'} q_t\left(a_t'|a^{t-1}\right)\exp\left(v_t\left(x_t, a_t', a^{t-1}\right)/\lambda\right)} \leq 1, \quad t = 1, 2, ..., T, \tag{B.3}$$

with equality if $q_t\left(a_t|a^{t-1}\right) > 0$. In (B.1), (B.2), *and* (B.3),

$$v_t\left(x_t, a^t\right) = u\left(x_t, a_t\right) + \beta\sum_{x_{t+1}}\pi\left(x_{t+1}|x_t, a_t\right)\widetilde{V}_{t+1}\left(x_{t+1}, a^t\right), \tag{B.4}$$

$$\widetilde{V}_t\left(x_t, a^{t-1}\right) = \lambda\ln\sum_{a_t} q_t\left(a_t|a^{t-1}\right)\exp\left(v_t\left(x_t, a^t\right)/\lambda\right), \tag{B.5}$$

$$\mu_{t+1}\left(x_{t+1}, a^t\right) = \sum_{x_t}\pi\left(x_{t+1}|x_t, a_t\right)p_t\left(a_t|x_t, a^{t-1}\right)\mu_t\left(x_t, a^{t-1}\right), \tag{B.6}$$

$$\mu_t\left(x_t|a^{t-1}\right) = \frac{\mu_t\left(x_t, a^{t-1}\right)}{\mu_t\left(a^{t-1}\right)} \text{ for } \mu_t\left(a^{t-1}\right) = \sum_{x_t}\mu_t\left(x_t, a^{t-1}\right) > 0. \tag{B.7}$$

(ii) The value functions satisfy

$$W_t\left(\left\{\mu_t\left(x_t, a^{t-1}\right)\right\}\right) = \sum_{x_t, a^{t-1}}\mu_t\left(x_t, a^{t-1}\right)\widetilde{V}_t\left(x_t, a^{t-1}\right), \tag{B.8}$$

$$V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right) = \sum_{x_t}\mu_t\left(x_t|a^{t-1}\right)\widetilde{V}_t\left(x_t, a^{t-1}\right). \tag{B.9}$$

In the finite-horizon case with $T < \infty$, there is a terminal condition $\widetilde{V}_{T+1}\left(x_{T+1}, a^T\right) = u_{T+1}\left(x_{T+1}\right)$.

Proof: We start with the finite-horizon case with $T < \infty$.

Step 1. In the last period at history a^{T-1}, define value function as

$$V_T\left(\mu_T\left(\cdot|a^{T-1}\right)\right) \equiv \max \sum_{x_T, a_T} q_T\left(a_T|a^{T-1}\right)\mu_T\left(x_T|a^T\right)\left[v_T\left(x_T, a_T\right) - \lambda \ln \frac{\mu_T\left(x_T|a^T\right)}{\mu_T\left(x_T|a^{T-1}\right)}\right] \tag{B.10}$$

subject to

$$\mu_T\left(x_T|a^{T-1}\right) = \sum_{a_T} q_T\left(a_T|a^{T-1}\right)\mu_T\left(x_T|a^T\right), \tag{B.11}$$

for all $x_T \in X$ such that $\mu_T\left(x_T|a^{T-1}\right) > 0$, where

$$v_T\left(x_T, a_T\right) = u\left(x_T, a_T\right) + \beta \sum_{x_{T+1}} \pi\left(x_{T+1}|x_T, a_T\right) u_{T+1}\left(x_{T+1}\right).$$

The choice variables are $q_T\left(\cdot|a^{T-1}\right) \in \Delta(A)$ and $\mu_T\left(\cdot|a_T, a^{T-1}\right) \in \Delta(X)$ for all $a_T \in A$. Viewing $\mu_T\left(\cdot|a^{T-1}\right)$ as the prior at history a^{T-1}, this problem is the same as the static RI problem studied in Section 3.1.

From Proposition 2, we have the solution

$$V_T\left(\mu_T\left(\cdot|a^{T-1}\right)\right) = \overline{V}_T(\mu_T(\cdot|a^{T-1})) - \lambda H(\mu_T(\cdot|a^{T-1})),$$

where

$$\begin{aligned} \overline{V}_T(\mu_T(\cdot|a^{T-1})) &= \sum_{x_T} \mu_T(x_T|a^{T-1})\widehat{V}_T\left(x_T, a^{T-1}\right), \\ \widehat{V}_T\left(x_T, a^{T-1}\right) &= \widetilde{V}_T\left(x_T, a^{T-1}\right) - \lambda \ln \mu_T(x_T|a^{T-1}). \end{aligned}$$

Here $\widetilde{V}_T\left(x_T, a^{T-1}\right)$ satisfies (B.5) for $t = T$. Equations (B.8) and (B.9) for $t = T$ also follow from Proposition 2.

Step 2. Now the decision problem at date $T-1$ at history a^{T-2} is given by

$$\begin{aligned} & V_{T-1}\left(\mu_{T-1}\left(\cdot|a^{T-2}\right)\right) \\ = \ & \max \sum_{x_{T-1}, a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right)\mu_{T-1}\left(x_{T-1}|a^{T-1}\right)\left[u\left(x_{T-1}, a_{T-1}\right) - \lambda \ln \frac{\mu_{T-1}\left(x_{T-1}|a^{T-1}\right)}{\mu_{T-1}\left(x_{T-1}|a^{T-2}\right)}\right] \\ & +\beta \sum_{a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right) V_T\left(\mu_T\left(\cdot|a^{T-1}\right)\right) \end{aligned}$$

subject to

$$\mu_{T-1}\left(x_{T-1}|a^{T-2}\right) = \sum_{a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right)\mu_{T-1}\left(x_{T-1}|a^{T-1}\right), \tag{B.12}$$

$$\mu_T\left(x_T|a^{T-1}\right) = \sum_{x_{T-1}} \pi\left(x_T|x_{T-1}, a_{T-1}\right)\mu_{T-1}\left(x_{T-1}|a^{T-1}\right), \tag{B.13}$$

for all $x_{T-1} \in X$ such that $\mu_{T-1}\left(x_{T-1}|a^{T-2}\right) > 0$. The choice variables are $q_{T-1}\left(a_{T-1}|a^{T-2}\right)$ and $\mu_{T-1}\left(x_{T-1}|a^{T-1}\right)$.

We rewrite the objective function in the problem above as

$$
\begin{aligned}
& \sum_{x_{T-1},a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right)\mu_{T-1}\left(x_{T-1}|a^{T-1}\right)\left[u\left(x_{T-1},a_{T-1}\right) - \lambda \ln \frac{\mu_{T-1}\left(x_{T-1}|a^{T-1}\right)}{\mu_{T-1}\left(x_{T-1}|a^{T-2}\right)}\right] \\
& +\beta \sum_{a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right) V_T\left(\mu_T\left(x_T|a^{T-1}\right)\right) \\
= & \sum_{a_{T-1}} q_{T-1}\left(a_{T-1}|a^{T-2}\right) N_G^{a_{T-1}}\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right) - \lambda H\left(\mu_{T-1}\left(\cdot|a^{T-2}\right)\right),
\end{aligned}
$$

where we define the net utility as

$$
N_G^{a_{T-1}}\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right) \equiv \sum_{x_{T-1}} \mu_{T-1}\left(x_{T-1}|a^{T-1}\right) u\left(x_{T-1},a_{T-1}\right) + \beta \overline{V}_T(\mu_T(\cdot|a^{T-1})) - \lambda G\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right).
$$

Here the cost function G is

$$
G\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right) \equiv \beta H\left(\mu_T\left(\cdot|a^{T-1}\right)\right) - H\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right).
$$

It follows from Lemma 2 that G is convex in $\mu_{T-1}(\cdot|a^{T-1})$. Moreover, Proposition 2 shows that $\overline{V}_T$ is concave in $\mu_T(\cdot|a^{T-1})$, and hence concave in $\mu_{T-1}(\cdot|a^{T-1})$ by (B.13). Therefore the net utility $N_G^{a_{T-1}}$ is concave in $\mu_{T-1}(\cdot|a^{T-1})$. We view the problem at $T-1$ as a static RI problem with the prior belief $\mu_{T-1}\left(\cdot|a^{T-2}\right)$. We can apply Lemmas 1 through 3 of CD (2013) to derive the following result:

Lemma 3 *Given the prior $\mu_{T-1}\left(\cdot|a^{T-2}\right) > 0$ at history a^{T-2}, the posterior $\left\{\mu_{T-1}(x_{T-1}|a^{T-1})\right\}$ and the default rule $\left\{q_{T-1}\left(a_{T-1}|a^{T-2}\right)\right\}$ are optimal for the RI problem in period $T-1$ at history a^{T-2} if and only if (i) (B.12) holds; (ii) for any $a_{T-1}, b_{T-1} \in A$ such that $q_{T-1}\left(a_{T-1}|a^{T-2}\right) > 0$ and $q_{T-1}\left(b_{T-1}|a^{T-2}\right) > 0$, and for any $x_{T-1} \in X$,*

$$
\frac{\mu_{T-1}(x_{T-1}|a^{T-1})}{\exp(v_{T-1}(x_{T-1},a^{T-1})/\lambda)} = \frac{\mu_{T-1}(x_{T-1}|b_{T-1},a^{T-2})}{\exp(v_{T-1}(x_{T-1},b_{T-1},a^{T-2})/\lambda)}, \tag{B.14}
$$

where $v_{T-1}(x_{T-1},a^{T-1})$ for all chosen actions $a_{T-1} \in A$ satisfies (B.4) for $t = T$; and (iii) for any a_{T-1} such that $q_{T-1}\left(a_{T-1}|a^{T-2}\right) > 0$ and b_{T-1} such that $q_{T-1}\left(b_{T-1}|a^{T-2}\right) = 0$,

$$
\sum_{x_{T-1}} \frac{\mu_{T-1}(x_{T-1}|a^{T-1})\exp(v_{T-1}(x_{T-1},b_{T-1},a^{T-2})/\lambda)}{\exp(v_{T-1}(x_{T-1},a^{T-1})/\lambda)} \leq 1, \tag{B.15}
$$

where $v_{T-1}(x_{T-1},b_{T-1},a^{T-2}) = u\left(x_{T-1},b_{T-1}\right)$.

We only sketch the key step of the proof, which involves computing the slope of the net utility function. Consider derivatives with respect to $\mu_{T-1}\left(1|a^{T-1}\right)$ for illustration. We compute

$$
\begin{aligned}
\frac{\partial G\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right)}{\partial \mu_{T-1}\left(1|a^{T-1}\right)} &= \ln \mu_{T-1}\left(1|a^{T-1}\right) - \ln \mu_{T-1}\left(m|a^{T-1}\right) \\
&\quad -\beta \sum_{x_T} \pi\left(x_T|1, a_{T-1}\right) \ln \mu_T\left(x_T|a^{T-1}\right) \\
&\quad +\beta \sum_{x_T} \pi\left(x_T|m, a_{T-1}\right) \ln \mu_T\left(x_T|a^{T-1}\right).
\end{aligned}
$$

Given $\lambda > 0$ and $\beta \in (0,1)$, it follows from the equation above that the partial derivative approaches $-\infty$ when $\mu_{T-1}(1|a^{T-1})$ approaches 0, holding $\mu_{T-1}(i|a^{T-1})$ fixed for $i = 2, \ldots, m-1$. Therefore, the optimal posterior $\mu_{T-1}(x_{T-1}|a^{T-1}) \in (0,1)$ for all $x_{T-1} \in X$, whenever $q_{T-1}\left(a_{T-1}|a^{T-2}\right) > 0$ and $q_{T-1}\left(b_{T-1}|a^{T-2}\right) > 0$.

By Proposition 2, we can compute

$$
\begin{aligned}
& \frac{\partial N_G^{a_{T-1}}\left(\mu_{T-1}\left(\cdot|a^{T-1}\right)\right)}{\partial \mu_{T-1}\left(1|a^{T-1}\right)} \\
= \; & \left[u(1, a_{T-1}) - \lambda \ln \mu_{T-1}(1|a^{T-1})\right] - \left[u(m, a_{T-1}) - \lambda \ln \mu_{T-1}(m|a^{T-1})\right] \\
& +\beta \sum_{x_T} \pi(x_T|1, a_{T-1}) \left[\widehat{V}_T(x_T, a^{T-1}) + \lambda \ln \mu_T(x_T|a^{T-1})\right] \\
& -\beta \sum_{x_T} \pi(x_T|m, a_{T-1}) \left[\widehat{V}_T(x_T, a^{T-1}) + \lambda \ln \mu_T(x_T|a^{T-1})\right] \\
= \; & \left[u(1, a_{T-1}) - \lambda \ln \mu_{T-1}(1|a^{T-1})\right] - \left[u(m, a_{T-1}) - \lambda \ln \mu_{T-1}(m|a^{T-1})\right] \\
& +\beta \sum_{x_T} \pi(x_T|1, a_{T-1}) \widetilde{V}_T(x_T, a^{T-1}) - \beta \sum_{x_T} \pi(x_T|m, a_{T-1}) \widetilde{V}_T(x_T, a^{T-1}) \\
= \; & \left[v_{T-1}(1, a^{T-1}) - \lambda \ln \mu_{T-1}(1|a^{T-1})\right] - \left[v_{T-1}(m, a^{T-1}) - \lambda \ln \mu_{T-1}(m|a^{T-1})\right],
\end{aligned}
$$

where $v_{T-1}\left(x_{T-1}, a^{T-1}\right)$ is defined in (B.4).

CD (2013) show that a necessary condition for optimality is that the slope of the net utility function is the same for each chosen action a_{T-1} with $q_{T-1}\left(a_{T-1}|a^{T-2}\right) > 0$ at its associated posterior. Moreover, the equality holds state by state. We then obtain (B.14). By Proposition 2 of CDL (2018a), the conditions in Lemma 3 are equivalent to the necessary and sufficient conditions for the optimality of $p_{T-1}(a_{T-1}|x_{T-1}, a^{T-2})$ and $q_{T-1}(a_{T-1}|a^{T-2})$, stated in Proposition 10.

Step 3. We continue this process by backward induction until the initial period $t = 1$, completing the proof of the conditions for the optimality of history-dependent choice rules $p_t(a_t|x_t, a^{t-1})$ and default rules $q_t(a_t|a^{t-1})$, $t = 1, \ldots, T$, for any finite horizon $T < \infty$.

Step 4. As $T \to \infty$, the posterior-based value function at any time t, $V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right)$, converges to the corresponding value function in the infinite-horizon case by the method of discounted dynamic programming as in the proof of Proposition 3. Similarly, $W_t\left(\left\{\mu_t\left(x_t, a^{t-1}\right)\right\}\right)$ also converges to the value function in the infinite-horizon case as $T \to \infty$. The limit points of the solution for the finite-horizon case are the solution for the infinite-horizon RI problem. These limiting choice rule and default rule satisfy their associated optimality conditions.

C Choice-based Approach

In this appendix we derive conditions as in Propositions 6 and 7 using the choice-based approach. We consider interior solutions in the finite-horizon case and sketch the key steps only. We focus on choice rules and default rules of the form $p_t\left(a_t|x_t, a_{t-1}\right)$ and $q_t\left(a_t|a_{t-1}\right)$.

The value function for the dynamic RI problem satisfies

$$W_t\left(\left\{\mu_t\left(x_t, a_{t-1}\right)\right\}\right) = \sum_{x_t, a_{t-1}} \mu_t\left(x_t, a_{t-1}\right) \widetilde{V}_t\left(x_t, a_{t-1}\right) = \sum_{x_t, a_{t-1}} \mu_t\left(a_{t-1}\right) \mu_t\left(x_t|a_{t-1}\right) \widetilde{V}_t\left(x_t, a_{t-1}\right),$$

where $\widetilde{V}_t\left(x_t, a_{t-1}\right)$ denotes the realized value function for $\mu_t\left(x_t, a_{t-1}\right) > 0$. Define the choice-based value function at history a_{t-1} as

$$\widetilde{W}_t\left(a_{t-1}\right) = \sum_{x_t} \mu_t\left(x_t|a_{t-1}\right) \widetilde{V}_t\left(x_t, a_{t-1}\right).$$

Then $\widetilde{W}_t\left(a_{t-1}\right)$ satisfies the Bellman equation

$$\begin{aligned} \widetilde{W}_t\left(a_{t-1}\right) &= \max_{p_t, q_t} \sum_{a_t, x_t} \mu_t\left(x_t|a_{t-1}\right) p_t\left(a_t|x_t, a_{t-1}\right) \left[u\left(x_t, a_t\right) - \lambda \ln \frac{p_t\left(a_t|x_t, a_{t-1}\right)}{q_t\left(a_t|a_{t-1}\right)}\right] \\ &\quad + \beta \sum_{a_t} q_t(a_t|a_{t-1}) \widetilde{W}_{t+1}\left(a_t\right). \end{aligned} \tag{C.1}$$

Notice that $\widetilde{W}_t\left(a_{t-1}\right)$ is the same as the posterior-based value function $V_t\left(\mu_t\left(\cdot|a_{t-1}\right)\right)$.

By backward induction and Proposition 2, we have

$$\widetilde{V}_{t+1}\left(x_{t+1}, a_t\right) = \widehat{V}_{t+1}\left(x_{t+1}\right) + \lambda \ln \mu_{t+1}\left(x_{t+1}|a_t\right), \tag{C.2}$$

where $\widehat{V}_{t+1}\left(x_{t+1}\right)$ is independent of the prior $\mu_{t+1}\left(x_{t+1}|a_t\right)$ at history a_t in the convex hull of the optimal posteriors for all chosen actions a_{t+1} and $\mu_{t+1}\left(x_{t+1}|a_t\right)$ satisfies

$$\mu_{t+1}\left(x_{t+1}|a_t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t, a_t\right) \mu_t\left(x_t|a_t\right) = \frac{\sum_{x_t} \pi\left(x_{t+1}|x_t, a_t\right) p_t\left(a_t|x_t, a_{t-1}\right) \mu_t\left(x_t|a_{t-1}\right)}{q_t(a_t|a_{t-1})}, \tag{C.3}$$

when $q_t(a_t|a_{t-1}) > 0$.

Rewrite the objective function in (C.1) as

$$\sum_{x_t} \mu_t\left(x_t|a_{t-1}\right) \left[\sum_{a_t} p_t\left(a_t|x_t, a_{t-1}\right) \hat{u}\left(x_t, a_t\right) + \lambda \Phi\left(p_t, q_t; x_t, a_{t-1}\right)\right], \tag{C.4}$$

where we define

$$\hat{u}\left(x_t, a_t\right) \equiv u\left(x_t, a_t\right) + \beta \sum_{x_{t+1}} \pi\left(x_{t+1}|x_t, a_t\right)\left[\widehat{V}_{t+1}\left(x_{t+1}\right) + \lambda \ln \mu_t\left(x_t|a_{t-1}\right)\right],$$

and

$$\Phi\left(p_t, q_t; x_t, a_{t-1}\right) \equiv \sum_{x_{t+1}, a_t} \pi\left(x_{t+1}|x_t, a_t\right) p_t\left(a_t|x_t, a_{t-1}\right) \ln \frac{\left[\sum_{x_t'} \pi\left(x_{t+1}|x_t', a_t\right) p_t\left(a_t|x_t', a_{t-1}\right)\right]^{\beta}}{p_t\left(a_t|x_t, a_{t-1}\right)\left[q_t\left(a_t|a_{t-1}\right)\right]^{\beta-1}}.$$

Lemma 4 *For any x_t, a_{t-1}, the function $\Phi\left(p_t, q_t; x_t, a_{t-1}\right)$ is jointly concave in $p_t\left(a_t|x_t, a_{t-1}\right)$ and $q_t\left(a_t|a_{t-1}\right)$ for any $\beta \in (0, 1]$.*

Proof: Write

$$\begin{aligned}\Phi\left(p_t, q_t; x_t, a_{t-1}\right) &= \beta \sum_{x_{t+1}, a_t} \pi\left(x_{t+1}|x_t, a_t\right) p_t\left(a_t|x_t, a_{t-1}\right) \ln \frac{\sum_{x_t'} \pi\left(x_{t+1}|x_t', a_t\right) p_t\left(a_t|x_t', a_{t-1}\right)}{p_t\left(a_t|x_t, a_{t-1}\right)} \\ &\quad - (1-\beta) \sum_{a_t} p_t\left(a_t|x_t, a_{t-1}\right) \ln \frac{p_t\left(a_t|x_t, a_{t-1}\right)}{q_t\left(a_t|a_{t-1}\right)}.\end{aligned}$$

The expression on the first line is concave in p_t by an argument similar to the proof for Lemma 2. The expression on the second line is equal to the negative of the relative entropy between p_t and q_t, which is jointly concave. Thus we obtain the desired result. □

By this lemma we deduce that the objective function in (C.4) is jointly concave in p_t and q_t. Thus the first-order Kuhn-Tucker conditions are necessary and sufficient for optimality. Consider the Lagrange function

$$\begin{aligned}&\sum_{a_t, x_t} \mu_t\left(x_t|a_{t-1}\right) p_t\left(a_t|x_t, a_{t-1}\right)\left[u\left(x_t, a_t\right) - \lambda \ln \frac{p_t\left(a_t|x_t, a_{t-1}\right)}{q_t\left(a_t|a_{t-1}\right)}\right] \\ &+ \beta \sum_{a_t} q_t(a_t|a_{t-1}) \widetilde{W}_{t+1}\left(a_t\right) + \xi_t\Big(\sum_{a_t} p(a_t|x_t, a_{t-1}) - 1\Big),\end{aligned}$$

where ξ_t is the Lagrange multiplier associated with the constraint $\sum_{a_t} p(a_t|x_t, a_{t-1}) = 1$. For a fixed q_t and a fixed $\mu_t\left(x_t|a_{t-1}\right)$, take the first-order condition with respect to $p_t(a_t|x_t, a_{t-1})$. We first focus on the second term in the Lagrange function,

$$\begin{aligned}&\frac{\partial}{\partial p_t(a_t|x_t, a_{t-1})} \beta \sum_{a_t} q_t(a_t|a_{t-1}) \widetilde{W}_{t+1}(a_t) \\ =&\beta \frac{\partial}{\partial p_t} \sum_{a_t, x_{t+1}} q(a_t|a_{t-1}) \mu_{t+1}(x_{t+1}|a_t)\left[\widehat{V}_{t+1}(x_{t+1}) + \lambda \ln \mu_{t+1}(x_{t+1}|a_t)\right] \\ =&\beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t) \mu_t(x_t|a_{t-1})\left[\widehat{V}_{t+1}(x_{t+1}) + \lambda \ln \mu_{t+1}(x_{t+1}|a_t)\right] + \beta\lambda \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t) \mu_t(x_t|a_{t-1}) \\ =&\beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t) \mu_t(x_t|a_{t-1}) \widetilde{V}_{t+1}(x_{t+1}) + \beta\lambda \mu_t(x_t|a_{t-1}),\end{aligned}$$

where the first and the third equalities follow from (C.2) and the second equality follows from

$$\frac{\partial \mu_{t+1}(x_{t+1}|a_t)}{\partial p_t(a_t|x_t, a_{t-1})} = \frac{\pi(x_{t+1}|x_t, a_t)\mu_t(x_t|a_{t-1})}{q_t(a_t|a_{t-1})}.$$

Derivatives with respect to $p_t(a_t|x_t, a_{t-1})$ in the first and the third terms of the objective function are

$$\mu_t(x_t|a_{t-1})\left[u(x_t, a_t) - \lambda \ln \frac{p_t(a_t|x_t, a_{t-1})}{q_t(a_t|a_{t-1})}\right] - \lambda\mu_t(x_t|a_{t-1}) + \xi_t.$$

Combining all above partial derivatives, we obtain the following first-order condition:

$$\mu_t(x_t|a_{t-1})\Big[u(x_t, a_t) + \beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t)\widetilde{V}_{t+1}(x_{t+1}, a_t) - \lambda \ln \frac{p_t(a_t|x_t, a_{t-1})}{q_t(a_t|a_{t-1})} + \beta\lambda - \lambda\Big] + \xi_t = 0.$$

Dividing $\mu_t(x_t|a_{t-1})$ and exponentiating both sides, we obtain

$$q_t(a_t|a_{t-1}) \exp(v_t(x_t, a_t)/\lambda) \exp\left(1 - \beta - \frac{\xi_t}{\lambda\mu_t(x_t|a_{t-1})}\right) = p_t(a_t|x_t, a_{t-1}).$$

Summing a_t on both sides, we confirm

$$p_t(a_t|x_t, a_{t-1}) = \frac{q_t(a_t|a_{t-1}) \exp(v_t(x_t, a_t)/\lambda)}{\sum_{a'_t} q_t(a'_t|a_{t-1}) \exp(v_t(x_t, a'_t)/\lambda)}.$$

Other statements in Proposition 6 can be confirmed as well. Using Proposition 1 and Lemma 4, we can also derive the sufficient conditions as in Proposition 7.

D Forward-Backward Arimoto-Blahut Algorithm

The algorithm consists of the following steps.

Step 1. Initialize:

$$\begin{aligned} \mu_1^{(T,1)}(x_1, a_0) &= \mu_1(x_1), \\ p_t^{(T,0)}(a_t|x_t, a_{t-1}) &= p^{(T,0)}(a_t|x_t, a_{t-1}) > 0 \text{ for } t = 1, 2, ..., T, \\ \phi_{T+1}^{(k)} &= 1 \text{ for } k = 1, 2, ..., K, \end{aligned}$$

for all $x_t \in X$ and $a_t, a_{t-1} \in A$, where T is the horizon and K is a large integer.

Step 2. For $k = 1, 2, ..., K$ until convergence do the following:

- (forward path) For $t = 1, 2, ..., T$ do

$$\mu_{t+1}^{(T,k)}(x_{t+1}, a_t) = \sum_{x_t, a_{t-1}} \pi(x_{t+1}|x_t, a_t)\, p_t^{(T,k-1)}(a_t|x_t, a_{t-1})\, \mu_t^{(T,k)}(x_t, a_{t-1}),$$

$$q_t^{(T,k)}(a_t|a_{t-1}) = \sum_{x_t} p_t^{(T,k-1)}(a_t|x_t, a_{t-1})\, \mu_t^{(T,k)}(x_t|a_{t-1}),$$

$$\mu_t^{(T,k)}(x_t|a_{t-1}) = \frac{\mu_t^{(T,k)}(x_t, a_{t-1})}{\sum_{x_t} \mu_t^{(T,k)}(x_t, a_{t-1})} \text{ if } \sum_{x_t} \mu_t^{(T,k)}(x_t, a_{t-1}) > 0.$$

- (backward path) For $t = T, T-1, ..., 1$ do

$$v_t^{(T,k)}(x_t, a_t) = u(x_t, a_t) + \beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t) \lambda \ln \phi_{t+1}^{(T,k)}(x_{t+1}, a_t),$$

$$\phi_t^{(T,k)}(x_t, a_{t-1}) = \sum_{a_t} q_t^{(T,k)}(a_t|a_{t-1}) \exp\left(v_t^{(T,k)}(x_t, a_t)/\lambda\right),$$

$$p_t^{(T,k)}(a_t|x_t, a_{t-1}) = \frac{q_t^{(T,k)}(a_t|a_{t-1}) \exp\left[v_t^{(T,k)}(x_t, a_t)/\lambda\right]}{\phi_t^{(T,k)}(x_t, a_{t-1})}.$$

Step 3. Return $p_t^{(T,K)}(a_t|x_t, a_{t-1})$ and other variables.

For the infinite horizon case, we increase T until convergence.

E UPS Information Cost

In this section we study RI problems with UPS information cost functions introduced by CD (2013) and CDL (2018b). We show that our approach can be applied to this general case.

E. 1 Static Case

Following CD (2013) and CDL (2018b), we define a uniformly posterior separable information cost function as follows

$$C_H(\mu, \mu(\cdot|\cdot), q) = H(\mu) - \sum_a q(a) H(\mu(\cdot|a)), \tag{E.1}$$

where $H : \Delta(X) \to \mathbb{R}_+$ is a concave function (called generalized entropy) and $\mu(x) = \sum_a q(a) \mu(x|a)$. Clearly, observing information reduces uncertainty so that $C_H(\mu, \mu(\cdot|\cdot), q) \geq 0$. The following specifications of H are interesting:

- Shannon entropy: $H(\nu) = -\sum_x \nu(x) \ln \nu(x)$.

- Weighted entropy (Belis and Guiasu 1968):

$$H(\nu) = -\sum_x w(x) \nu(x) \ln \nu(x),$$

  where the weighting function satisfies $w(x) \geq 0$ and $\sum_x w(x) = 1$.

- Tsallis entropy (Havrda and Charvat (1967) and Tsallis (1988)):

$$H(\nu) = \frac{1}{\sigma - 1} \sum_x \nu(x) \left(1 - \nu(x)^{\sigma-1}\right), \quad \sigma > 0.$$

- Rényi entropy (Rényi (1961)):

$$H(\nu) = \frac{1}{1-\alpha} \ln\left(\sum_x \nu(x)^\alpha\right), \ \alpha \in (0,1).$$

The restriction of $\sigma > 0$ and $\alpha \in (0,1)$ ensures the concavity of H. The Shannon entropy is obtained as the limit as $\sigma \to 1$ and $\alpha \to 1$. Notice that the information cost function based on the weighted ϕ-divergence,

$$C\left(\mu, \mu\left(\cdot|\cdot\right), q\right) \equiv \sum_{a,x} q\left(a\right) D_{\phi}^{w}\left(\mu\left(\cdot|a\right)||\mu\right),$$

may not satisfy the posterior-separability property, where

$$D_{\phi}^{w}\left(\mu\left(\cdot|a\right)||\mu\right) \equiv \sum_{x} w\left(x\right)\mu\left(x\right)\phi\left(\frac{\mu\left(x|a\right)}{\mu\left(x\right)}\right)$$

is called a w-weighted ϕ-divergence. The function $\phi : \mathbb{R}_{+} \to \mathbb{R}_{+}$ satisfies $\phi'' < 0$, $\phi\left(1\right) = \phi'\left(1\right) = 0$. The Kullback-Laibler relative entropy is a special case with $\phi\left(z\right) = z\ln z - z + 1$ and $w\left(x\right) \equiv 1$. In this case the cost function becomes the Shannon mutual information.

Problem 5 *(static RI problem with UPS cost)*

$$V\left(\mu\right) = \max_{q,\mu\left(\cdot|\cdot\right)} \mathbb{E}\left[u\left(x,a\right)\right] - \lambda C_{H}\left(\mu, \mu\left(\cdot|\cdot\right), q\right)$$

subject to $\mu\left(x\right) = \sum_{a} q\left(a\right)\mu\left(x|a\right)$*, where* $\lambda > 0$.

Define the net utility function as

$$N_{H}^{a}\left(\mu\left(\cdot|a\right)\right) = \sum_{x}\mu\left(x|a\right)u\left(x,a\right) + \lambda H\left(\mu\left(\cdot|a\right)\right).$$

Then the problem becomes

$$V\left(\mu\right) = \max_{q,\mu\left(\cdot|\cdot\right)} \sum_{a} q\left(a\right) N_{H}^{a}\left(\mu\left(\cdot|a\right)\right) - \lambda H\left(\mu\right). \tag{E.2}$$

Since H is concave, the net utility function N_{H}^{a} is concave so that Lemmas 1 through 3 in CD (2013) can be applied to characterize optimal solutions.

E. 2 Two-period Case

We write the objective function in Problem 4 for the UPS cost case as

$$\begin{aligned}
& J\left(q_1, \mu_1\left(\cdot|\cdot\right), q_2, \mu_2\left(\cdot|\cdot\right)\right) \\
\equiv\; & \mathbb{E}\left[u\left(x_1,a_1\right) + \beta u\left(x_2,a_2\right)\right] - \lambda C_{H}\left(\mu_1, \mu_1\left(\cdot|\cdot\right), q_1\right) \\
& -\beta\lambda\sum_{a_1} q_1\left(a_1\right) C_{H}\left(\mu_2\left(\cdot|a_1\right), \mu_2\left(\cdot|\cdot, a_1\right), q_2\left(\cdot|a_1\right)\right) \\
=\; & \sum_{a_1,x_1} q_1\left(a_1\right)\left[\mu_1\left(x_1|a_1\right)u\left(x_1,a_1\right) + \lambda H\left(\mu_1\left(\cdot|a_1\right)\right)\right] - \lambda H\left(\mu_1\right) \\
& +\beta\sum_{a_1,a_2,x_2} q_1\left(a_1\right)\left\{q_2\left(a_2|a_1\right)\left[\mu_2\left(x_2|a^2\right)u\left(x_2,a_2\right) + \lambda H\left(\mu_2\left(\cdot|a^2\right)\right)\right] - \lambda H\left(\mu_2\left(\cdot|a_1\right)\right)\right\},
\end{aligned}$$

where $\mu_1(x_1) = \sum_{a_1} q_1(a_1)\mu_1(x_1|a_1)$ and

$$\mu_2(x_2|a_1) \equiv \sum_{a_2} \mu_2\left(x_2|a^2\right) q_2(a_2|a_1) = \sum_{x_1} \pi(x_2|x_1, a_1)\mu_1(x_1|a_1). \tag{E.3}$$

We consider interior solutions with $q_1(a_1) > 0$ and $q_2(a_2|a_1) > 0$ for simplicity. We solve Problem 4 by dynamic programming using the predictive distribution as the state variable. First consider the RI problem in period 2 for $q_1(a_1) > 0$:

$$V_2(\mu_2(\cdot|a_1)) = \max \sum_{a_2} q_2(a_2|a_1) N_H^a\left(\mu_2\left(\cdot|a^2\right)\right) - \lambda H(\mu_2(\cdot|a_1)) \tag{E.4}$$

subject to

$$\mu_2(x_2|a_1) = \sum_{a_2} \mu_2\left(x_2|a^2\right) q_2(a_2|a_1),$$

where we define the net utility as

$$N_H^{a_2}\left(\mu_2\left(\cdot|a^2\right)\right) = \sum_{x_2} \mu_2\left(x_2|a^2\right) u(x_2, a_2) + \lambda H\left(\mu_2\left(\cdot|a^2\right)\right).$$

The choice variables are $\mu_2\left(\cdot|a^2\right)$ and $q_2(\cdot|a_1)$. Taking the predictive distribution $\mu_2(\cdot|a_1)$ as the prior at history a_1, we view this problem as a static RI problem and apply Lemmas 1 through 3 in CD (2013) to characterize the solution in period 2.

As in Proposition 2, we can show that

$$V_2(\mu_2(\cdot|a_1)) = \sum_{x_2} \mu_2(x_2|a_1)\widehat{V}_2(x_2, a_1) - \lambda H(\mu_2(\cdot|a_1)), \tag{E.5}$$

where $\widehat{V}_2(x_2, a_1)$ and the optimal posterior $\mu_2\left(\cdot|a^2\right) \in (0,1)$ are independent of the prior $\mu_2(x_2|a_1)$ in the convex hull of the optimal posteriors $\mu_2\left(\cdot|a^2\right)$ for all a_2 such that $q_2(a_2|a_1) > 0$. Since the history a_1 enters the problem in (16) through $\mu_2(\cdot|a_1)$ only, $\mu_2\left(\cdot|a^2\right)$ is independent of a_1 and can be written as $\mu_2(\cdot|a_2)$. Moreover $\widehat{V}_2(x_2, a_1)$ is also independent of $\mu_2(x_2|a_1)$ and can be written as $\widehat{V}_2(x_2)$. Since H is concave, V_2 is convex in $\mu_2(\cdot|a_1)$ in the preceding convex hull.

By dynamic programming, the problem in period 1 is to choose $\mu_1(x_1|a_1)$ and $q_1(a_1)$ to solve:

$$\begin{aligned} V_1(\{\mu_1(x_1)\}) \quad &= \quad \max \sum_{a_1, x_1} q_1(a_1)\left[\mu_1(x_1|a_1) u(x_1, a_1) + \lambda H(\mu_1(\cdot|a_1))\right] - \lambda H(\mu_1) \\ &\quad + \beta \sum_{a_1} q_1(a_1) V_2(\mu_2(\cdot|a_1)) \end{aligned} \tag{E.6}$$

subject to $\mu_1(x_1) = \sum_{a_1} q_1(a_1)\mu_1(x_1|a_1)$ for all $x_1 \in X$ and

$$\mu_2(x_2|a_1) = \sum_{x_1} \pi(x_2|x_1, a_1)\mu_1(x_1|a_1) \text{ for all } x_2 \in X. \tag{E.7}$$

The link between the problems in the two periods is through the predictive distribution $\mu_2\left(x_2|a_1\right)$ in (E.3).

Substituting (E.3) and (E.5) into (E.6) yields

$$V_1\left(\{\mu_1\left(x_1\right)\}\right)=\max_{q_1,\mu_1(\cdot|\cdot)}\sum_{a_1}q_1\left(a_1\right)N_G^{a_1}\left(\mu_1\left(\cdot|a_1\right)\right)-\lambda H\left(\mu_1\right), \tag{E.8}$$

where we define the net utility associated with action a_1 as

$$N_G^{a_1}\left(\mu_1\left(\cdot|a_1\right)\right)\equiv\sum_{x_1}\mu_1\left(x_1|a_1\right)\hat{u}\left(x_1,a_1\right)-\lambda G\left(\mu_1\left(\cdot|a_1\right)\right).$$

Here the new utility function is given by

$$\hat{u}\left(x_1,a_1\right)=u\left(x_1,a_1\right)+\beta\sum_{x_2}\pi\left(x_2|x_1,a_1\right)\widehat{V}_2\left(x_2\right),$$

and the information cost is given by

$$G\left(\mu_1\left(\cdot|a_1\right)\right)=\beta H\left(\mu_2\left(\cdot|a_1\right)\right)-H\left(\mu_1\left(\cdot|a_1\right)\right),$$

where $\mu_2\left(\cdot|a_1\right)$ satisfies (E.3).

Since $\widehat{V}_2\left(x_2\right)$ is independent of $\mu_1\left(x_1|a_1\right)$, $\hat{u}\left(x_1,a_1\right)$ is also independent of $\mu_1\left(\cdot|a_1\right)$. Thus the concavity of $N_G^{a_1}$ is equivalent to the convexity of G. For the posterior-based approach to work, we need the following condition to hold:

Condition 1 *For any $\beta\in(0,1]$ and $a_1\in A$ with $q_1\left(a_1\right)>0$ the function $G\left(\mu_1\left(\cdot|a_1\right)\right)$ is convex in $\mu_1\left(\cdot|a_1\right)$.*

Since H is a concave function, it is not obvious that G is a convex function. We have shown in Lemma 2 that G is convex for the Shannon entropy. We next show that this is also true for the weighted entropy under a condition for the weighting function.

Lemma 5 *For the weighted entropy, suppose that the transition kernel $\pi(x_2|x_1)$ is independent of actions and $w(x_1)=\sum_{x_2}w(x_2)\pi(x_2|x_1)$ for any x_1 and x_2. Then for any $\beta\in(0,1]$, the function $G\left(\mu_1\left(\cdot|a_1\right)\right)$ is convex in $\mu_1\left(\cdot|a_1\right)$ and the net utility $N_G^{a_1}\left(\mu_1\left(\cdot|a_1\right)\right)$ is concave in $\mu_1\left(\cdot|a_1\right)$.*

Proof: We define the function

$$\begin{aligned}&\widetilde{G}\left(\mu_1\left(\cdot|a_1\right),\widetilde{\mu}_2\left(\cdot|a_1\right)\right)\\&=\sum_{x_1}w\left(x_1\right)\mu_1\left(x_1|a_1\right)\ln\mu_1\left(x_1|a_1\right)-\beta\sum_{x_1,x_2}w\left(x_2\right)\pi\left(x_2|x_1,a_1\right)\mu_1\left(x_1|a_1\right)\ln\widetilde{\mu}_2\left(x_2|a_1\right)\\&=\sum_{x_1,x_2}w(x_2)\pi(x_2|a_1,a_1)\mu_1(x_1|a_1)\ln\frac{\mu_1(x_1|a_1)}{[\widetilde{\mu}_2(x_2|a_1)]^{\beta}}.\end{aligned}$$

The remaining proof is the same as that of Lemma 2. $\square$

Let w be a stationary distribution of the state process. Then the condition in the lemma is satisfied. It follows from this lemma that we can apply Lemmas 1 through 3 of CD (2013) to characterize the optimal solution in period 1. For other generalized entropy functions, we need to verify Condition 1 case by case.

E. 3 General Case

For the general dynamic case, we define the discounted UPS information cost as:

Definition 3 *The discounted UPS information cost is given by*

$$\sum_{t=1}^{T} \beta^{t-1} I\left(x_t; a_t | a^{t-1}\right), \ I\left(x_t; a_t | a^{t-1}\right) \equiv \sum_{a^{t-1}} \mu_t\left(a^{t-1}\right) C_H\left(\mu_t\left(x_t | a^{t-1}\right), \mu_t\left(x_t | a^t\right), q_t\left(a_t | a^{t-1}\right)\right), \tag{E.9}$$

where C_H is given in (E.1), the sequences of marginal distributions $\left\{\mu_t\left(a^{t-1}\right)\right\}$, predictive distributions $\left\{\mu_t\left(x_t | a^{t-1}\right)\right\}$, posteriors $\left\{\mu_t\left(x_t | a^t\right)\right\}$, and default rules $\left\{q_t\left(a_t | a^{t-1}\right)\right\}$ are consistent with the joint distribution $\mu_T\left(x^T, a^T\right)$, and the predictive distributions and posteriors satisfy

$$\mu_t\left(x_t | a^{t-1}\right) = \sum_{a_t} q\left(a_t | a^{t-1}\right) \mu_t\left(x_t | a^t\right) = \sum_{x_{t-1}} \pi\left(x_t | x_{t-1}, a_{t-1}\right) \mu_{t-1}\left(x_{t-1} | a^{t-1}\right), \ t \geq 1.$$

For the Shannon entropy case in Definition 1, we allow the DM to learn the whole history of states x^t at each time t. In Proposition 3 we have shown that the history of states does not matter so that we can focus on the Shannon information cost of the form $I\left(x_t; a_t | a^{t-1}\right)$. Here we directly impose the assumption in (E.9) for the UPS information cost. This is analogous to the transfer entropy defined in TSS (2018). The transfer entropy is widely applied in information theory, physics, climatology, and neuroscience and is useful to limit the amount of information that can be learned. We will show later that the assumption in (E.9) is critical for the optimal choice rule to take the simple form $p_t\left(a_t | x_t, a^{t-1}\right)$, which facilitates computations.

Problem 6 *(posterior-based dynamic RI problem with UPS cost)*

$$\max \ \mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} u\left(x_t, a_t\right) + \beta^{T-1} u_{T+1}\left(x_{T+1}\right)\right] - \lambda \sum_{t=1}^{T} \beta^{t-1} I\left(x_t; a_t | a^{t-1}\right), \tag{E.10}$$

where $I\left(x_t; a_t | a^{t-1}\right)$ is given in (E.9), the expectation is taken with respect to the joint distribution induced by the transition kernel π, $\left\{\mu_t\left(x_t | a^t\right)\right\}$, and $\left\{q_t\left(a_t | a^{t-1}\right)\right\}$. The choice variables are sequences of $\left\{\mu_t\left(x_t | a^t\right)\right\}$ and $\left\{q_t\left(a_t | a^{t-1}\right)\right\}$.

We can solve this problem using the posterior-based dynamic programming

$$\begin{aligned} V_t\left(\mu_t\left(\cdot|a^{t-1}\right)\right) &= \max_{\mu_t\left(\cdot|\cdot,a^{t-1}\right),q_t\left(\cdot|a^{t-1}\right)} \sum_{x_t,a_t} q_t\left(a_t|a^{t-1}\right)\mu_t\left(x_t|a^t\right)u\left(x_t,a_t\right) \\ & -\lambda C_H\left(\mu_t\left(x_t|a^{t-1}\right),\mu_t\left(x_t|a^t\right),q_t\left(a_t|a^{t-1}\right)\right) \\ & +\beta\sum_{a_t} q_t\left(a_t|a^{t-1}\right)V_{t+1}\left(\mu_{t+1}\left(\cdot|a^t\right)\right), \end{aligned}$$

subject to (38) and (39). Starting with a finite-horizon T, we can solve this problem by backward induction. We use a similar method discussed earlier to derive an optimal solution. We omit the detailed derivations here.

After obtaining a solution for $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$, we use the Bayes rule to derive the choice rule $p_t\left(a_t|x_t,a^{t-1}\right)$ as in (40). This choice rule does not depend on the history x^{t-1}.

E. 4 Signal-Based Formulation

Following SSM (2017), we consider the signal-based formulation. Suppose that there is a signal space S satisfying $|A| \leq |S| < \infty$. At time t, the DM can choose any signal about the state x_t with realizations s_t in S. A strategy is a pair (f,σ) composed of

1. an information strategy f consisting of a system of signal distributions $f_t\left(s_t|x^t,s^{t-1}\right)$, for all $s^t \in S^t$, $x^t \in X^t$, and $t \geq 1$;

2. an action strategy σ consisting of a system of mappings $\sigma_t : S^t \to A$, which give an action $a_t = \sigma_t\left(s^t\right)$, for $t \geq 1$.

Given an action strategy σ, we denote by $\sigma^t\left(s^t\right)$ the history of actions up to time t given the realized signals s^t. The state transition kernel π and the strategy (f,σ) induce a sequence of joint distributions for x^{t+1} and s^t recursively

$$\mu_{t+1}\left(x^{t+1},s^t\right) = \pi\left(x_{t+1}|x_t,\sigma_t\left(s^t\right)\right)f_t\left(s_t|x^t,s^{t-1}\right)\mu_t\left(x^t,s^{t-1}\right),$$

where $\mu_1\left(x^1,s^0\right) = \mu_1\left(x_1\right)$ is given. Using this sequence of distributions, we can compute the predictive distributions $\mu_t\left(x_t|s^{t-1}\right)$ and the posteriors $\mu_t\left(x_t|s^t\right)$ and hence we can define the discounted UPS information cost $\sum_{t=1}^T \beta^{t-1} I\left(x_t;s_t|s^{t-1}\right)$, similar to (E.9).

Problem 7 *(signal-based dynamic RI problem with UPS costs)*

$$\max_{f,\sigma}\ \mathbb{E}\left[\sum_{t=1}^T \beta^{t-1}u\left(x_t,\sigma_t\left(s^t\right)\right)+\beta^T u_{T+1}\left(x_{T+1}\right)\right]-\lambda\sum_{t=1}^T\beta^{t-1}I\left(x_t;s_t|s^{t-1}\right)$$

where the expectation is taken with respect to the joint distribution over sequences x^{T+1} *and* s^T *induced by the transition kernel* π *and the strategy* (f,σ).

We say that a strategy (f,σ) generates a choice rule $\mathbf{p}$ if

$$p_t\left(a_t|x^t,a^{t-1}\right)=\Pr\left(\sigma_t\left(s^t\right)=a_t|x^t,\sigma^{t-1}\left(s^{t-1}\right)=a^{t-1}\right)$$

for all a_t, x^t, and a^{t-1}. Conversely, a choice rule $\mathbf{p}$ of the form $p_t\left(a_t|x^t,a^{t-1}\right)$ can induce a strategy (f,σ) as described by SSM (2017).

We have the following recommendation lemma or the revelation principle similar to Lemma 1 of SSM (2017) or Lemma 2 of Ravid (2019).

Lemma 6 *Suppose that the function*

$$G(\mu_t(\cdot|a^t))\equiv\beta H\left(\mu_{t+1}(\cdot|a^t)\right)-H(\mu_t(\cdot|a^t))$$

is convex in $\mu_t(\cdot|a^t)$ for any chosen a^t and $t\geq 1$, where $\mu_{t+1}(x_{t+1}|a^t)=\sum_{x_t}\pi\left(x_{t+1}|x_t,a_t\right)\mu_t(x_t|a^t)$. Then any strategy (f,σ) solving the dynamic RI Problem 7 generates sequences of posteriors $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and default rules $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$ solving Problem 6. Conversely any sequences of posteriors $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and default rules $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$ solving Problem 6 induce a strategy (f,σ) solving Problem 7.

Proof: We focus on the finite-horizon case with $T<\infty$. The result for the infinite-horizon case can be obtained by taking limits as $T\to\infty$.

First, using the constructed $\mathbf{p}$ from a strategy (f,σ), we can define a sequence of joint distributions $\mu_t\left(x^t,a^{t-1}\right)$ as in (1). The distribution induced by the strategy (f,σ) and the sequence of distributions $\mu_t\left(x^t,a^{t-1}\right)$ give the same stream of expected utility. Next we show that the discounted information cost associated with $\mathbf{p}$, $\sum_{t=1}^T\beta^{t-1}I(x_t;a_t|a^{t-1})$, is not larger than that associated with (f,σ). These information costs can be computed using the posteriors and predictive distributions (priors) induced by the corresponding joint distributions.

By the definition of the discounted UPS cost in (E.9), we compute

$$\begin{aligned}I\left(x_t;a_t|a^{t-1}\right) &= \sum_{a^{t-1}}\mu_t\left(a^{t-1}\right)C_H\left(\mu_t\left(x_t|a^{t-1}\right),\mu_t\left(x_t|a^t\right),q_t\left(a_t|a^{t-1}\right)\right)\\ &= \sum_{a^{t-1}}\mu_t\left(a^{t-1}\right)H\left(\mu_t\left(\cdot|a^{t-1}\right)\right)-\sum_{a^{t-1}}\mu_t\left(a^{t-1}\right)\sum_{a_t}q_t\left(a_t|a^{t-1}\right)H\left(\mu_t\left(\cdot|a^t\right)\right)\\ &= \sum_{a^{t-1}}\mu_t\left(a^{t-1}\right)H\left(\mu_t\left(\cdot|a^{t-1}\right)\right)-\sum_{a^t}\mu_t\left(a^t\right)H\left(\mu_t\left(\cdot|a^t\right)\right).\end{aligned}$$

Since both $\mu_{t+1}(a^t)$ and $\mu_t(a^t)$ are marginal distributions of a^t, they are the same. Rearranging the terms in the discounted UPS cost yields

$$\begin{aligned}
&\sum_{t=1}^{T} \beta^{t-1} I(x_t; a_t|a^{t-1}) \\
= \quad & H(\mu_1) + \sum_{t=1}^{T-1} \sum_{a^t} \beta^{t-1} \mu_t(a^t) \left[\beta H\left(\mu_{t+1}(\cdot|a^t)\right) - H(\mu_t(\cdot|a^t))\right] \\
& -\beta^{T-1} \sum_{a^T} \mu_T(a^T) H(\mu_T(\cdot|a^T)) \\
= \quad & H(\mu_1) + \sum_{t=1}^{T-1} \sum_{a^t} \beta^{t-1} \mu_t(a^t)\, G(\mu_t(\cdot|a^t)) - \beta^{T-1} \sum_{a^T} \mu_T(a^T) H(\mu_T(\cdot|a^T)). \qquad \text{(E.11)}
\end{aligned}$$

We can derive a similar decomposition for $\sum_{t=1}^{T} \beta^{t-1} I(x_t; s_t|s^{t-1})$.

Now we prove that

$$\sum_{a^t} \mu_t(a^t)\, G(\mu_t(\cdot|a^t)) \leq \sum_{s^t} \mu_t(s^t)\, G(\mu_t(\cdot|s^t)).$$

Since $a^t = \sigma^t\left(s^t\right)$, we have

$$\mu_t(x_t|a^t) = \sum_{s^t} \mu_t(x_t|s^t) \Pr(s^t|a^t), \;\; x_t \in X.$$

Since G is convex, it follows from Jensen's inequality that

$$G(\mu_t(\cdot|a^t)) \leq \sum_{s^t} \Pr(s^t|a^t) G(\mu_t(\cdot|s^t)).$$

Multiplying both sides by $\mu_t(a^t)$ and summing over a^t, we obtain

$$\sum_{a^t} \mu_t(a^t)\, G(\mu_t(\cdot|a^t)) \leq \sum_{s^t} \sum_{a^t} \Pr(s^t|a^t) \mu_t(a^t)\, G(\mu_t(\cdot|s^t)) = \sum_{s^t} \mu_t(s^t)\, G(\mu_t(\cdot|s^t)).$$

Since the generalized entropy H is concave, we can similarly prove that

$$\sum_{a^T} \mu_T(a^T) H(\mu_T(\cdot|a^T)) \geq \sum_{s^T} \mu_T(s^T)\, H(\mu_T(\cdot|s^T)).$$

Applying the preceding two inequalities to the second and the third terms on the right-hand side of (E.11), we obtain

$$\sum_{t=1}^{T} \beta^{t-1} I(x_t; a_t|a^{t-1}) \leq \sum_{t=1}^{T} \beta^{t-1} I(x_t; s_t|s^{t-1}).$$

We have shown that the discounted expected payoff from any strategy (f, σ) is not larger than the value of the objective function in Problem 6 given the sequences of posteriors and default rules consistent with the joint distribution $\mu_T\left(x^{T+1}, a^T\right)$, that is induced by the choice rule generated by (f, σ). Conversely, using the Bayes rule in (40) to construct the choice rule $\mathbf{p}$, we follow the same argument as in SSM (2017) to construct a strategy (f, σ). Notice that the choice rule takes the form $p_t\left(a_t|x_t, a^{t-1}\right)$ so that the signal distribution f_t depends only on $\left(x_t, s^{t-1}\right)$, but not on x^{t-1}. The discounted expected payoff from this strategy is identical to the value of the objective function in Problem 6 given the sequences of posteriors and default rules. These two relationships together imply the result. $\square$

The convexity assumption on G is essentially the same as Condition 1 discussed before, which is critical to ensure the concavity of the net utility function so that the posterior-based approach can work.

F Markovian versus History-Dependent Solutions

In this appendix we present two numerical examples to illustrate Markovian solutions and history-dependent solutions. For both examples, we use both the fully history-dependent forward-backward Arimoto-Blahut algorithm and the Markovian version described in Appendix D to compute numerical solutions. For the first example, let $T = 3$, $u_{T+1} = 0$, $X = A = \{1, 2, 3\}$, and the transition kernel satisfy $\pi\left(x_{t+1}|x_t, a_t\right) = 1 - \gamma$ if $x_{t+1} = x_t$; $\pi\left(x_{t+1}|x_t, a_t\right) = \gamma/2$ if $x_{t+1} \neq x_t$, for all a_t. Let $\mu_1(1) = 0.2$, $\mu_1(2) = \mu_1(3) = 0.4$, $\beta = \lambda = 1$, $\gamma = 0.2$, $u(x, a) = x - 1$ if $x = a$; $u(x, a) = 0$, otherwise. Figure 7 presents the solution for this dynamic RI problem. History may matter only in period 3. We find that $q_3(a_3 = 2|a_2 = 2, a_1 = 2) = q_3(2|2, 3) = 0.8723$, $q_3(3|2, 2) = q_3(3|2, 3) = 0.1277$, and $q_3(3|3, 2) = q_3(3|3, 3) = 1$. The corresponding predictive distributions satisfy $\mu_3(x_3|a_2 = 2, a_1 = 2) = \mu_3(x_3|2, 3)$ and $\mu_3(x_3|3, 2) = \mu_3(x_3|3, 3)$ for all $x_3 \in X$. Thus the solution is Markovian. Using our algorithm in Appendix D gives an almost identical solution. Notice that this solution is not interior, a case not covered by SSM (2017).

[Insert Figures 7 and 8 Here]

For the second example, let $T = 3$, $u_{T+1} = 0$, $X = A = \{1, 2\}$, and the transition kernel satisfy $\pi_t\left(x_{t+1}|x_t, a_t\right) = 1 - \gamma_t$ if $x_{t+1} = x_t$; $\pi_t\left(x_{t+1}|x_t, a_t\right) = \gamma_t$ if $x_{t+1} \neq x_t$, for all a_t. Let $\lambda = 10$, $\beta = 1$, $\mu_1(1) = 0.7$, $\gamma_1 = 0.15$, $\gamma_2 = 0.9$, $u(x, a) = 5x$ if $x = a$; $u(x, a) = 0$, otherwise. Figures 8 presents the solution for this dynamic RI problem. We find that the default rules are history dependent as $q_3(1|2, 1) \neq q_3(1|2, 2)$ and $q_3(2|2, 1) \neq q_3(2|2, 2)$.The predictive distributions are also history dependent as $\mu_3(x_3|2, 2) \neq \mu_3(x_3|2, 1)$ for $x_3 \in X$. Using our algorithm in Appendix D gives a suboptimal Markovian solution, which is different from the optimal history-dependent solution. We find that the welfare loss is very small. In particular, the optimal payoff in period 1 is 14.4372, and the payoff implied by the Markovian solution is 14.4362.

Reference

Arimoto, Suguru, 1972, An Algorithm for Computing the Capacity of Arbitrary Discrete Memoryless Channels, *IEEE Transactions on Information Theory* 18, 14-20.

Belis, Mariana, and Silviu Guiasu, 1968, A Quantitative and Qualitative Measure of Information in Cybernetic Systems, *IEEE Transactions on Information Theory* 14, 593–594.

Berger, Toby, 1971, *Rate Distortion Theory*, Prentice-Hall, New Jersey.

Bertsekas, Dimitri, 2016, *Nonlinear Programming*, Athena Scientific.

Blahut, Richard, E., 1972, Computation of Channel Capacity and Rate-Distortion Functions, *IEEE Transactions on Information Theory* 18, 460-473.

Caplin, Andrew, and Mark Dean, 2013, Behavioral Implications of Rational Inattention with Shannon Entropy, NBER working paper 19318.

Caplin, Andrew, Mark Dean, and John Leahy, 2018a, Rational Inattention, Optimal Consideration Sets, and Stochastic Choice, forthcoming in *Review of Economic Studies*.

Caplin, Andrew, Mark Dean, and John Leahy, 2018b, Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy, working paper, Columbia University.

Cover, Thomas M., and Joy A. Thomas, 2006, *Elements of Information Theory*, 2th edition, John Wiley & Sons, Inc.

Denti, Tommaso, Massimo Marinacci, and Luigi Montrucchio, 2019, A Note on Rational Inattention and Rate Distortion Theory, forthcoming in *Decisions in Economics and Finance.*

Fudenberg, Drew, and Tomasz Strzalecki, 2015, Dynamic Logit with Choice Aversion, *Econometrica* 83, 651-691.

Gallager, Robert G., 1968, *Information Theory and Reliable Communication*, John Wiley & Sons, Inc.

Havrda, Jan, and František Charvat, 1967, Quantification Method of Classification Processes: Concept of Structural a-Entropy, *Kybernetika* 3, 30-35.

Hébert, Benjamin, and Michael Woodford, 2018, Rational Inattention in Continuous Time, working paper, Columbia University.

Jung, Junehyuk, Jeong-Ho Kim, Filip Matějka, and Christopher A. Sims, 2018, Discrete Actions in Information-constrained Decision Problems, working paper, Princeton University.

Kamenica, Emir, 2018, Bayesian Persuasion and Information Design, forthcoming in *Annual Review of Economics.*

Luo, Yulei, 2008, Consumption Dynamics under Information Processing Constraints, *Review of Economic Dynamics* 11, 366-385.

Maćkowiak, Bartosz, and Mirko Wiederholt, 2009, Optimal Sticky Prices under Rational Inattention, *American Economic Review* 99, 769–803.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, 2018, Survey: Rational Inattention, a Disciplined Behavioral Model, working paper, Sciences Po.

Massey, James, L., 1990, Causality, Feedback and Directed Information, *International Symposium on Information Theory and its Applications.*

Matějka, Filip, and Alisdair McKay, 2015, Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model, *American Economic Review* 105, 272-298.

Mattsson, Lars-Göran and Jörgen W. Weibull, 2002, Probabilistic Choice and Procedurally Bounded Rationality, *Games and Economic Behavior* 41, 61-78.

Miao, Jianjun, 2019, Multivariate LQG Control under Rational Inattention in Continuous Time, working paper, Boston University.

Miao, Jianjun, Jieran Wu, and Eric Young, 2019, Multivariate Rational Inattention, working paper, Boston University.

Mondria, Jordi, 2010, Portfolio Choice, Attention Allocation, and Price Comovement, *Journal of Economic Theory* 145, 1837-1864.

Peng, Lin, and Wei Xiong, 2006, Investor Attention, Overconfidence and Category Learning, *Journal of Financial Economics* 80, 563-602.

Puterman, Martin L., 2005, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley&Sons, New Jersey.

Ravid, Doron, 2019, Bargaining with Rational Inattention, working paper, University of Chicago.

Rényi, Alfred, 1961, On Measures of Entropy and Information, Fourth Berkeley Symposium on Mathematical Statistics and Probability, 547–561, 1961.

Rust, John, 1987, Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, *Econometrica* 55, 999-1033.

Rust, John, 1994, Structural Estimation of Markov Decision Processes, in R. Engle and D. McFadden (eds.) *Handbook of Econometrics* Volume 4, 3082–3139, North Holland.

Shannon, Claude, E., 1948, A Mathematical Theory of Communication, *Bell System Technical Journal* 27, 379–423.

Sims, Christopher A., 1998, Stickiness, *Carnegie-Rochester Conference Series On Public Policy* 49, 317–356.

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

Sims, Christopher A., 2011, Rational Inattention and Monetary Economics, in *Handbook of Monetary Economics*, V. 3A, Benjamin M. Friedman and Michael Woodford (ed.), North-Holland.

Steiner, Jakub, Colin Stewart, and Filip Matějka, 2017, Rational Inattention Dynamics: Inertia and Delay in Decision-Making, *Econometrica* 85, 521-553.

Stokey, Nancy L., Robert E. Lucas with Edward C. Prescott, 1989, *Recursive Methods in Economic Dynamics*, Harvard University Press.

Tanaka, Takashi, Henrik Sandberg, Mikael Skoglund, 2018, Transfer-Entropy-Regularized Markov Decision Processes, working paper.

Tsallis, Constantino, 1988, Possible Generalization of Boltzmann-Gibbs Statistics, *Journal of Statistical Physics* 52, 479–487.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information Acquisition and Under-diversification, *Review of Economic Studies* 77, 571-626.

Woodford, Michael, 2009, Information Constrained State Dependent Pricing. *Journal of Monetary Economics* 56(S), S100–S124.

Zhong, Weijie, 2019, Optimal Dynamic Information Acquisition, working paper, Columbia University.

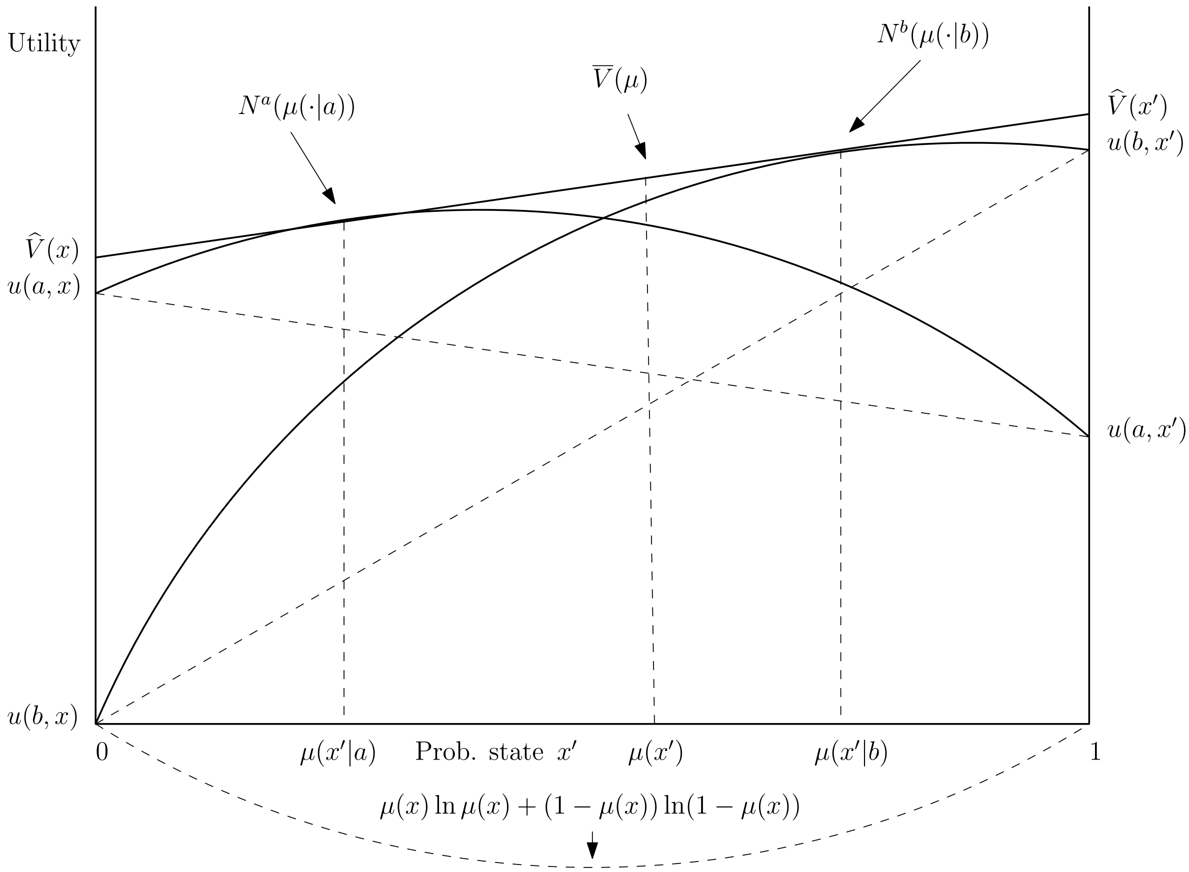

Figure 1: The net utility function and concavification.

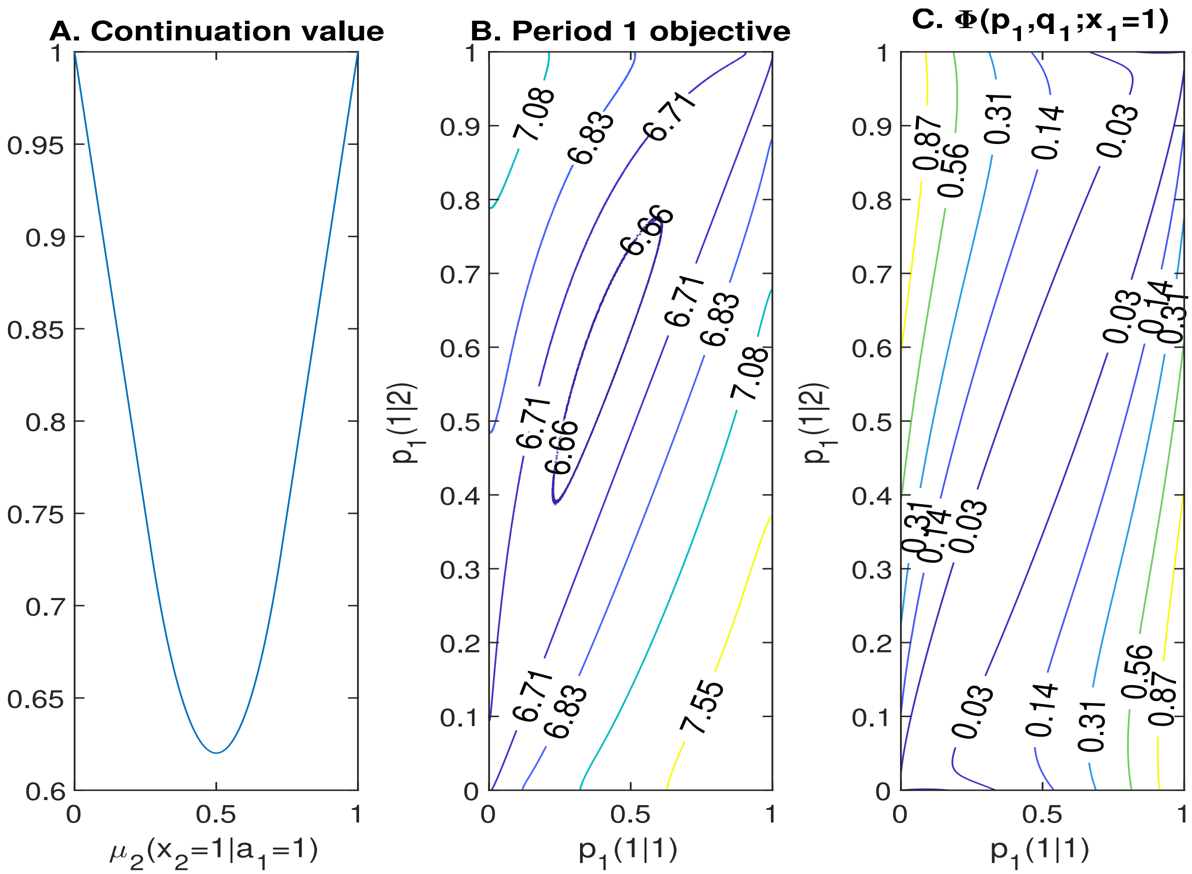

Figure 2: Panel A plots the continuation value as a function of $\mu_2(x_2 = 1|a_1 = 1)$. Panel B shows the contour plot of the objective function in period 1 as a function of $p_1(1|1)$ and $p_1(1|2)$. Panel C shows the contour plot of $\Phi(p_1, q_1; x_1 = 1)$ as a function $p_1(1|1)$ and $p_1(1|2)$. For the last two functions, we replace $q_1(a_1)$ by $\sum_{x_1} p_1(a_1|x_1)\mu(x_1)$.

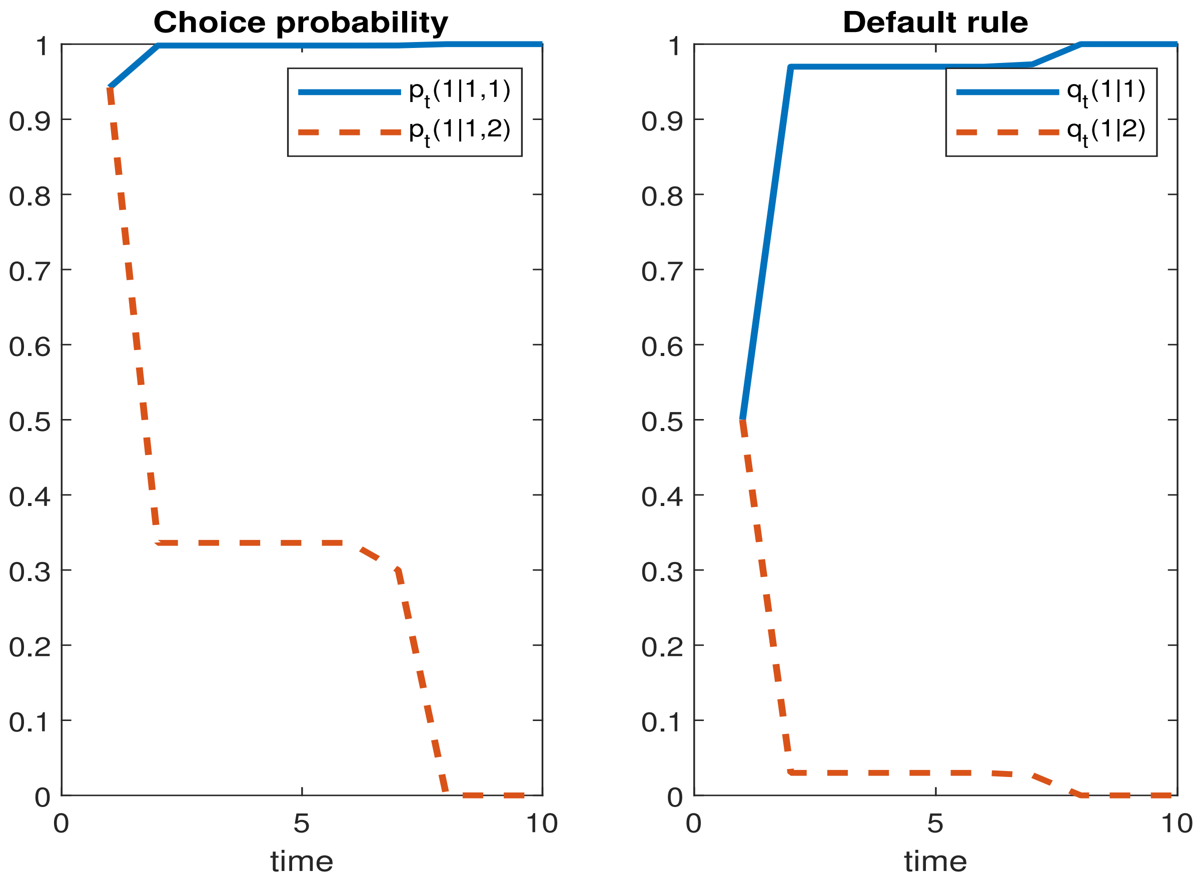

Figure 3: Choice probabilities and default rule for $T = 10$. Parameter values are $\mu_1(0) = 0.5, \pi(x_{t+1}|x_t, a_t) = \gamma = 0.03$ if $x_{t+1} \neq x_t$, $\beta = 0.8$, and $\lambda = 1$.

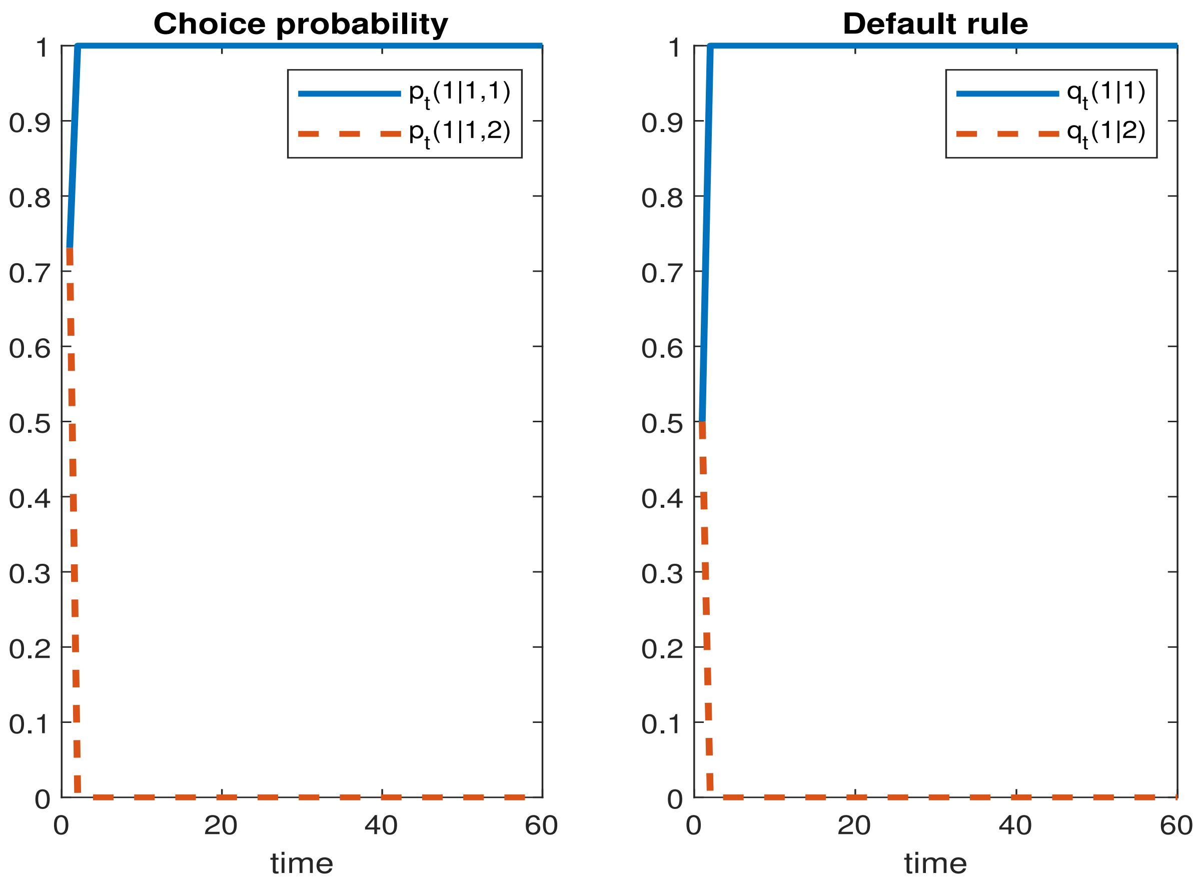

Figure 4: Choice probabilities and default rule for $T = \infty$. Parameter values are $\mu_1(0) = 0.5, \pi(x_{t+1}|x_t, a_t) = \alpha = 0.9$ if $x_{t+1} = a_t$, $\beta = 0.8$, and $\lambda = 1$.

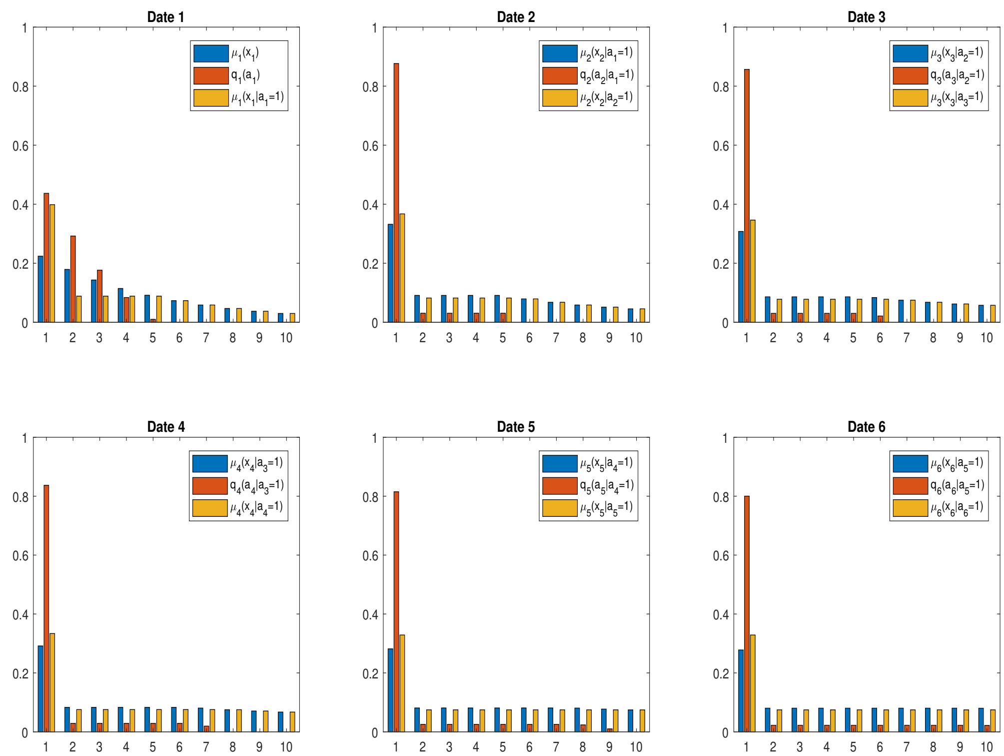

Figure 5: Choice probabilities and default rule for the model with $T = \infty$ and the state transition kernel being independent of actions.

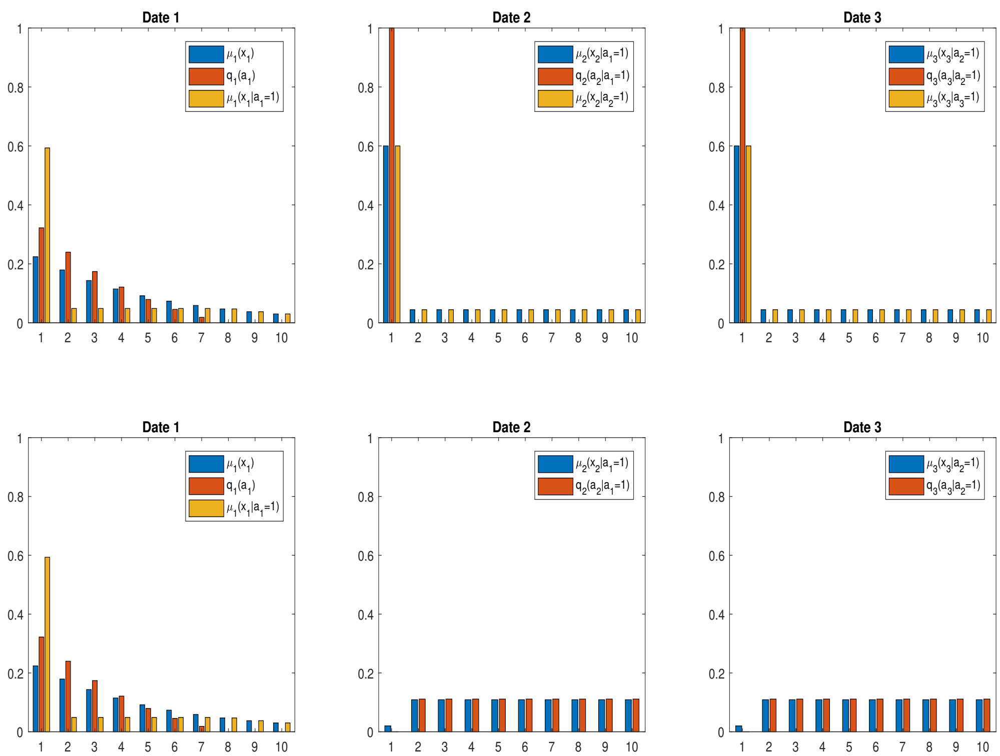

Figure 6: Choice probabilities and default rule for the model with $T = 3$ and the state transition kernel depending on actions. The top 3 panels are for the case of $\alpha = 0.6$, and the bottom three panels are for the case of $\alpha = 0.2$.

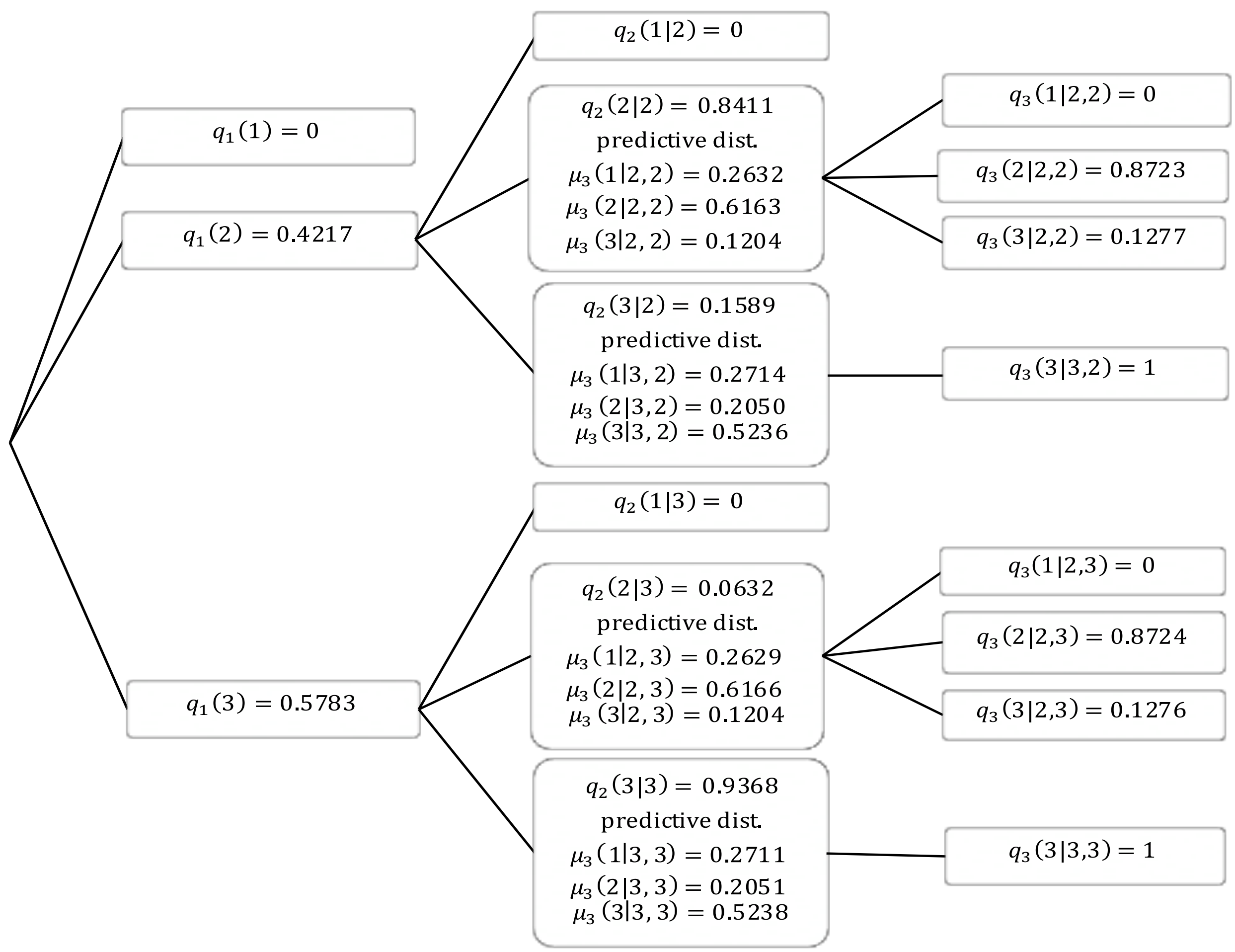

Figure 7: Markovian solution.

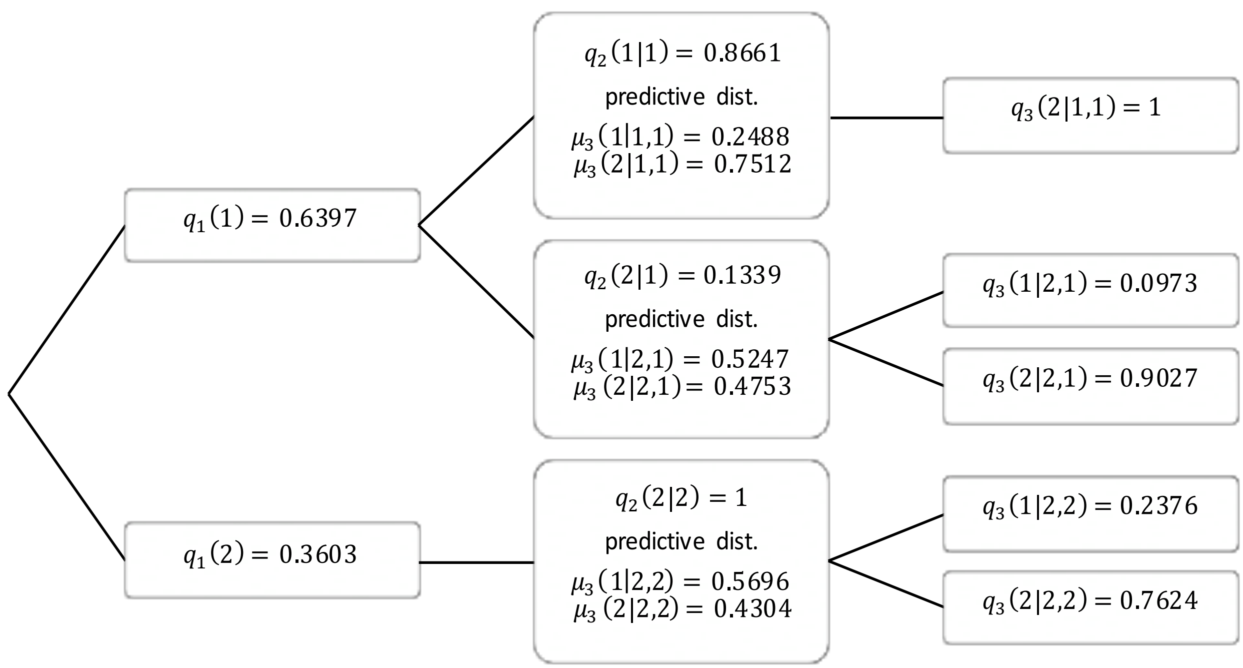

Figure 8: History-dependent solution.