Economic Theory
https://doi.org/10.1007/s00199-023-01512-1

RESEARCH ARTICLE

Dynamic discrete choice under rational inattention

Jianjun Miao[1] · Hao Xing[2]

Received: 3 August 2022 / Accepted: 17 July 2023

Abstract
We adopt the posterior-based approach to study dynamic discrete choice problems under rational inattention. We provide necessary and sufficient conditions to characterize the solution for general uniformly posterior-separable cost functions. We propose an efficient algorithm to solve these conditions and apply our model to explain phenomena such as perceptual distance, status quo bias, confirmation bias, and belief polarization. A key condition for our approach to work is the concavity of the difference between the generalized entropy of the current posterior and the discounted generalized entropy of the prior beliefs about the future states.

Keywords Rational inattention · Endogenous information acquisition · Entropy · Dynamic discrete choice · Dynamic programming

JEL Classification D11 · D81 · D83

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

We thank Andrew Caplin, Mark Dean, John Leahy, Bart Lipman, Stephen Morris, Jakub Steiner, Colin Stewart, and Mike Woodford for helpful discussions.

✉ Jianjun Miao
miaoj@bu.edu

Hao Xing
haoxing@bu.edu

[1] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA

[2] Department of Finance, Questrom School of Business, Boston University, 595 Commonwealth Ave, Boston, MA 02215, USA

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 a generalized entropy function of his beliefs. The DM's objective is to maximize the expected discounted utility less the utility cost of the information he acquires. We call this problem the dynamic RI problem.

The existing literature typically adopts the Shannon (1948) entropy cost function. Despite many appealing features of this specification, the experimental literature in economics and psychology establishes some behavior that violates key features of the Shannon model (see, e.g., Woodford (2012), Caplin and Dean (2013) (henceforth CD), Dean and Neligh (2019), and Dewan and Neligh (2020)). Motivated by this evidence, CD (2013), Caplin et al. (2022) (henceforth CDL)), Pomatto et al. (2023), Hébert and Woodford (2021), and Bloedel and Zhong (2021) propose more flexible cost functions. While they provide solutions in a static setup given these cost functions, how to extend their analysis to a dynamic setup is still an open question. The goal of our paper is to fill this gap.

We make three contributions to the literature. First, we characterize the solution to the dynamic RI problem using the posterior-based approach. To apply this approach, we focus on the class of uniformly posterior-separable (UPS) cost functions proposed by CD (2013) and CDL (2022). 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. The continuation value may not be concave in the revised prior beliefs following any history reached with positive probabilities.[1] 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.

Steiner et al. (2017) (henceforth SSM) solve the dynamic RI problem in the special case of the Shannon entropy using the choice-based approach. They first transform the problem into an unconstrained control problem and then take coordinate-wise first-order conditions to provide a dynamic logit characterization. We argue that this approach does not work for general UPS cost functions. 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.

We reduce the dynamic RI problem to a collection of interconnected static problems using the Bellman equation. The static problem in each period is to solve the

[1] We can show that it is actually convex for the Shannon entropy case, see Proposition 7 in Appendix A.

concavification of a collection of net utilities as functions of the posterior. Each net utility function consists of the current net utility and the continuation value. It is critical for all net utility functions to be concave for the concavification problem to be solvable. We show that the overall net utilities are concave under the assumption that the difference between the generalized entropy of the current posterior and the discounted generalized entropy of the prior belief about the future states is concave. This assumption reflects the convexity of the intertemporal information cost in a dynamic model. It is also important for us to establish a recommendation lemma similar to those in SSM (2017) and Ravid (2020), which states that the signal-based formulation is equivalent to our posterior-based formulation with actions as signals.

Under the above assumption, we provide a tractable first-order characterization for the dynamic RI problem using the result in CD (2013). This characterization gives necessary and sufficient conditions for optimal solutions. It is reduced to the dynamic logit characterization of SSM (2017) in the special Shannon entropy case.

Our second contribution is to propose a characterization of Markovian solutions and an efficient algorithm to find such a solution. For a Markovian solution, the predictive distribution of the next-period states depends only on the current action, the default rule (i.e., the distribution of actions conditioned on a history of actions) depends only on the last period action, and the choice rule (i.e., the distribution of actions conditioned on a history of both states and actions) depends only on the current state and the last period action. Our characterization generalizes that of SSM (2017) by allowing corner solutions and UPS cost functions.

Our algorithm extends the forward-backward Arimoto–Blahut algorithm of Tanaka et al. (2022) to infinite-horizon models with discounting and UPS cost functions. This algorithm is based on the Arimoto–Blahut algorithm for solving static channel capacity and rate distortion problems with Shannon entropy in information theory in the engineering literature (Arimoto (1972) and Blahut (1972)).

Our third contribution is to apply our theoretical results and numerical methods to solve some economic examples based on the perception task problem and the matching state problem often studied in the literature. We show that RI can help explain some phenomena documented in the psychology literature, such as perceptual distance, status quo bias, confirmation bias, and belief polarization. As a starting point, we prove that the dynamic RI solution is the same as the repeated static solution if the initial prior and the transition kernel are the same. For the static perception task problem we adopt the total information cost of Bloedel and Zhong (2021) and show that it can generate sigmoid-shaped choice responses unlike the step function for the Shannon entropy cost. This solution is repeated in the dynamic case if the preceding condition is satisfied.

For the matching state problem, we follow CD (2013) to adopt the Shorrocks (1980) entropy function that includes the Shannon entropy as a special case. We find that the status quo bias discussed by SSM (2017) does not arise when the decision horizon is sufficiently long. 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.

The Shorrocks entropy function incorporates a curvature parameter that affects the marginal information cost, thereby affecting both static and dynamic choice probabilities as well as the timing of choices. We find that the status quo bias can occur earlier and the confirmation bias is more likely to occur when the curvature parameter is larger because it induces a larger marginal information cost.

Our paper is closely related to CD (2013, 2015), Matějka and McKay (2015) (henceforth MM), SSM (2017), and CDL (2019, 2022). SSM (2017) is the first paper that extends the static model of MM (2015) to a dynamic setting and derives the dynamic logit rule.[2] Their solution method does not apply to the general UPS cost functions adopted in our paper. Our generalization permits us to study a wide range of economic and psychological behavior in a dynamic setting.

Our paper is also related to Hébert and Woodford (2018) and Zhong (2022), who adopt the posterior-based approach to study optimal stopping problems under RI with general information cost functions in the continuous-time setup.[3] Unlike their papers, ours is the first to study optimal control problems under RI where the concavity of the objective function is important for the optimality of the first-order conditions. More importantly, unlike their papers with fixed states over time, states in our model follow a Markov chain.

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), Miao and Su (2023), and Miao et al. (2022)). 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. (2019) 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 et al. (2018) for surveys and references cited therein.

2 Model

In this section we present the model setup, discounted information costs, and decision problems. We then establish an important recommendation lemma.

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 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 (probability) distributions on any finite set Z and $\Delta(Y|Z)$ to denote the set of conditional

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

[3] The posterior-based approach is often applied in the Bayesian persuasion literature. See Kamenica (2019) for a survey and the references cited therein.

distributions on any finite set Y given any $z \in Z$.[4] We also use a bold case letter to denote any random variable such as $\mathbf{x}$ with its realization denoted by a normal case letter x.

The decision maker (DM) makes choices from a finite action set denoted by A satisfying $|A| \geq 2$. For simplicity we assume that the action set A does not depend on the current state and has no identical elements. The state transition kernel is given by $\pi\left(x_{t+1}|x_t, a_t\right)$, which defines the probability of any state $x_{t+1} \in X$ given any state $x_t \in X$ and any action $a_t \in A$ for $t \geq 1$. 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 explicitly 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 $u : X \times A \to \mathbb{R}$. For the finite-horizon case with $T < \infty$, we allow u to be time dependent and include a terminal utility function $U : X \to \mathbb{R}$. SSM (2017) allow u to depend on the entire history of states and actions, which can generate history-dependent solutions.

Prior to choosing an action in any period $t \geq 1$, the DM can acquire costly information about the history of the state x^t, where we use x^t to denote the history $\{x_1, x_2, ..., x_t\}$ and x_k^t to denote the history $\{x_k, x_{k+1}, ..., x_t\}$ for $k < t$. More accurate information will lead to better choices, but are more costly, with information costs to be discussed later.

We first 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. By convention, set $s_0 = s^0 = \varnothing$. A strategy is a pair (d, σ) composed of

1. an information strategy d consisting of a system of signal distributions $d_t\left(s_t|x_t, s^{t-1}\right)$, for all $s^t \in S^t$, $x_t \in X$, 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 (d, σ) induce a sequence of joint distributions for x^{t+1} and s^t recursively

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

for $t \geq 1$, 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 prior/predictive distribution $\mu_t\left(x_t|s^{t-1}\right)$ and the posterior $\mu_t\left(x_t|s^t\right)$ for all t.

[4] As convention we define a conditional probability $P(C|B) = P(C \cap B)/P(B)$ whenever $P(B) > 0$; otherwise, set $P(C|B) = 0$, which does not affect our analysis, but simplifies notation.

The joint distribution $\mu_{T+1} \in \Delta\left(X^{T+1} \times S^T\right)$ constructed above induces an expected discounted utility value

$$\mathbb{E}\left[\sum_{t=1}^{T} \beta^{t-1} u\left(\mathbf{x}_t, \sigma_t\left(\mathbf{s}^t\right)\right) + \beta^T U\left(\mathbf{x}_{T+1}\right)\right],$$

where $\mathbb{E}$ denotes an expectation operator and $\beta \in (0, 1)$ denotes the subjective discount factor. In the next subsection, we define the information cost function.

2.2 Information costs

In a static setup we follow CD (2013) and CDL (2019, 2022) to define a UPS information cost function as follows[5]

$$C_H\left(\mu, \mu\left(\cdot|\cdot\right), q\right) \equiv H\left(\mu\right) - \sum_s q\left(s\right) H\left(\mu\left(\cdot|s\right)\right),$$

where $\mu \in \Delta\left(X\right)$ is a prior distribution, $\mu\left(\cdot|\cdot\right) \in \Delta\left(X|S\right)$ is a posterior, and $q \in \Delta\left(S\right)$ is a marginal distribution of signals that satisfies $\mu\left(x\right) = \sum_s q\left(s\right) \mu\left(x|s\right)$. Assume that H is a differentiable concave function on $\Delta\left(X\right)$ (called *generalized entropy*).[6] The term $H\left(\mu\right)$ measures the amount of prior uncertainty and the term $\sum_s q\left(s\right) H\left(\mu\left(\cdot|s\right)\right)$ measures the amount of uncertainty after acquiring information s. The concavity of H implies $C_H\left(\mu, \mu\left(\cdot|\cdot\right), q\right) \geq 0$ and the value of $C_H(\mu, \mu(\cdot|\cdot), q)$ represents the magnitude of uncertainty reduction by observing information $\mathbf{s}$ about the state $\mathbf{x}$.

Our approach applies to any differentiable and concave function H that is implied by all recent specifications of UPS cost functions such as the Shannon mutual information of Sims (2003), the Tsallis (1988) entropy cost of CDL (2022), the Shorrocks (1980) entropy cost of CD (2013) and Dean and Neligh (2019), the neighborhood-based cost of Hébert and Woodford (2021), and the total information cost of Bloedel and Zhong (2021). The Tsallis and Shorrocks entropy cost functions are useful to explain the behavioral responses to changing incentives (CD (2013)) and the last two cost functions are useful to explain the perceptual distance effect from the experimental evidence Dean and Neligh (2019).

In this paper we focus on (i) the Shorrocks entropy index

$$H\left(\nu\right) = \frac{1 - \sum_x \nu\left(x\right)^{2-\rho}}{\left(\rho - 1\right)\left(\rho - 2\right)}, \ \rho \neq 1, 2, \tag{1}$$

[5] See CDL (2022), Pomatto et al. (2023), Bloedel and Zhong (2021), and Denti (2022) for axiomatization of general information cost functions and foundations of UPS cost functions.

[6] For any finite distribution $\mu = \left(\mu_1, ..., \mu_M\right) \in \Delta\left(X\right)$, we write $H\left(\mu\right) = H\left(\mu_1, ..., \mu_M\right)$. We say H is differentiable on $\Delta\left(X\right)$ if $H\left(\mu_1, ..., \mu_M\right)$ is differentiable on the interior of the set $\left\{\left(\mu_1, ..., \mu_M\right) \in \mathbb{R}_+^M : \sum_{i=1}^M \mu_i = 1\right\}$.

where the Shannon entropy $H(\nu) = -\sum_x \nu(x) \ln \nu(x)$ is a special case as $\rho \to 1$, and (ii) the total information cost function with

$$H(\nu) = -\sum_{x,y \in X} \omega_{x,y} \nu(x) \ln \frac{\nu(x)}{\nu(y)}. \tag{2}$$

A tractable specification of $\omega_{x,y}$ is

$$\omega_{x,y} = \frac{1}{(x-y)^2} \text{ for } x \neq y; \ \omega_{x,y} = 0 \text{ for } x = y, \tag{3}$$

from Pomatto et al. (2023). Special cases of the total information cost function include the Wald cost function of Morris and Strack (2019) with two states and the Fisher information cost function of Hébert and Woodford (2021) in the continuous-state limit. The total information cost function is the only UPS cost function that exhibits the constant marginal cost property of Pomatto et al. (2023) and is a special case of their Bayesian log-likelihood ratio cost function.

In our dynamic setup, for the predictive distribution (prior belief) $\mu_t\left(\cdot|s^{t-1}\right)$ given history s^{t-1}, we define the conditional information cost in period t of acquiring information $\mathbf{s}_t$ about the state $\mathbf{x}_t$ as

$$\begin{aligned} &C_H\left(\mu_t\left(\cdot|s^{t-1}\right), \mu_t\left(\cdot|\cdot, s^{t-1}\right), q_t\left(\cdot|s^{t-1}\right)\right) \\ &= H\left(\mu_t\left(\cdot|s^{t-1}\right)\right) - \sum_{s_t} q_t\left(s_t|s^{t-1}\right) H\left(\mu_t\left(\cdot|s^t\right)\right), \end{aligned}$$

where the prior/predictive distribution $\mu_t\left(\cdot|s^{t-1}\right)$, the posterior distribution $\mu_t\left(\cdot|\cdot, s^{t-1}\right)$, and the one-step-ahead conditional distribution $q_t\left(\cdot|s^{t-1}\right)$ satisfy

$$\mu_t\left(x_t|s^{t-1}\right) = \sum_{s_t} q_t\left(s_t|s^{t-1}\right) \mu_t\left(x_t|s^t\right), \ t \geq 1, \tag{4}$$

and

$$\mu_{t+1}\left(x_{t+1}|s^t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t, \sigma_t\left(s^t\right)\right) \mu_t\left(x_t|s^t\right), \ t \geq 1. \tag{5}$$

Equation (4) shows that the posterior $\mu_t\left(x_t|s^t\right)$ weighted by $q_t\left(s_t|s^{t-1}\right)$ must average to the prior $\mu_t\left(x_t|s^{t-1}\right)$ given history s^{t-1}. Equation (5) shows that the prior $\mu_{t+1}\left(x_{t+1}|s^t\right)$ in the next period given history s^t is generated from the posterior $\mu_t\left(\cdot|s^t\right)$ and the state transition kernel π.

The unconditional information cost in period t of acquiring information $\mathbf{s}^t$ about the state $\mathbf{x}_t$ is defined as

$$\mathcal{I}\left(\mathbf{x}_t; \mathbf{s}_t|\mathbf{s}^{t-1}\right) \equiv \sum_{s^{t-1}} q_{t-1}\left(s^{t-1}\right) C_H\left(\mu_t\left(\cdot|s^{t-1}\right), \mu_t\left(\cdot|\cdot, s^{t-1}\right), q_t\left(\cdot|s^{t-1}\right)\right), \tag{6}$$

where

$$q_{t-1}\left(s^{t-1}\right)=q_0\left(s^0\right)q_1\left(s_1|s^0\right)q_2\left(s_2|s^1\right)\cdots q_{t-1}\left(s_{t-1}|s^{t-2}\right),\ q_0\left(s^0\right)\equiv 1.$$

The discounted information cost of acquiring information $\mathbf{s}^T$ about $\mathbf{x}^T$ is given by

$$\sum_{t=1}^{T}\beta^{t-1}\mathcal{I}\left(\mathbf{x}_t;\mathbf{s}_t|\mathbf{s}^{t-1}\right).$$

The undiscounted Shannon mutual information between $\mathbf{s}^T$ and $\mathbf{x}^T$ is the special case where $\beta=1$ and H is the Shannon entropy function.

2.3 Decision problems

We are ready to formulate the DM's decision problem:

Problem 1 (signal-based dynamic RI problem)

$$\max_{d,\sigma}\ \mathbb{E}\left[\sum_{t=1}^{T}\beta^{t-1}u\left(\mathbf{x}_t,\sigma_t\left(\mathbf{s}^t\right)\right)+\beta^T U\left(\mathbf{x}_{T+1}\right)\right]-\lambda\sum_{t=1}^{T}\beta^{t-1}\mathcal{I}\left(\mathbf{x}_t;\mathbf{s}_t|\mathbf{s}^{t-1}\right)$$

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

The parameter $\lambda>0$ measures the shadow price 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 signal to be statistically more dependent on the state, which generates a larger information cost.

We are interested in the DM's action choices. A (stochastic) *choice rule* $\{p_t\}$ is a sequence of distributions $p_t\left(a_t|x_t,a^{t-1}\right)$ over A conditional on x_t and a^{t-1} for $t\geq 1$, interpreted as the probability of choosing a_t given the state x_t at the history a^{t-1}. By convention, set $a_0=a^0=\varnothing$. We say that a strategy (d,σ) generates a choice rule $\{p_t\}$ 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 $\{p_t\}$ of the form $p_t\left(a_t|x_t,a^{t-1}\right)$ can induce a strategy (d,σ). Specifically, take any finite set S such that $|S|=|A|$. Fix any bijection $\phi:A\to S$, and for any t, let ϕ^t denote the mapping from A^t to S^t by applying ϕ coordinate-by-coordinate. Define

$$d_t\left(s_t|x_t,s^{t-1}\right)=d_t\left(\phi\left(a_t\right)|x_t,\phi^{t-1}\left(a^{t-1}\right)\right)=p_t\left(a_t|x_t,a^{t-1}\right),$$

$$\sigma_t\left(s^t\right) = \sigma_t\left(\phi^t\left(a^t\right)\right) = a_t.$$

Instead of solving the complicated Problem 1, we focus on a special class of information strategies in which signals correspond directly to actions. We replace a stochastic signal $\mathbf{s}_t$ by a stochastic action $\mathbf{a}_t$ with its realization denoted by a_t, but still keep other notations in Sect. 2.2 without risk of confusion. Following SSM (2017), we call $\{q_t\}$ a *default rule* , which is a sequence of distributions $q_t\left(a_t|a^{t-1}\right)$ that assign action probabilities at date t given history a^{t-1}. We then consider another problem related to Problem 1:

Problem 2 (posterior-based dynamic RI problem)

$$\max \ \mathbb{E}\left[\sum_{t=1}^{T}\beta^{t-1}u\left(\mathbf{x}_t,\mathbf{a}_t\right)+\beta^T U\left(\mathbf{x}_{T+1}\right)\right]-\lambda\sum_{t=1}^{T}\beta^{t-1}\mathcal{I}\left(\mathbf{x}_t;\mathbf{a}_t|\mathbf{a}^{t-1}\right), \tag{7}$$

where the choice variables are sequences of distributions $\left\{\mu_t\left(x_t|a^t\right)\right\}$ *and* $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$ *that satisfy (4) and (5) with* s_t *replaced by* a_t *for all* $t \geq 1$*, and the expectation is taken with respect to the joint distribution induced by* π, $\left\{\mu_t\left(x_t|a^t\right)\right\}$*, and* $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$.

To establish the equivalence of the above two problems, we introduce the following assumption for an important recommendation lemma or the revelation principle similar to Lemma 1 of SSM (2017) or Lemma 2 of Ravid (2020).

Assumption 1 For any $a \in A$, the function defined by

$$G^a\left(\nu\right) \equiv H\left(\nu\right) - \beta H\left(\sum_x \pi\left(\cdot|x,a\right)\nu\left(x\right)\right), \tag{8}$$

is concave in $\nu \in \Delta\left(X\right)$.

This assumption is specific to dynamic RI problems. It is trivially satisfied if $\beta = 0$, which corresponds to the static case. It is also satisfied for any concave H and any $\beta \in (0, 1)$ in the following two cases: (i) the transition kernel is independent of the past state, i.e., $\pi\left(\cdot|x,a\right)$ is independent of x for any $a \in A$, so that $\sum_x \pi\left(\cdot|x,a\right)\nu\left(x\right)$ is independent of ν; and (ii) the transition kernel is the identity kernel, i.e., $\pi\left(x'|x,a\right) = 1$ for $x = x'$; $= 0$ for $x' \neq x$, so that $G^a\left(\nu\right) = (1-\beta) H\left(\nu\right)$. In general, we have to take into account of discounting β and the state transition kernel π. The function $G^a\left(\nu\right)$ is equal to the difference of two concave functions and thus may not be concave. In the two-state case, $G^a\left(\nu\right)$ can be reduced to a univariate function and hence the concavity is easy to check. In particular, in Appendix A.4, we apply the Shorrocks entropy index and the total information cost function to check Assumption 1 in the two-state case. More generally, we show that Assumption 1 is satisfied for the Shannon entropy for any π and $\beta \in [0, 1]$.[7] In that appendix, we also give other sufficient conditions for Assumption 1.

[7] Miao et al. (2022) prove a related result in the linear-quadratic Gaussian framework.

The intuition behind Assumption 1 can be best understood in the two-period case with $T = 2$. Then we can write the discounted information cost as

$$\begin{aligned}
&\mathcal{I}\left(\mathbf{x}_1;\mathbf{a}_1\right)+\beta\mathcal{I}\left(\mathbf{x}_2;\mathbf{a}_2|\mathbf{a}_1\right)\\
&\quad=H\left(\mu_1\left(\cdot\right)\right)-\sum_{a_1}q_1\left(a_1\right)\left[H\left(\mu_1\left(\cdot|a_1\right)\right)-\beta H\left(\sum_x\pi\left(\cdot|x,a_1\right)\mu_1\left(x|a_1\right)\right)\right]\\
&\qquad-\beta\sum_{a^2}q_2(a^2)H(\mu_2(\cdot|a^2)).
\end{aligned}$$

The expression in the third line of the above equation describes the information cost reduction for $\mu_2(\cdot|a^2)$ chosen in period 2. In period 1, the DM chooses $\mu_1\left(\cdot|a_1\right)$ for each a_1 to be better informed, but incurs information costs. In a dynamic model, these costs are reflected by two components. The first is $H\left(\mu_1\left(\cdot|a_1\right)\right)$ which reduces the current prior generalized entropy $H\left(\mu_1\left(\cdot\right)\right)$. As the posterior $\mu_1\left(\cdot|a_1\right)$ today becomes the prior $\sum_x\pi\left(\cdot|x,a_1\right)\mu_1\left(x|a_1\right)$ in the next period after mixing with the transition kernel π, the discounted future generalized entropy $\beta H\left(\sum_x\pi\left(\cdot|x,a_1\right)\mu_1\left(x|a_1\right)\right)$ is the second component, which raises information costs. Then $G^{a_1}\left(\mu_1\left(\cdot|a_1\right)\right)$ captures the intertemporal information cost reduction for $\mu_1\left(\cdot|a_1\right)$. Intuitively, there is an intertemporal element of when to pay a cost that effectively balances (say) more learning today against less tomorrow.

The concavity of $H\left(\cdot\right)$ and $G^a\left(\cdot\right)$ ensures that the information cost reduction in each period is a concave function,[8] so that we can apply Jensen's inequality to show that restricting to a special class of information strategies $\left\{a^t\right\}$ does not raise the information cost. This is analogous to the data processing inequality for the Shannon entropy in information theory (see Theorem 2.8.1 and its corollary in Cover and Thomas (2006)). Intuitively, as $a_t=\sigma_t\left(s^t\right)$ is a function of data signals s^t, post-processing cannot increase information.

Lemma 1 *Let Assumption 1 hold. Then any strategy (d,σ) solving the dynamic RI Problem 1 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 2. 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 2 induce a strategy (d,σ) solving Problem 1.*

The intuition behind this lemma is as follows: Restricting to actions as information signals can achieve the same utility level without raising the information cost due to the 'data processing inequality' discussed earlier. Thus the maximal utility value net of information costs for Problem 2 is not smaller than that for Problem 1. On the other hand, as Problem 2 only focuses on a restricted class of information strategies, the associated maximal net utility value is not larger than that for Problem 1. As a result, the two problems can achieve the same maximum. The formal proof along with proofs for other results in the main text is in Appendix B.

By Lemma 1, we will focus on Problem 2 by solving sequences of optimal posteriors $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and default rules $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$. In the literature, there is another approach

[8] Equivalently, the discounted cost of information is convex in the choice variable, the posterior distribution.

that solves for the choice rules $\{p_t\}$ and default rules $\{q_t\}$ in the Shannon entropy case (e.g., MM (2015) and SSM (2017)). Using a static example in Appendix A.2, we show that such a choice-based approach does not apply to RI problems with UPS information costs.

3 Static case

Before presenting our main results for the dynamic case, starting with the static case will help us better understand the intuition behind our approach. The static solution also provides the foundation for the dynamic solution.

When $T = 1$ and $U = 0$, Problem 2 is reduced to the following static RI problem:

Problem 3 (static RI problem with UPS cost)

$$V(\mu) \equiv \max_{q\in\Delta(A),\mu(\cdot|\cdot)\in\Delta(X|A)} \mathbb{E}\left[u(\mathbf{x},\mathbf{a})\right] - \lambda C_H\left(\mu,\mu(\cdot|\cdot),q\right)$$

subject to

$$\mu(x) = \sum_a \mu(x|a)\, q(a), \;\; x \in X. \tag{9}$$

Following CD (2013) and CDL (2019), we rewrite this problem as

$$\overline{V}(\mu) \equiv \max_{q\in\Delta(A),\mu(\cdot|\cdot)\in\Delta(X|A)} \sum_a q(a)\, N_H^a\left(\mu(\cdot|a)\right), \quad V(\mu) = \overline{V}(\mu) - \lambda H(\mu) \tag{10}$$

subject to (9), where $N_H^a\left(\mu(\cdot|a)\right)$ denotes the net utility of action a defined as

$$N_H^a\left(\mu(\cdot|a)\right) \equiv \sum_x \mu(x|a)\, u(x,a) + \lambda H\left(\mu(\cdot|a)\right), \tag{11}$$

and $\overline{V}(\mu)$ is the concave envelope of net utilities for all actions.

CD (2013) and CDL (2022) provide a first-order characterization for the solution using a geometric approach from convex analysis. It is challenging to generalize their characterization in the static case to the dynamic case. We now show that their characterization can be simplified to a form that allows such an extension and facilitates computation in applications. In particular, we show that their three first-order conditions are equivalent to two conditions, one for chosen actions and the other for unchosen actions.

Before presenting our conditions, we define a function:

$$f(\nu) = H(\nu) - \sum_{x=1}^{M} H_x(\nu)\nu(x), \quad \nu \in \Delta(X), \tag{12}$$

where $H_x(\nu)$ denotes the partial derivative $\partial H(\nu)/\partial\nu(x)$ for $x = 1, ..., M$, without the restriction $\sum_x \nu(x) = 1$. Note that we view $H(\nu)$ as a function $H(\nu(1), \ldots, \nu(M))$.

Definition 1 The pair of $q \in \Delta(A)$ and $\mu(\cdot|\cdot) \in \Delta(X|A)$ satisfies

(i) FOC-CA, if for any chosen actions $a, b \in A$ with $q(a) > 0$ and $q(b) > 0$,

$$u(x,a) + \lambda H_x(\mu(\cdot|a)) + \lambda f(\mu(\cdot|a)) = u(x,b) + \lambda H_x(\mu(\cdot|b)) + \lambda f(\mu(\cdot|b)), \tag{13}$$

for any $x \in X$, where this common value is defined as $\widehat{V}(x)$;

(ii) FOC-UA, if for any unchosen action $b \in A$ with $q(b) = 0$ and for $\mu^b \in \Delta(X)$ and $x \in X$ such that

$$u(x,b) + \lambda H_x\left(\mu^b\right) - \left[u(M,b) + \lambda H_M\left(\mu^b\right)\right] = \widehat{V}(x) - \widehat{V}(M), \tag{14}$$

we have

$$\sum_x I_x\left(\widehat{V}(x)/\lambda - u(x,b)/\lambda - f(\mu^b); \mu^b\right) \leq 1.^{9} \tag{15}$$

Condition (i) derives from the fact that the net utility functions of the chosen actions must support the same hyperplane at their associated posteriors. It says that posterior probabilities of a given state x following two chosen actions a and b depend on relative payoffs of the two actions in that state and possibly information costs in other states. States other than x do not matter for the Shannon entropy case with $H_x(\mu(\cdot|a)) = -\ln \mu(x|a) - 1$ and $f(\cdot) = 1$. Condition (ii) derives from the fact that the net utility functions of all unchosen actions must lie weakly below this hyperplane.

To better understand (15), suppose that action b is chosen with $q(b) > 0$. By the definition of $\widehat{V}(x)$ and Eq. (13), Eq. (14) is satisfied with μ^b equal to the posterior distribution $\mu(\cdot|b)$. By the definition of $\widehat{V}(x)$, we also have

$$I_x\left(\widehat{V}(x)/\lambda - u(x,b)/\lambda - f(\mu(\cdot|b)); \mu(\cdot|b)\right) = \mu(x|b).$$

Therefore inequality (15) is satisfied as an equality because $\mu(\cdot|b)$ is a probability distribution. When b is not chosen with $q(b) = 0$, the posterior $\mu(\cdot|b)$ is not well defined. Then for $\mu^b \in \Delta(X)$ satisfying (14), (15) holds as an inequality.

Condition (15) is analogous to condition (3) of CDL (2019) for the Shannon entropy case. CDL (2019) argue that their condition (3) plays an important role in the convergence of the Arimoto–Blahut algorithm. Our (15) plays a similar role in our algorithm presented in Appendix C.1.

In Appendix B, we prove that condition FOC-CA is equivalent to conditions (ED) and (CT) in Lemma 3 of CD (2013), and condition FOC-UA is equivalent to condition (UB) in that lemma. Because conditions (ED), (CT), and (UB) are necessary and sufficient for the optimality in the static RI problem, FOC-CA and FOC-UA are equivalent necessary and sufficient conditions as well. See also the equivalent Lagrangian lemma in CDL (2022) Lemma 1.

[9] Here the function $I_x(\cdot;\nu) : \mathbb{R} \to [0,1]$ is the inverse function of $H_x(\nu)$ at its x^{th} component defined as

$$y = H_x(\nu(1), \ldots, I_x(y;\nu), \ldots, \nu(M)),$$

where the inverse function exists because $H_x(\nu(1), \ldots, \nu(x), \ldots, \nu(M))$ is decreasing in $\nu(x)$ due to the concavity of H. Note that $I_x(\cdot;\nu)$ may depend on $\nu\left(x'\right)$ for any $x' \neq x$.

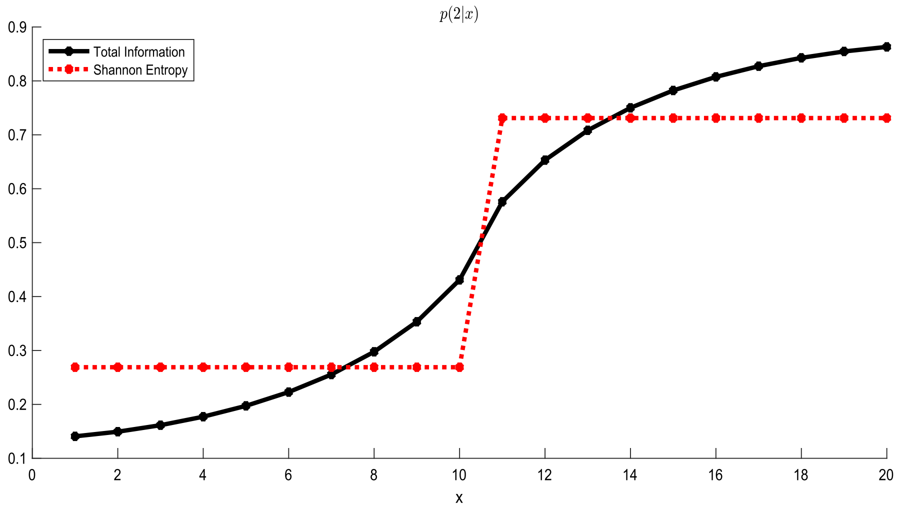

Fig. 1 Choice probabilities for the total information cost and Shannon entropy cost

Proposition 1 *The pair of* $\mu(\cdot|\cdot)$ *and* q *satisfying* (9) *is optimal for Problem* 3 *if and only if conditions FOC-CA and FOC-UA hold. Moreover, the value function* V *satisfies*

$$V(\mu) = \overline{V}(x) - \lambda H(\mu) = \sum_x \mu(x)\widehat{V}(x) - \lambda H(\mu). \tag{16}$$

Because $V(\mu)$ is equal to the difference of two concave functions $\overline{V}(x)$ and $\lambda H(\mu)$, it is unclear whether $V(\mu)$ is concave. In Proposition 7 of Appendix A, we show that $V(\mu)$ is convex for the Shannon entropy case. We also provide a geometric characterization of the function $\widehat{V}(x)$ and establish some properties that are useful for analyzing the dynamic RI problem. The function $\widehat{V}(x)$ can also be interpreted as the realized net value function after prior uncertainty is resolved. The connection of the definition of $\widehat{V}(x)$ in Definition 1 and its alternative characterizations are our new discovery in this paper.

Now we apply our first-order conditions to compute an example with the total information cost (2) and (3), that can help explain the perceptual distance phenomenon. Suppose that the prior is uniform, M is an even number, and there are two actions $a = 1, 2$. The payoff satisfies: $u(x, 2) = 1$ if $x > M/2$, $u(x, 1) = 1$ if $x \leq M/2$, and $u(x, a) = 0$, otherwise. Figure 1 plots the choice probabilities $p(a = 2|x)$ against the state x for $M = 20$ and $\lambda = 1$. As in Pomatto et al. (2023), this figure shows that the choice probabilities for the total information cost have a sigmoid shape similar to psychometric curves, unlike a step function for the Shannon entropy cost.

4 Main results

In this section we analyze dynamic RI problems using dynamic programming, provide necessary and sufficient conditions for optimality, and describe a numerical algorithm to solve these conditions.

4.1 Dynamic programming

To study Problem 2, let $V^T(\mu_1)$ denote the maximized value in (7) for $T < \infty$. Choosing the predictive/prior distribution as the state variable, we obtain a value function $V^T : \Delta(X) \to \mathbb{R}$. For $T = \infty$, we simply use $V(\cdot)$ to denote the corresponding value function.

Endow $\Delta(X)$ with the weak topology. Let $\mathbb{V}$ denote the set of all continuous functions on $\Delta(X)$. Then $\mathbb{V}$ is a Banach space. Define a Bellman operator $\mathcal{T}$ on this space as

$$\begin{aligned}\mathcal{T}v(\mu) = \max_{\mu_p \in \Delta(X|A), q \in \Delta(A)} & \sum_{x,a} q(a)\mu_p(x|a)u(x,a) - \lambda C_H\left(\mu, \mu_p, q\right) \\ & + \beta \sum_a q(a) v\left(\mu'(\cdot|a)\right) \end{aligned} \tag{17}$$

subject to

$$\mu(\cdot) = \sum_a q(a)\mu_p(\cdot|a), \tag{18}$$

$$\mu'(\cdot|a) = \sum_x \pi(\cdot|x,a)\mu_p(x|a). \tag{19}$$

Analogous to (9) in the static case, Eq. (18) shows that the posterior distribution μ_p weighted by q must average to the prior distribution μ. Equation (19) describes the law of motion for the state variable. The state variable $\mu'(\cdot|a)$ in the next period is generated from the posterior μ_p and the state transition kernel π.

By the principle of optimality, V and $\{V^s\}_{s=1}^T$ satisfy the Bellman equations $V = \mathcal{T}V$, $V^s = \mathcal{T}V^{s-1}$, and

$$V^0(\mu) = \sum_x \mu(x)U(x) \text{ for } T < \infty.$$

Proposition 2 *For dynamic RI Problem 2 with $T = \infty$, there exists a unique function V on $\mathbb{V}$ that satisfies the Bellman equation $V = \mathcal{T}V$. Moreover,* $\lim_{T\to\infty} V^T(\mu) = V(\mu)$ *for any $\mu \in \Delta(X)$.*

After deriving an optimal policy function from the Bellman equation, we can generate sequences of optimal posteriors $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and default rules $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$

in a standard way. Unfortunately, both the infinite- and finite-horizon dynamic programming problems are difficult to analyze. The optimization problem in (17) is also difficult to solve, because the value function may not be concave as discussed in the static case. In particular, the first-order conditions may not be sufficient. The standard value function iteration method is inefficient and inaccurate because one has to discretize the state space and optimize globally over a feasible set of $\mu_p(\cdot|a)$ and $q(a)$ for all a that satisfy constraint (18). To the best of our knowledge, we have not seen the implementation of the value function iteration method in the RI literature.

4.2 First-order characterization

In this subsection we use our dynamic programming equation to provide a first-order characterization for an optimal solution and establish sufficiency conditions. The idea is to work backward and treat the dynamic RI problem as a sequence of two-period (or essentially static) problems. We can then apply the analysis in the static case of Sect. 3.

There are two crucial differences from the static first-order conditions by considering future utility and information costs: (i) the function f in (12) is replaced by a dynamic version f_t; (ii) the one-period utility function u is replaced by continuation utility v_t.

Specifically, f_t is defined as

$$f_t(\nu, a) \equiv H(\nu) - \sum_x \nu(x)\ H_x(\nu) - \beta \left[H(\tilde{\nu}(\cdot|a)) - \sum_x \tilde{\nu}(x|a) H_x(\tilde{\nu}(\cdot|a)) \right], \tag{20}$$

for $\nu \in \Delta(X)$, $a \in A$, $\tilde{\nu}(\cdot|a) = \sum_x \pi(\cdot|x, a)\nu(x)$, $t = 1, ..., T-1$, and $f_T(\nu, a) \equiv H(\nu) - \sum_x \nu(x) H_x(\nu)$. In the final period $T < \infty$, the problem is static and f_T is the same as in (12).

Continuation utility v_t is defined as

$$\begin{aligned} v_t(x_t, a^t) &= u(x_t, a_t) \\ &+ \beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t) \left[\widehat{V}_{t+1}(x_{t+1}|a^t) - \lambda H_{x_{t+1}}(\mu_{t+1}(\cdot|a^t)) \mathbf{1}_{\{t<T\}} \right], \end{aligned} \tag{21}$$

for any a_t with $q_t\left(a_t|a^{t-1}\right) > 0$ and for $t = 1, \ldots, T$, with $\widehat{V}_{T+1}(x_{T+1}|a^T) = U(x_{T+1})$, $x_{T+1} \in X$.[10] The function $\widehat{V}_{t+1}(\cdot|a^t) : X \to \mathbb{R}$ corresponds to $\widehat{V}$ in (16) and must be jointly solved from the system of first-order conditions, For the dynamic RI problem, continuation utility depends on the future realized net value function $\widehat{V}_{t+1}(x_{t+1}|a^t)$ and the future predictive distribution $\mu_{t+1}(\cdot|a^t)$.

We now introduce the system of first-order conditions for the dynamic case.

Definition 2 The sequences of $\{\mu_t(\cdot|a^t)\}_{t=1}^T$ and $\{q_t(\cdot|a^{t-1})\}_{t=1}^T$ satisfy

[10] $\mathbf{1}_{\{t<T\}}$ is an indicator function taking value 1 if $t < T$ and 0 if $t = T$,

(i) FOC$_t$-CA, if for any chosen action $a_t \in A$ with $q_t(a_t|a^{t-1}) > 0$ and for any $x_t \in X$,

$$\widehat{V}_t(x_t|a^{t-1}) = v_t(x_t, a^t) + \lambda H_{x_t}(\mu_t(\cdot|a^t)) + \lambda f_t(\mu_t(\cdot|a^t), a_t); \tag{22}$$

(ii) FOC$_t$-UA, if for any unchosen action $a_t \in A$ with $q_t\left(a_t|a^{t-1}\right) = 0$ and $\mu_t^{a_t} \in \Delta(X)$ such that

$$\left[v_t(x_t, a^t) + \lambda H_{x_t}(\mu_t^{a_t})\right] - \left[v_t(M, a_t) + \lambda H_M(\mu_t^{a_t})\right] = \widehat{V}_t\left(x_t|a^{t-1}\right) - \widehat{V}_t\left(M|a^{t-1}\right), \tag{23}$$

for any $x_t \in X$, we have

$$\sum_{x_t} I_{x_t}\left(\widehat{V}_t(x_t|a^{t-1})/\lambda - v_t(x_t, a^t)/\lambda - f_t(\mu_t^{a_t}, a_t); \mu_t^{a_t}\right) \leq 1. \tag{24}$$

We are ready to present our first main result:

Theorem 1 *Suppose that Assumption 1 holds. Then the sequences of* $\{q_t(a_t|a^{t-1})\}_{t=1}^T$ *and* $\{\mu_t(x_t|a^t)\}_{t=1}^T$ *are the optimal solution to dynamic RI Problem 2 if and only if they satisfy (i) Eqs. (4) and (5) with* s *replaced by* a*; and (ii) conditions FOC*$_t$*-CA and FOC*$_t$*-UA for* $t = 1, \ldots, T$*. The value function given history* a^{t-1} *reached with positive probabilities satisfies*

$$V^{T+1-t}(\mu_t(\cdot|a^{t-1})) = \sum_{x_t} \mu_t(x_t|a^{t-1})\widehat{V}_t(x_t|a^{t-1}) - \lambda H(\mu_t(\cdot|a^{t-1})),\ t = 1, ..., T. \tag{25}$$

Assumption 1 is critical not only for the recommendation lemma, but also for the applicability of the posterior-based approach and the sufficiency of the first-order conditions. As discussed in Sect. 2.3, the choice of posterior $\mu_t\left(\cdot|a^t\right)$ affects information costs in both periods t and $t+1$ because the posterior $\mu_t\left(\cdot|a^t\right)$ becomes the prior $\mu_{t+1}\left(\cdot|a^t\right)$ after mixing with the transition kernel π. Assumption 1 ensures the concavity of the intertemporal entropy cost reduction and hence the concavity of the net utility in $\mu_t\left(\cdot|a^t\right)$ given history a^{t-1}. We can then compute the concave envelope of the net utilities to solve the RI problem as in CD (2013) and CDL (2019, 2022). We illustrate this point in a two-period RI problem in Appendix A.3.

The modified Eqs. (4) and (5) correspond to (18) and (19) and describe the Bayes consistency of beliefs and their evolution over time. The first-order conditions FOC$_t$-CA and FOC$_t$-UA resemble their static counterpart, with the utility function u and the function f replaced by v_t and f_t, respectively. It is crucial to recognize that these first-order conditions are coupled across time: first-order conditions at time t and the resulting optimal posterior $\mu_t(\cdot|a^t)$ depend on the continuation utility v_t, which takes into account the future realized value $\widehat{V}_{t+1}$ and the predictive distribution $\mu_{t+1}(\cdot|a^t)$ for the next period; meanwhile the realized value $\widehat{V}_{t+1}$ and the predictive distribution $\mu_{t+1}(\cdot|a^t)$ at time $t+1$ depend on the optimal posterior $\mu_t(\cdot|a^t)$ at time t. All of the first-order conditions across time must be solved jointly rather than separately. This coupling among the first-order conditions also motivates our algorithm in Sect. 4.4.

Equation (25) corresponds to (16) in the static case. This form of the value function allows us to combine the continuation value with the current utility u and information costs to compute the current net utility in the Bellman equation. We can then apply the static analysis in Proposition 1.

For the Shannon entropy case with $H(\nu) = -\sum_x \nu(x) \ln \nu(x)$, we can verify that Assumption 1 is satisfied so that we can apply Theorem 1 to derive a dynamic logit characterization, which is first obtained by SSM (2017) using the choice-based approach. Unlike their result, we also provide explicit sufficiency conditions. Moreover, as shown in Appendix A.2, their choice-based approach does not work for general UPS cost functions beyond the Shannon entropy.

4.3 Markovian solution

Although our first-order characterization is useful for numerical analysis, it is still complicated and even infeasible to compute for the infinite-horizon case, because the solution is generally history dependent. For example, the optimal posterior $\mu_t(\cdot|a^t)$ and default rule $q_t\left(\cdot|a^{t-1}\right)$ may depend on the entire history of actions a^{t-1}. To simplify dynamic solutions, we turn to a Markovian characterization. We define the following notion of Markovian solution.

Definition 3 An optimal solution to dynamic RI Problem 2 is Markovian if, for any two different histories a^{t-1} and $\left\{b^{t-2}, a_{t-1}\right\}$ reached with positive probabilities and any $t = 1, 2, ..., T$, 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 independent of history of actions, then there is an optimal solution that is also independent of history of actions. We thus have our second main result.

Theorem 2 *For a Markovian solution to dynamic RI Problem 2, the posterior distribution* $\mu_t\left(x_t|a^t\right)$, *the choice rule* $p_t(a_t|x_t, a^{t-1})$, *and the default rule* $q_t(a_t|a^{t-1})$ *at any history* a^{t-1} *reached with positive probabilities take the form of* $\mu_t\left(x_t|a^t_{t-1}\right)$, $p_t(a_t|x_t, a_{t-1})$, *and* $q_t(a_t|a_{t-1})$, *respectively, for any* $t = 1, ..., T$.

By this theorem, we simply write the Markovian solution for $\mu_{t+1}\left(x_{t+1}|a^t\right)$, $\mu_t\left(x_t|a^t\right)$, and $q_t\left(a_t|a^{t-1}\right)$ as $\mu_{t+1}(x_{t+1}|a_t)$, $\mu_t\left(x_t|a^t_{t-1}\right)$, and $q_t(a_t|a_{t-1})$, respectively. We can then modify (4) and (5) as

$$\mu_t(x_t|a_{t-1}) = \sum_{a_t} \mu_t(x_t|a^t_{t-1}) q_t(a_t|a_{t-1}), \tag{26}$$

$$\mu_{t+1}(x_{t+1}|a_t) = \sum_{x_t} \pi(x_{t+1}|x_t, a_t)\, \mu_t\left(x_t|a^t_{t-1}\right), \tag{27}$$

for any a_{t-1} leading to a_t with positive probabilities.

We say that a solution to the dynamic RI problem in (7) is interior if $q_t(a_t|a^{t-1}) > 0$ for any action $a_t \in A$ and any history a^{t-1}, $t \geq 1$. The following result shows that any interior solution is Markovian.

Proposition 3 *Every interior solution to dynamic RI Problem 2 is Markovian, for which the optimal posterior distribution $\mu_t\left(x_t|a^t\right)$ takes the form $\mu_t\left(x_t|a_t\right)$ for any $t \geq 1$.*

This result relies on the locally invariant posteriors property discovered by CD (2013), which implies that the optimal posteriors $\mu_t\left(\cdot|a^t\right)$ for all chosen actions a_t with $q_t(a_t|a^{t-1}) > 0$ are independent of the prior $\mu_t\left(\cdot|a^{t-1}\right)$ in the convex hull of these posteriors. Thus $\mu_t\left(\cdot|a^t\right)$ for each a_t is independent of a^{t-1}. It follows from forward induction that the prior $\mu_{t+1}\left(\cdot|a^t\right)$ is also independent of a^{t-1}. We then obtain a Markovian solution for which $\mu_t\left(x_t|a^t\right)$ is also independent of a^{t-1}.

Combining Theorems 1 and 2, we can provide necessary and sufficient first-order conditions for a Markovian solution, which may not be interior, using the posterior-based approach (see Proposition 8 in Appendix D). This result generalizes Proposition 3 and Lemma 6 of SSM (2017) for the Shannon entropy case to allow corner solutions and UPS cost functions. Notice that not every dynamic RI problem admits an optimal Markovian solution. We have solved numerical examples to illustrate this point for the Shannon entropy case. We have solved numerical examples to illustrate this point for the Shannon entropy case in Appendix D.

4.4 Numerical methods

In the Markovian case, we have 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 Tanaka et al. (2022) to our dynamic RI model with UPS costs. We present the algorithm in Appendix C. Here we sketch the key idea.

For any given prior $\mu_1(x_1)$, we compute the sequence of prior distributions $\{\mu_{t+1}(x_{t+1}|a_t)\}$ forward starting from $\mu_1(x_1|a_0) = \mu_1(x_1)$. Because the terminal value $\widehat{V}_{T+1}(x_{T+1}|a^T) = U(x_{T+1})$ is given in the finite-horizon case, we compute the sequence of values $\{\widehat{V}_t(x_t|a_{t-1})\}$ backward starting from the terminal time. All other variables are jointly computed along the forward and backward paths. Taking limits as $T \to \infty$, we obtain the infinite-horizon solution.

To facilitate numerical implementation, we replace (27) with the following equations:

$$\mu_{t+1}(x_{t+1}, a_t) = \sum_{x_t, a_{t-1}} \pi(x_{t+1}|x_t, a_t)\, \mu_t(x_t|a_{t-1}^t) q_t(a_t|a_{t-1}) q_{t-1}(a_{t-1}),\ q_0(a_0) = 1, \tag{28}$$

$$\mu_{t+1}(x_{t+1}|a_t) = \frac{\mu_{t+1}(x_{t+1}, a_t)}{q_t(a_t)},\ q_t(a_t) = \sum_{x_{t+1}} \mu_{t+1}(x_{t+1}, a_t) > 0. \tag{29}$$

Notice that the two $\mu_{t+1}(x_{t+1}|a_t)$ defined in (29) and (27) may not be identical even if we use the same notation. If we use Eq. (27) to compute $\mu_{t+1}(x_{t+1}|a_t)$, then the expression on the right-hand side of (27) may depend on a_{t-1} during iterations and thus the algorithm may fail to converge to an Markovian solution. By contrast, Eqs. (28) and (29) ensure that a_{t-1} is averaged out during each iteration.

If the solution for sequences of posteriors $\left\{\mu_t\left(x_t|a_{t-1}^t\right)\right\}$ and conditional default rules $\{q_t(a_t|a_{t-1})\}$ is known, Eqs. (28) and (29) can be solved forward in time to obtain a sequence of predictive distributions $\{\mu_{t+1}(x_{t+1}|a_t)\}$. On the other hand, if the solution for $\{\mu_{t+1}(x_{t+1}|a_t)\}$ is known, the Markovian version of Eqs. (21) and (22), which may be viewed as Bellman equations, can be solved backward in time. At each time t we use the static algorithm described in Appendix C to solve for $\left\{\mu_t\left(x_t|a_{t-1}^t\right)\right\}$ and $\{q_t(a_t|a_{t-1})\}$. We solve all equations iteratively until convergence and check whether (27) is satisfied.

Clearly, $\mu_{t+1}(x_{t+1}|a_t)$ defined in (27) satisfies (29). But conversely $\mu_{t+1}(x_{t+1}|a_t)$ defined in (29) may not satisfy (27) unless $\mu_t(\cdot|a_{t-1}^t)$ is the same for any two different actions a_{t-1} and b_{t-1} reaching the same a_t or there is only one action leading to a_t. The following proposition gives sufficient conditions to ensure that our algorithm generated Markovian solution is optimal and this solution can be either interior or at the corner.

Proposition 4 *If the algorithm converges to a limiting solution such that for any* $a_t \in A$, $t \geq 2$ *one of the following two conditions is satisfied: (i)* $\mu_t(x_t|a_t, a_{t-1}) = \mu_t(x_t|a_t, b_{t-1})$ *for any* $a_{t-1} \neq b_{t-1}$ *with* $q_t(a_t|a_{t-1}) > 0$ *and* $q_t(a_t|b_{t-1}) > 0$; *(ii) there exists a unique* a_{t-1} *such that* $q_t(a_t|a_{t-1}) = 1$ *and* $q_t(a_t|b_{t-1}) = 0$ *for any* $b_{t-1} \neq a_{t-1}$, *then it is an optimal Markovian solution.*

We will use the above algorithm to solve some numerical examples in the next section. By Proposition 4, all numerical solutions for these examples are optimal Markovian solutions. We can design a similar algorithm for the history-dependent solution in Theorem 1. This algorithm becomes complicated for long-horizon problems as the history increases with the horizon and becomes infeasible under infinite horizon.

5 Applications

To apply our theoretical results, we first establish that the dynamic solution is the repeated static solution if the transition kernel and the initial prior are uniform in Sect. 5.1. We then apply this result by extending the static perception task problem in Sect. 3 to a dynamic setting. Next we study a dynamic matching state problem often studied in the literature (SSM (2017), CD (2013) and CDL (2019)). This problem can be used to describe many economic decisions, e.g., consumer choices, project selection, and job search. We assume that the transition kernel is independent of actions in Sect. 5.2 and allow it to depend on actions in Sect. 5.3. For simplicity, set $|X| = |A| = 2$, $\mu_1(x_1 = 1) = 0.5$, $u(x_t, a_t) = 1$ if $x_t = a_t$; and $u(x_t, a_t) = 0$, otherwise, throughout Sects. 5.2 and 5.3.[11] We adopt the Shorrocks entropy in (1) and compare the solution with that for the Shannon entropy.[12] We will show how the parameter ρ affects behavioral responses to incentives in a dynamic setting.

5.1 Repeated static solution

It is nontrivial to derive analytical results for general dynamic RI problems. The simplest possible dynamic solution is the repeated static solution in the sense that $p_t\left(a_t|x_t, a^{t-1}\right) = p_1\left(a_t|x_t\right),\ q_t\left(a_t|a^{t-1}\right) = q_1\left(a_t\right)$, and $\mu_t\left(x_t|a^t\right) = \mu_1\left(x_t|a_t\right)$ for any $t \geq 1$, where $p_1\left(\cdot|\cdot\right), q_1\left(\cdot\right)$, and $\mu_1\left(\cdot|\cdot\right)$ are the static solution for $T = 1$. We are able to derive sufficient conditions for such a solution.

Proposition 5 *Suppose that $U = 0$ and the transition kernel $\pi\left(\cdot|x, a\right)$ is the same as the initial prior $\mu_1\left(\cdot\right)$ for any x and a. Then any solution to the dynamic RI Problem 2 is the repeated static solution.*

The assumption in the proposition implies that future states are drawn independently and identically from the same initial prior μ_1.[13] Then the prior beliefs about the next-period state given any history reached with positive probabilities is the same as the initial prior μ_1, independent of the current choice variable. Because the prior beliefs are the only state variable for the dynamic RI problem, the continuation value does not depend on the current choice variable. By dynamic programming, the solution in any period only depends on the current payoff, independent of the future continuation value.

We now apply Proposition 5 by extending the static perception task problem in Sect. 3 to a dynamic setting. For simplicity, we take $\pi\left(x'|x, a\right) = \mu_1\left(x'\right) = 1/M$ for any x'. Then the dynamic solution is the repeated static solution illustrated in Fig. 1. This figure shows that the dynamic logit solution for the choice probabilities in any period for the Shannon entropy cost is a step function of the current state. By contrast, the dynamic solution for the total information cost has a sigmoid shape.

5.2 Transition kernel independent of actions

In the remaining two subsections, we study the matching state problem. Assuming $\pi\left(x_{t+1}|x_t, a_t\right) = \gamma$ whenever $x_{t+1} \neq x_t$ for any $a_t \in A$, we use this example to illustrate that rationally inattentive behavior exhibits status quo bias over a short horizon, but not over an infinite horizon. Moreover, the infinite-horizon behavior exhibits inertia.[14] As a benchmark, the optimal solution for the case without information cost ($\lambda = 0$) is to choose an action to match the state in each period.

[11] Our algorithm works for larger state and action spaces. Computation codes for all numerical examples are available upon request.

[12] We have verified that Assumption 1 is satisfied for all our numerical examples in this section.

[13] As discussed in Sect. 2.3, Assumption 1 is satisfied in this case.

[14] Following SSM (2017), we define status quo bias as the situation with a zero probability of changing initial actions and inertia in actions as the situation where actions satisfy $\Pr\left(a_t = 1|a_{t-1} = 1\right) + \Pr\left(a_t = 2|a_{t-1} = 2\right) > 1$ for all t sufficiently large. We can similarly define inertia in states.

With information cost $\lambda > 0$, we first consider the interior Markovian solution in 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 Eqs. (4) and (5), we have

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

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

$$q(a_t = 1|a_{t-1} = 1) = q(a_t = 2|a_{t-1} = 2) = 1 - \gamma, \tag{30}$$

as long as $\mu(2|1) \neq \mu(1|1)$. By prior symmetry the initial default rule satisfies $q_1(a_1 = 1) = q_1(a_1 = 2) = 1/2$.

Using the forward-backward algorithm, we numerically solve for the whole transition path. We can verify the above interior solution and find that there is no transition in this example. In particular, 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 in the Shannon entropy case. 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 $1 - \gamma$.

It is more interesting to analyze the finite-horizon case that admits corner solutions. SSM (2017) study the two-period case with the Shannon entropy and their Proposition 4 shows that when γ is sufficiently small, $q_2(a_2 = 1|a_1 = 1) = q_2(a_2 = 2|a_1 = 2) = 1$ 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. By contrast, this result does not hold in the infinite-horizon case. In particular, we always have the interior solution described in (30) for any $\gamma \in (0, 1)$ given $\mu_1(x_1 = 1) = 0.5$.

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 2 illustrates the analysis above for different values of the curvature parameter ρ for the Shorrocks entropy. We set $T = 6$, $\mu_1(x_1 = 1) = 0.5$, $\lambda = 1$, $\gamma = 0.03$, $U = 0$, and $\beta = 0.8$. For the baseline Shannon entropy case with $\rho = 1$, we find that $q_1(a_1 = 1) = 1/2$ by symmetry and

$$q_t(a_t = 2|a_t = 2) = q_t(a_t = 1|a_{t-1} = 1) = \begin{cases} 0.97 & \text{for } t = 2, \\ 0.9728 & \text{for } t = 3, \\ 1 & \text{for } t = 4, 5, 6. \end{cases}$$

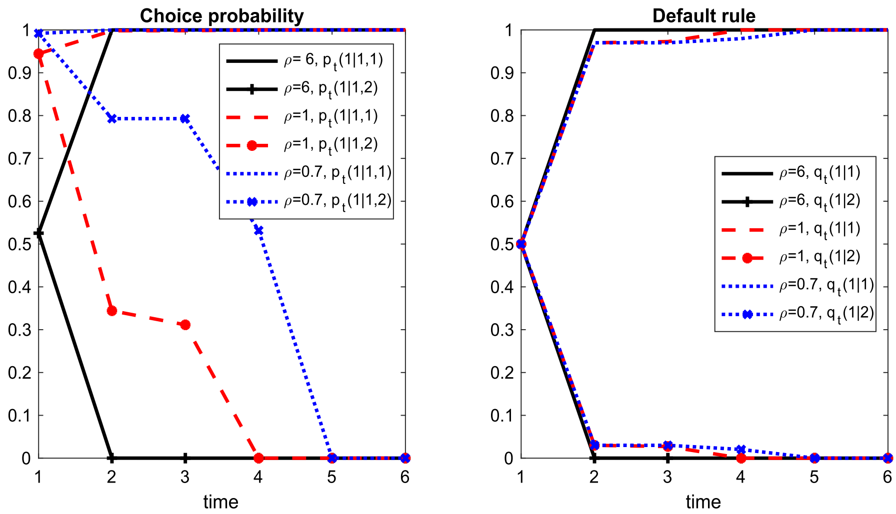

Fig. 2 Choice probabilities and default rules for different values of ρ. Parameter values are $T = 6$, $\mu_1(0) = 0.5$, $\pi(x_{t+1}|x_t, a_t) = \gamma = 0.03$ if $x_{t+1} \neq x_t$, $\beta = 0.8$, and $\lambda = 1$

Thus the status quo bias behavior occurs starting from period 4 on. The left panel of Fig. 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 $t = 1$, they are the same because $a_0 = \varnothing$. 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). Proposition 4 ensures that our computed Markovian solution is optimal. We also find that, when $T \to \infty$, there is no terminal time and $q_t(a_t = 1|a_{t-1} = 1) = 0.96 = 1 - \gamma$ for all $t \geq 2$.

Next we consider solutions when $\rho \neq 1$. We find that the status quo bias behavior occurs earlier when ρ is larger. Moreover, the state dependent choice probabilities $p_t(a_t = x_t|x_t, a_{t-1})$ decrease as ρ increases. The intuition is that the marginal cost of information is larger for a larger ρ so that the DM has less incentive to acquire new information. Thus the DM is more likely to make mistakes and stick to the old information.

Our analysis indicates that it is rational inattention combined with the short horizon that generates the status quo bias. This bias does not exist under infinite horizon. But the inertia behavior exists in both finite- and infinite-horizon settings. Moreover, the timing of the status quo bias depends on the marginal cost of information, which depends on the specification of the information cost function.

5.3 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 ($\lambda > 0$).

Proposition 6 *Consider the two-period RI model with Shannon entropy. Let* $U = 0$, $\beta = 1$, *and* $\pi(x_{t+1}|x_t, a_t) = \alpha \in [0, 1]$ *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 for reasons other than those discussed in the previous subsection. 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 $t > 1$.

We are unable to derive an analytical result similar to Proposition 6 for general UPS cost functions. We thus solve numerical examples using the Shorrocks entropy. Figure 3illustrates the transition dynamics for different values of ρ. We set the parameter values $T = 6$, $\lambda = 1$, $\beta = 0.8$, $U = 0$ and $\alpha = 0.7$. We find that there is no transition and the solution becomes stationary from the second period on. For this value of α, we have $\Pr(a_2 = a_1) = 1$ for $\rho = 1.8$**.** For $\rho = 1$ and 0.7, the solutions are interior, but $\rho = 1.8$ generates a corner solution. The difference in ρ is also reflected in the initial choice probabilities $p_1(a_1 = 1|x_1 = 1)$. Figure 3 shows that $p_1(a_1 = 1|x_1 = 1)$ declines as ρ increases. For a larger value of ρ, the marginal cost of information is larger and the DM has less incentive to acquire new information. Thus the DM is more likely to make mistakes.

Imagine that there is a unit mass of ex ante identical individuals in the case of $\rho = 1.8$. These individuals face the same decision problem with the same prior beliefs and parameter values. Our numerical results above show that half of the individuals choose action 1 in the first period and then will stick to this action forever with probability 1 from period 2 on. By contrast, the other half of the individuals choose action 2 in the first period and then will stick to this action forever with probability 1 from period 2 on.

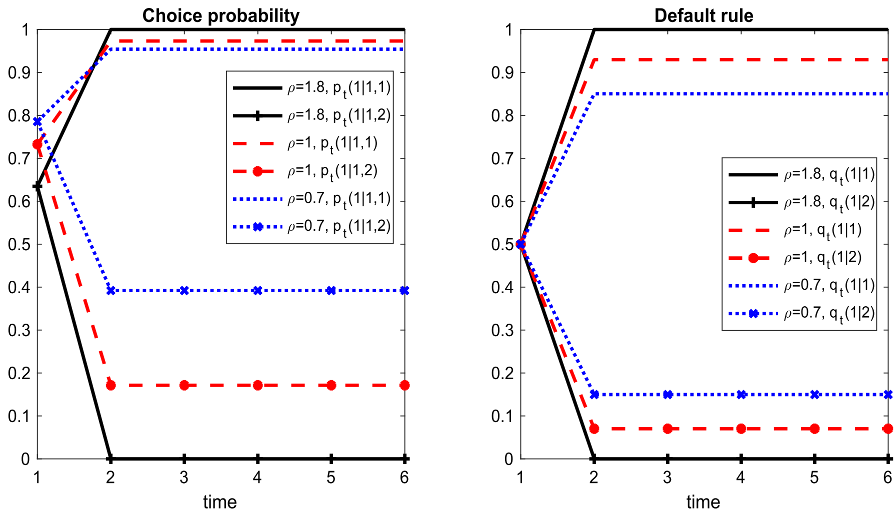

Fig. 3 Choice probabilities and default rules for different values of ρ. Parameter values are $T = 6$, $\mu_1(0) = 0.5$, $\pi(x_{t+1}|x_t, a_t) = \alpha = 0.7$ if $x_{t+1} = a_t$, $\beta = 0.8$, and $\lambda = 1$

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 ($\rho = 1.8$) 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 has implications for confirmation bias and 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 can occur. Suppose that all individuals with the same prior about the states have the same beliefs about state transition probabilities. In the case of $\rho = 1.8$, if they all believe the future state is more likely to be consistent with their current actions, then after the same states are realized over time, half of the individuals will choose only one action forever with probability 1, but the other half will choose the other action forever with probability 1. In the case of $\rho = 1$ and 0.7, they do acquire information beyond the first period. Because $p_t(1|1, 1) > p_t(2|1, 1)$ and $p_t(2|1, 2) > p_t(1|1, 2)$ for any $t > 0$, an individual acquires information that is more consistent with his previous action and his beliefs.

Our interpretation of the confirmation bias and belief polarization is similar to that of Nimark and Sundaresan (2019) (also see Catonini and Mayskaya (2023) for a different model). In their model, the state does not change over time. Agents update their prior beliefs about the state by acquiring endogenous signals. They show that the beliefs of ex ante identical agents over time can cluster in two distinct groups at opposite ends of the belief space. Unlike their model, we assume that states follow a Markov chain.

6 Conclusion

We adopt the posterior-based approach to study dynamic RI problems and provide necessary and sufficient conditions for optimal solutions for general UPS cost functions. We propose an efficient algorithm to solve these conditions and apply our model to explain some behavioral biases. Because the class of cost functions considered in our paper can help explain some behavior that violates the predictions of the RI models with the Shannon entropy cost, our approach will find wide applications in dynamic settings.

Appendix

A Preliminaries

In this appendix we first present the solution in the static case related to CD (2013), MM (2015), and CDL (2019). We then show that the choice-based approach of MM (2015) and SSM (2017) does not work for the general UPS cost functions. Next we study the two-period case and illustrate the difficulty of the dynamic model and our solution approach. Finally, we verify that our Assumption 1 is satisfied by the Shannon entropy so that the dynamic logit solution of SSM (2017) can be derived by our approach.

A.1 Static case

Notice that $N_H^a(\mu(\cdot|a))$ is concave in $\mu(\cdot|a)$ by the concavity of H, but the problem in (10) 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) and CDL (2019) instead propose a geometric approach from convex analysis and derive necessary and sufficient conditions for optimality.

Before stating some properties of the solution, let us treat the posterior probability of state M as the residual and define the set

$$Y = \left\{ (\mu_1, ..., \mu_{M-1}) \in \mathbb{R}_+^{M-1} : \sum_{m=1}^{M-1} \mu_m \leq 1 \right\}$$

as the domain of functions of probability distributions over X and $\widehat{V}(x)$ as the height of the supporting hyperplane of the net utilities in $Y \times \mathbb{R}$ at the point with $\mu(x) = 1$ and $\mu(x') = 0$ for all $x' \neq x$.[15]

[15] Without risk of confusion, we use the same notation $\widehat{V}(x)$ as in Definition 1 because the two functions are identical as shown in the proof of Proposition 1.

Proposition 7 *Consider Problem 3. (i) The optimal posteriors $\mu(\cdot|a)$ for all chosen actions a with $q(a) \in (0,1)$ 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*

$$V(\mu) = \overline{V}(\mu) - \lambda H(\mu) = \sum_x \mu(x) \widehat{V}(x) - \lambda H(\mu), \tag{A.1}$$

where $\widehat{V}(x)$ 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 with $q(a) \in (0,1)$.[16] *(iii) $\overline{V}(\mu)$ is concave in μ and for $x = 1, \ldots, M-1$ and $\mu(x) \in (0,1)$,*

$$\frac{\partial \overline{V}(\mu)}{\partial \mu(x)} = \widehat{V}(x) - \widehat{V}(M). \tag{A.2}$$

For the Shannon entropy case, $V(\mu)$ is convex in μ.[17]

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

(ii) By definition, $\overline{V}(\mu) = \sum_x \mu(x) \widehat{V}(x)$. By part (i), $\widehat{V}(x)$ is independent of the prior μ in that convex hull spanned by the optimal posteriors $\mu(\cdot|a)$ for all chosen actions a with $q(a) \in (0,1)$. We obtain the desired result.

(iii) Let q_i^* and $\mu_i^*(\cdot|\cdot)$ be the optimal solution corresponds to any prior μ_i for $i = 1, 2$. Let $q(a) = \theta q_1^*(a) + (1-\theta) q_2^*(a)$ for $\theta \in (0,1)$ for any chosen action a. Then we can derive

$$\begin{aligned}
&\theta \overline{V}(\mu_1) + (1-\theta) \overline{V}(\mu_2) \\
&\quad = \theta \sum_a q_1^*(a) N_H^a\left(\mu_1^*(\cdot|a)\right) + (1-\theta) \sum_a q_2^*(a) N_H^a\left(\mu_1^*(\cdot|a)\right) \\
&\quad = \sum_a q(a) \left[\frac{\theta q_1^*(a)}{q(a)} N_H^a\left(\mu_1^*(\cdot|a)\right) + \frac{(1-\theta) q_2^*(a)}{q(a)} N_H^a\left(\mu_2^*(\cdot|a)\right)\right] \\
&\quad \leq \sum_a q(a) N_H^a\left(\frac{\theta q_1^*(a)}{q(a)} \mu_1^*(\cdot|a) + \frac{(1-\theta) q_2^*(a)}{q(a)} \mu_2^*(\cdot|a)\right) \\
&\quad \leq \overline{V}(\theta \mu_1 + (1-\theta) \mu_2),
\end{aligned}$$

where the first inequality follows from the concavity of the net utility N_H^a, and the second inequality from the following:

[16] 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.

[17] Propositions 4 and 5 of Denti et al. (2022) show that $V(\mu)$ is convex for any general information cost function as long as it is experimental.

$$\sum_a q(a)\left(\frac{\theta q_1^*(a)}{q(a)}\mu_1^*(\cdot|a)+\frac{(1-\theta)q_2^*(a)}{q(a)}\mu_2^*(\cdot|a)\right)$$
$$=\theta\sum_a q_1^*(a)\mu_1^*(\cdot|a)+(1-\theta)q_2^*(a)\mu_2^*(\cdot|a)$$
$$=\theta\mu_1(\cdot)+(1-\theta)\mu_2(\cdot).$$

Thus $\overline{V}(\mu)$ is concave in μ.

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 then obtain (A.2). CDL (2019) 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 (A.2) applies to each set of priors and crosses boundaries of neighboring sets continuously.

Finally consider the Shannon entropy case. Then Problem 3 can be reformulated as in (A.3) and (A.4) presented in Appendix A.2 (see MM (2015)). 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 Shannon entropy is a concave function, we deduce that

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

Thus $V(\mu)$ is convex. □

Part (i) is the locally invariant posteriors (LIP) property discovered by CD (2013). Part (ii) can be best understood using the geometric approach. Specifically, the optimal posterior $\mu(\cdot|a)$ is the tangent point of the net utility associated with the chosen action a and $\widehat{V}(x)$ satisfies

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

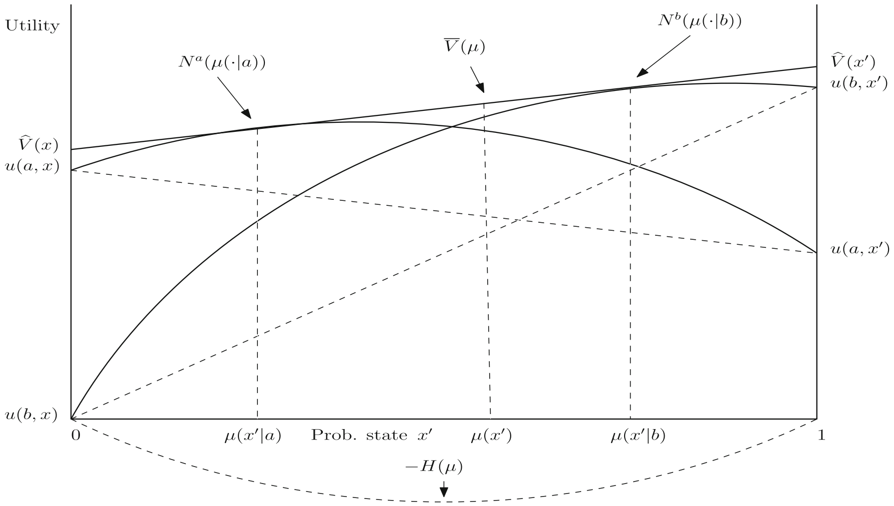

Fig. 4 The net utility function and concavification

at the optimum. The value $\overline{V}(\mu)$ is the height above $\mu(x)$ of the convex hull connecting $N_H^a(\mu(\cdot|a))$ for all chosen actions a. The optimal posterior $\mu(\cdot|a)$ is the tangent point of $\overline{V}(\mu)$ and $N_H^a(\mu(\cdot|a))$ for each a with $q(a) \in (0, 1)$. The value $\widehat{V}(x)$ is the height of the supporting hyperplane at the point with $\mu(x) = 1$ and $\mu\left(x'\right) = 0$ for all $x' \neq x$. This value is independent of the prior μ in the convex hull of the optimal posteriors. This result does not appear in the literature and is critical for the analysis of the dynamic model. In particular, this result establishes a useful property of the value function in the dynamic model, which implies that Eq. (A.2) also holds for the value function. This equation will be repeatedly applied when we derive first-order conditions for the dynamic model.

Figure 4 is similar to Figure 5 of CDL (2019) in the case with two states $\left\{x, x'\right\}$ 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 $\widehat{V}\left(x'\right)$ are invariant to changes of $\mu\left(x'\right)$ within the interval $\left(\mu\left(x'|a\right), \mu\left(x'|b\right)\right)$. If $\mu\left(x'\right) \in [0, \mu\left(x'|a\right)]$, then $q(a) = 1$ and $\mu\left(x'|a\right) = \mu\left(x'\right)$. If $\mu\left(x'\right) \in [\mu\left(x'|b\right), 1]$, then $q(b) = 1$ and $\mu\left(x'|b\right) = \mu\left(x'\right)$.

Part (iii) of Proposition 7 shows that $\overline{V}(\mu)$ is a concave function because it is the concave envelope of net utilities. It is also differentiable and satisfies an envelope condition. It is unclear whether $V(\mu)$ is concave as it is equal to the difference of two concave functions by (A.1). Part (iii) shows that $V(\mu)$ is convex if H is the Shannon entropy function. This issue poses a difficulty when solving the dynamic RI problem.

A.2 Failure of the choice-based approach

MM (2015) and SSM (2017) solve RI problems with the Shannon entropy cost using the choice-based approach. To understand this approach, we notice that Problem 3 for the Shannon entropy case can be rewritten as

$$V(\mu) = \max_{q\in\Delta(A),\, p\in\Delta(X|A)} \sum_{x,a} p(a|x)\,\mu(x)\left[u(x,a) - \lambda \ln \frac{p(a|x)}{q(a)}\right], \tag{A.3}$$

subject to

$$q(a) = \sum_{x} p(a|x)\,\mu(x),\ \ a \in A. \tag{A.4}$$

Let $F(p,q)$ denote the objective function in (A.3). We can verify that $F(p,q)$ is jointly concave in (p,q). Blahut (1972, Theorem 4) establishes the following result:

Lemma 2 *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 (A.3) is equivalent to the following unconstrained optimization problem:

$$\max_{p\in\Delta(A|X),\ q\in\Delta(A)} F(p,q). \tag{A.5}$$

Taking first-order conditions with respect to p and q yields the choice-based characterization as in MM (2015) and CDL (2019). CDL (2019) also provide sufficient conditions for optimality.

To illustrate why the choice-based approach may not work for general UPS cost functions, we let H be the weighted entropy. Then the cost function becomes

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

for some weight function $w : X \to [0,1]$. Following the choice-based approach described above, 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 2 does not hold in general so that the static RI problem is not equivalent to the unconstrained problem in (A.5) for general UPS cost functions. Similarly, Lemma 2 in SSM (2017) also fails for general UPS cost functions in dynamic RI models. Thus the unconstrained coordinate-wise first-order conditions for p and q cannot be used to characterize the solutions to dynamic RI problems.

A.3 Two-period case

We study a two-period problem with $T = 2$ and $U = 0$ by dynamic programming. First, consider the problem in period 2 conditional on the history of a chosen action a_1 :

$$V_2\left(\mu_2\left(\cdot|a_1\right)\right) = \max_{q_2(\cdot|a_1),\mu_2(\cdot|\cdot,a_1)} \sum_{a_2} q_2\left(a_2|a_1\right) N_H^{a_2}\left(\mu_2\left(\cdot|a^2\right)\right) - \lambda H\left(\mu_2\left(\cdot|a_1\right)\right) \tag{A.6}$$

subject to

$$\mu_2\left(\cdot|a_1\right) = \sum_{a_2} \mu_2\left(\cdot|a^2\right) q_2\left(a_2|a_1\right), \tag{A.7}$$

where the net utility $N_H^{a_2}$ is given by

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

The prior distribution $\mu_2\left(\cdot|a_1\right)$ also satisfies

$$\mu_2\left(x_2|a_1\right) = \sum_{x_1} \pi\left(x_2|x_1, a_1\right) \mu_1\left(x_1|a_1\right), \quad a_1 \in A, x_2 \in X. \tag{A.8}$$

It follows from Proposition 1 that $q_2(\cdot|a_1)$ and $\mu_2(\cdot|\cdot, a_1)$ are optimal if and only if: (i) Equation (A.7) holds. (ii) For any $x_2 \in X$, and chosen actions a_1 and a_2 with $q_1\left(a_1\right) > 0$ and $q_2\left(a_2|a_1\right) > 0$, we have

$$\widehat{V}_2(x_2|a_1) = u(x_2, a_2) + \lambda H_{x_2}(\mu_2(\cdot|a^2)) + \lambda f_2(\mu_2\left(\cdot|a^2\right)), \tag{A.9}$$

where

$$f_2(\nu) \equiv H(\nu) - \sum_{x} \nu(x) H_x(\nu), \quad \nu(x) \in \Delta\left(X\right). \tag{A.10}$$

(iii) For any unchosen action a_2 and $\mu_2^{a_2} \in \Delta\left(X\right)$ such that

$$u\left(x_2, a_2\right) + \lambda H_{x_2}\left(\mu_2^{a_2}\right) - \left[u\left(M, a_2\right) + \lambda H_M\left(\mu_2^{a_2}\right)\right] = \widehat{V}_2(x_2|a_1) - \widehat{V}_2(M|a_1)$$

for $x_2 = 1, 2, ..., M - 1$, we have

$$\sum_{x_2} I_{x_2}\left(\widehat{V}_2(x_2|a_1)/\lambda - u(x_2, a_2)/\lambda - f_2(\mu_2^{a_2}); \mu_2^{a_2}\right) \leq 1.$$

By Proposition 7, the value function in period 2 satisfies

$$
\begin{aligned}
V_2\left(\mu_2(\cdot|a_1)\right) &= \overline{V}_2(\mu_2(\cdot|a_1)) - \lambda H(\mu_2(\cdot|a_1)) \\
&= \sum_{x_2} \mu_2\left(x_2|a_1\right) \widehat{V}_2\left(x_2|a_1\right) - \lambda H\left(\mu_2\left(\cdot|a_1\right)\right).
\end{aligned} \tag{A.11}
$$

Next consider the problem in period 1. By dynamic programming, we use (A.11) to derive

$$
V_1\left(\mu_1\right) = \max_{q_1, \mu_1(\cdot|\cdot)} \sum_{a_1} q_1\left(a_1\right) N_G^{a_1}\left(\mu_1\left(\cdot|a_1\right)\right) - \lambda H\left(\mu_1\right), \tag{A.12}
$$

where

$$
\mu_1(x_1) = \sum_{a_1} q_1(a_1)\mu_1(x_1|a_1), \quad x_1 \in X. \tag{A.13}
$$

The net utility $N_G^{a_1}$ associated with action a_1 is

$$
N_G^{a_1}(\mu_1(\cdot|a_1)) = \sum_{x1} \mu_1(x_1|a_1)u(x_1, a_1) + \beta\overline{V}_2(\mu_2(\cdot|a_1)) + \lambda G^{a_1}(\mu_1(\cdot|a_1)), \tag{A.14}
$$

where $\overline{V}_2$ is given in (A.11), G^{a_1} is defined in (8), and $\mu_2\left(\cdot|a_1\right)$ satisfies (A.7).

It follows from Proposition 7 that $\overline{V}_2$ is concave in $\mu_2(\cdot|a_1)$, and hence concave in $\mu_1(\cdot|a_1)$ by (A.7). Moreover, Assumption 1 ensures the concavity of G^{a_1} in $\mu_1(\cdot|a_1)$. Therefore the net utility $N_G^{a_1}$ is concave in $\mu_1(\cdot|a_1)$. We view the problem in period 1 as a static RI problem with the prior belief μ_1. Applying our Proposition 1, we derive the following result.

Lemma 3 *Let Assumption 1 hold. Then the pair $\mu_1(\cdot|\cdot)$ and q_1 is optimal for problem (A.12) if any only if it satisfies the following conditions: (i) Equation (A.13) is satisfied. (ii) There exists a function $\widehat{V}_1\left(x_1\right)$ such that for any chosen action a_1 with $q_1(a_1) > 0$ and for any $x_1 \in X$,*

$$
\widehat{V}_1\left(x_1\right) = v_1(x_1, a_1) + \lambda H_{x_1}(\mu_1(\cdot|a_1)) + \lambda f_1(\mu_1(\cdot|a_1), a_1), \tag{A.15}
$$

where f_1 and v_1 satisfy (20) and (21) with $t = 1$. (iii) For any unchosen action a_1 and $\mu_1^{a_1} \in \Delta\left(X\right)$ such that

$$
\left[v_1(x_1, a_1) + \lambda H_{x_1}(\mu_1^{a_1})\right] - \left[v_1(M, a_1) + \lambda H_M(\mu_1^{a_1})\right] = \widehat{V}_1\left(x\right) - \widehat{V}_1\left(M\right),
$$

for $x_1 \in X$, we have

$$
\sum_{x_1} I_{x_1}\left(\widehat{V}_1(x_1)/\lambda - v_1(x_1, a_1)/\lambda - f_1(\mu_1^{a_1}, a_1); \mu_1^{a_1}\right) \leq 1. \tag{A.16}
$$

Moreover, the optimal value for the two-period RI problem satisfies

$$
V_1(\mu_1) = \sum_{x_1} \mu(x_1)\widehat{V}_1(x_1) - \lambda H(\mu_1).
$$

Proof First, we rewrite (A.8) as

$$\mu_2(x_2|a_1) = \sum_{x_1=1}^{M-1} \pi(x_2|x_1, a_1)\mu_1(x_1|a_1) + \pi(x_2|M, a_1)\left(1 - \sum_{x_1=1}^{M-1} \mu_1(x_1|a_1)\right), \tag{A.17}$$

and (8) as

$$G^{a_1}(\mu_1(\cdot|a_1)) = H(\mu_1(\cdot|a_1)) - \beta H(\mu_2(\cdot|a_1)) \tag{A.18}$$

Using (A.17), (A.18), and the chain rule, we calculate

$$\begin{aligned}\frac{\partial G^{a_1}(\mu_1(\cdot|a_1))}{\partial \mu_1(x_1|a_1)} =& H_{x_1}(\mu_1(\cdot|a_1)) - H_M(\mu_2(\cdot|a_1)) \\ & - \beta \sum_{x_2=1}^{M} H_{x_2}(\mu_2(\cdot|a_1))\left[\pi(x_2|x_1, a_1) - \pi(x_2|M, a_1)\right],\end{aligned} \tag{A.19}$$

for $x_1 = 1, \ldots, M-1$.

Second, using Proposition 7 (iii), Eq. (A.17), and the chain rule, we can derive for $x_1 = 1, \ldots, M-1$,

$$\begin{aligned}\frac{\partial \overline{V}_2(\mu_2(\cdot|a_1))}{\partial \mu_1(x_1|a_1)} &= \sum_{x_2=1}^{M-1} \frac{\partial \overline{V}_2(\mu_2(\cdot|a_1))}{\partial \mu_2(x_2|a_1)} \frac{\partial \mu_2(x_2|a_1)}{\partial \mu_1(x_1|a_1)} \\ &= \sum_{x_2=1}^{M-1} \left[\widehat{V}_2(x_2|a_1) - \widehat{V}_2(M|a_1)\right]\left[\pi(x_2|x_1, a_1) - \pi(x_2|M, a_1)\right] \\ &= \sum_{x_2=1}^{M-1} \widehat{V}_2(x_2|a_1)\left[\pi(x_2|x_1, a_1) - \pi(x_2|M, a_1)\right] \\ &\quad - \sum_{x_2=1}^{M-1} \widehat{V}_2(M|a_1)\left[\pi(x_2|x_1, a_1) - \pi(x_2|M, a_1)\right] \\ &= \sum_{x_2=1}^{M} \widehat{V}_2(x_2|a_1)[\pi(x_2|x_1, a_1) - \pi(x_2|M, a_1)],\end{aligned} \tag{A.20}$$

where the last equality uses the equations

$$\sum_{x_2=1}^{M-1} \pi(x_2|x_1, a_1) = 1 - \pi(M|x_1, a_1), \quad \sum_{x_2=1}^{M-1} \pi(x_2|M, a_1) = 1 - \pi(M|M, a_1).$$

Therefore, combining (A.19) and (A.20), we have

$$\begin{aligned}\frac{\partial N_G^{a_1}(\mu_1(\cdot|a_1))}{\partial\mu_1(x_1|a_1)} &= u(x_1,a_1)-u(M,a_1)+\beta\frac{\partial\overline{V}_2(\mu_2(\cdot|a_1))}{\partial\mu_1(x_1|a_1)}+\lambda\frac{\partial G^{a_1}(\mu_1(\cdot|a_1))}{\partial\mu_1(x_1|a_1)}\\ &= [u(x_1,a_1)+\lambda H_{x_1}(\mu_1(\cdot|a_1))]-[u(M,a_1)+\lambda H_M(\mu_1(\cdot|a_1))]\\ &\quad+\beta\sum_{x_2=1}^{M}\left[\widehat{V}_2(x_2|a_1)-\lambda H_{x_2}(\mu_2(\cdot|a_1))\right][\pi(x_2|x_1,a_1)-\pi(x_2|M,a_1)]\\ &= \left[v_1(x_1,a_1)+\lambda H_{x_1}(\mu_1(\cdot|a_1))\right]-[v_1(M,a_1)+\lambda H_M(\mu_1(\cdot|a_1))].\end{aligned} \tag{A.21}$$

We now use Lemma 3 of CD (2013), it is also equivalent to the Lagrangian lemma in CDL (2022) Lemma 1. Condition (ED) in Lemma 3 of CD (2013) is equivalent to

$$\begin{aligned}&\left[v_1(x_1,a_1)+\lambda H_{x_1}(\mu_1(\cdot|a_1))\right]-[v_1(M,a_1)+\lambda H_M(\mu_1(\cdot|a_1))]\\ &\quad=\left[v_1(x_1,b_1)+\lambda H_{x_1}(\mu_1(\cdot|b_1))\right]-[v_1(M,b_1)+\lambda H_M(\mu_1(\cdot|b_1))],\end{aligned} \tag{A.22}$$

for any chosen actions a_1, b_1 and $x_1 = 1,\ldots,M-1$. Using (A.14) and (A.21), we calculate

$$\begin{aligned}&N_G^{a_1}(\mu_1(\cdot|a_1))-\sum_{x_1=1}^{M-1}\frac{\partial N_G^{a_1}(\mu_1(\cdot|a_1))}{\partial\mu_1(x_1|a_1)}\mu_1(x_1|a_1)\\ &\quad=v_1(M,a_1)+\lambda H_M(\mu_1(\cdot|a_1))+\lambda f_1(\mu_1(\cdot|a_1),a_1).\end{aligned} \tag{A.23}$$

Then condition (CT) in CD (2013) is equivalent to

$$\begin{aligned}&v_1(M,a_1)+\lambda H_M(\mu_1(\cdot|a_1))+\lambda f_1(\mu_1(\cdot|a_1),a_1)\\ &\quad=v_1(M,b_1)+\lambda H_M(\mu_1(\cdot|b_1))+\lambda f_1(\mu_1(\cdot|b_1),b_1),\end{aligned} \tag{A.24}$$

for any chosen actions a_1 and b_1.

The rest of the proof is the same as the proof of Proposition 1 with u, $\mu(\cdot|\cdot)$, and f replaced by v_1, $\mu_1(\cdot|\cdot)$, and f_1, respectively. After $\widehat{V}_1$ is identified, the expression for the optimal value follows from Proposition 7 directly. ◻

A.4 Sufficient conditions for Assumption 1

A.4.1 Shannon entropy case

We first verify Assumption 1 for the Shannon entropy case, $H(\mu(x)) = -\sum_x \mu(x)\ln\mu(x)$ and provide a dynamic logit characterization.

For any $\mu,\tilde{\mu}\in\Delta(X)$, define $\tilde{G}^a(\mu,\tilde{\mu})$ as

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

Therefore,

$$G^a(\nu) = \tilde{G}^a\left(\nu, \sum_x \pi(\cdot|x,a)\nu(x)\right).$$

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

For any $\theta \in [0, 1]$ and $\nu, \nu' \in \Delta(X)$,

$$\begin{aligned}
&G^a\left(\theta\nu + (1-\theta)\nu'\right) \\
&\quad= \tilde{G}^a\left(\theta\nu + (1-\theta)\nu', \theta\sum_x \pi(\cdot|x,a)\nu(x) + (1-\theta)\sum_x \pi(\cdot|x,a)\nu'(x)\right) \\
&\quad\geq \theta\tilde{G}^a\left(\nu, \sum_x \pi(\cdot|x,a)\nu(x)\right) + (1-\theta)\tilde{G}^a\left(\nu, \sum_x \pi(\cdot|x,a)\nu'(x)\right) \\
&\quad= \theta G^a(\nu) + (1-\theta)G^a(\nu'),
\end{aligned}$$

where the inequality follows from the definition of a jointly concave function. Thus Assumption 1 is satisfied for Shannon entropy.

We can then apply Theorem 1 to derive a dynamic logit solution.

Corollary 1 *The solution to dynamic RI Problem 2 with the Shannon entropy cost satisfies the following necessary and sufficient conditions: (i) For $t = 1, \ldots, T$,*

$$\mu_t(x_t|a^t) = \frac{\mu_t(x_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, a^t)/\lambda\right)}{\sum_{b_t} q_t(b_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, b_t, a^{t-1})/\lambda\right)}, \tag{A.25}$$

$$p_t\left(a_t|x_t, a^{t-1}\right) = \frac{q_t(a_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, a^t)/\lambda\right)}{\sum_{b_t} q_t(b_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, b_t, a^{t-1})/\lambda\right)}, \tag{A.26}$$

$$q_t\left(a_t|a^{t-1}\right) = \sum_{x_t} p_t\left(a_t|x_t, a^{t-1}\right)\mu_t\left(x_t|a^{t-1}\right), \tag{A.27}$$

$$\tilde{v}_t(x_t, a^t) = u(x_t, a_t) + \beta\sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t)\widetilde{V}_{t+1}\left(x_{t+1}, a^t\right), \tag{A.28}$$

$$\widetilde{V}_t\left(x_t, a^{t-1}\right) = \lambda\ln\left(\sum_{a_t} q_t(a_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, a^t)/\lambda\right)\right), \tag{A.29}$$

with the terminal condition $\widetilde{V}_{T+1}\left(x_{T+1}, a^T\right) = U(x_{T+1})$. (ii) For any $a_t \in A$,

$$\sum_{x_t} \frac{\mu_t(x_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, a^t)/\lambda\right)}{\sum_{b_t} q_t(b_t|a^{t-1})\exp\left(\tilde{v}_t(x_t, b_t, a^{t-1})/\lambda\right)} \leq 1, \tag{A.30}$$

with equality if $q_t(a_t|a^{t-1}) > 0$*. Moreover, the value function satisfies*

$$V^{T-t+1}\left(\mu_t\left(\cdot|a^{t-1}\right)\right) = \sum_{x_t}\mu_t\left(x_t|a^{t-1}\right)\widetilde{V}_t\left(x_t, a^{t-1}\right),\ t = 1, 2, ..., T. \quad \text{(A.31)}$$

Proof We provide a sketch of the proof here. In the Shannon entropy case, we have $f_t(\mu_t\left(\cdot|a^t\right), a_t) = 1 - \beta$ for $t = 1, \ldots, T-1$, and $f_T(\mu_T\left(\cdot|a^T\right), a_T) = 1$. We obtain from (22) that

$$\mu_t(x_t|a^t) = \exp\left(-\frac{\widehat{V}_t(x_t|a^{t-1})}{\lambda} + \frac{v_t(x_t, a^t)}{\lambda} - \beta\mathbf{1}_{\{t<T\}}\right),$$

where $\mathbf{1}$ is an indicator function. Using this equation and (4), we can solve for $\widehat{V}_t(x_t|a^{t-1})$:

$$\widehat{V}_t(x_t|a^{t-1}) = -\lambda\ln\left(\frac{\mu_t(x_t|a^{t-1})}{\sum_{b_t} q_t(b_t|a^{t-1})\exp\left(\frac{v_t(x_t,b_t,a^{t-1})}{\lambda} - \beta\mathbf{1}_{\{t<T\}}\right)}\right).$$

Define $\tilde{v}_t(x_t, a^t) = v_t(x_t, a^t) - \lambda\beta\mathbf{1}_{\{t<T\}}$ and define $\widetilde{V}_t\left(x_t, a^{t-1}\right)$ as in (A.29). Combining the previous two equations, we confirm (A.25). Plugging the previous expression of $\widehat{V}_t(x_t|a^{t-1})$ into (21), we confirm (A.28) for $\tilde{v}_t$. Equations (A.26) and (A.27) follow from the usual probability rules. Inequality (A.30) follows from (24) and (A.31) follows from (25). □

By this corollary and Theorem 2, we can provide a first-order characterization of a Markovian solution for the Shannon entropy cost. For space limitations, we will not state this result.

A.4.2 Two-state case

Consider the two-state case, i.e., $M = 2$. For any distribution $\nu = (\nu(1), \nu(2))$, define

$$h(\nu(1)) \equiv H(\nu) \quad \text{and}$$
$$g^a(\nu(1)) \equiv G^a(\nu) = H(\nu) - \beta H\left(\sum_{x=1,2}\pi(\cdot|x, a)\nu(x)\right), \quad a \in A.$$

Both h and g^a are univariate functions on [0, 1]. It follows from the concavity of H that h is concave. We also denote

$$\widetilde{\nu}^a(x) \equiv \sum_{x'}\pi(x|x', a)\nu(x'), \quad x = 1, 2.$$

For notation simplicity, we remove the superscript a.

Assume that the transition matrix is *symmetric* for any $a \in A$ in the sense that

$$\pi(1|2,a) = \pi(2|1,a), \quad \text{for any } a \in A. \tag{A.32}$$

The symmetric property of the transition matrix implies that

$$\min\{\nu(1), \nu(2)\} \le \widetilde{\nu}(i) \le \max\{\widetilde{\nu}(1), \widetilde{\nu}(2)\}, \quad i = 1, 2. \tag{A.33}$$

This means that the distribution $\widetilde{\nu} = (\widetilde{\nu}(1), \widetilde{\nu}(2))$ is closer to the uniform distribution $(1/2, 1/2)$ than the distribution $\nu = (\nu(1), \nu(2))$.

To prove (A.33), consider $i = 1$ and derive

$$\begin{aligned}\widetilde{\nu}(1) &= \pi(1|1,a)\nu(1) + \pi(1|2,a)\nu(2) = \pi(1|1,a)\nu(1) + \pi(2|1,a)\nu(2)\\ &\ge \pi(1|1,a)\min\{\nu(1),\nu(2)\} + \pi(2|1,a)\min\{\nu(1),\nu(2)\} = \min\{\nu(1),\nu(2)\},\end{aligned}$$

where the second equality follows from (A.32). Similarly, we can prove $\widetilde{\nu}(1) \le \max\{\nu(1), \nu(2)\}$.

Lemma 4 *Suppose that* $\beta \in (0, 1)$, $h(\nu(1)) = h(1 - \nu(1))$ *for any* $\nu(1) \in [0, 1]$, *and* $h^{''}$ *increases on* $[0, 1/2]$. *(i) If* π *is symmetric, i.e.,* (A.32) *holds, then Assumption 1 holds. (ii) Suppose further that the transition kernel* π *satisfies* $\pi(x'|x, a) \ge \epsilon$ *for some* $\epsilon > 0$ *and any* $a \in A$, $x, x' \in X$, *and that* $h^{''}(\nu(1)) \le -\eta$ *for some* $\eta > 0$ *and any* $\nu(1) \in [0, 1]$. *If* π *is sufficiently close to be symmetric under some matrix norm, then Assumption 1 also holds.*

Proof (i) Because

$$g(\nu(1)) = h(\nu(1)) - \beta h(\widetilde{\nu}(1)),$$

we can compute

$$\begin{aligned}g^{''}(\nu(1)) &= h^{''}(\nu(1)) - \beta h^{''}(\widetilde{\nu}(1))\,[\pi(1|1,a) - \pi(1|2,a)]^2\\ &\le h^{''}(\nu(1)) - h^{''}(\widetilde{\nu}(1)),\end{aligned} \tag{A.34}$$

where the inequality follows from $h^{''}(\widetilde{\nu}(1)) \le 0$ and $\beta\,(\pi(1|1,a) - \pi(1|2,a))^2 < 1$. First consider the case of $\nu(1) \le 1/2$. If $\widetilde{\nu}(1) \le 1/2$, because $\widetilde{\nu}(1) \ge \nu(1)$ from (A.33) and $h^{''}$ increases on $[0, 1/2]$, we have $h^{''}(\nu(1)) - h^{''}(\widetilde{\nu}(1)) \le 0$. If $\widetilde{\nu}(1) > 1/2$, it follows from $h(\nu(1)) = h(1 - \nu(1))$ that $h^{''}(\widetilde{\nu}(1)) = h^{''}(1 - \widetilde{\nu}(1))$. Due to $1 - \widetilde{\nu}(1) \ge \nu(1)$ from (A.33) and the fact that $h^{''}$ increases on $[0, 1/2]$, we still have $h^{''}(\nu(1)) - h^{''}(\widetilde{\nu}(1)) = h^{''}(\nu(1)) - h^{''}(1 - \widetilde{\nu}(1)) \le 0$. Next for the case of $\nu(1) > 1/2$, we can work with the variable $\nu(2) = 1 - \nu(1)$ by the symmetry of h. The same argument follows.

(ii) Given the additional assumption, we have

$$\widetilde{\nu}^{\pi}(x) = \sum_{x'} \pi(x|x', a)\nu(x') \ge \epsilon \sum_{x'} \nu(x') = \epsilon, \quad \text{for } x \in X,\ a \in A,$$

where we introduce the superscript to specify the dependence of $\widetilde{\nu}(x)$ on π. This inequality implies that $\widetilde{\nu}^{\pi}(1) \in [\epsilon, 1-\epsilon]$. Because h'' is continuous in the bounded closed interval $[\epsilon, 1-\epsilon]$, it is also uniformly continuous in the same interval. It follows that

$$\left| h''(\widetilde{\nu}^{\pi}(1)) - h''(\widetilde{\nu}^{\pi_s}(1)) \right| \leq \delta \quad \text{for any } \delta > 0, \tag{A.35}$$

when π is sufficiently close to some symmetric transition kernel π_s. Meanwhile, for any $\delta < \frac{1-\beta}{\beta}\eta$, we can derive

$$\begin{aligned} g''(\nu(1)) &\leq h''(\nu(1)) - \beta h''(\widetilde{\nu}^{\pi}(1)) = (1-\beta)h''(\nu(1)) + \beta \left[h''(\nu(1)) - h''(\widetilde{\nu}^{\pi}(1)) \right] \\ &\leq (1-\beta)h''(\nu(1)) + \beta\delta + \beta \left[h''(\nu(1)) - h''(\widetilde{\nu}^{\pi_s}(1)) \right] \\ &\leq (1-\beta)h''(\nu(1)) + \beta\delta \\ &\leq -\eta(1-\beta) + \beta\delta < 0, \end{aligned}$$

where the first inequality follows from (A.34), the second inequality from (A.35), the third inequality from $h''(\nu(1)) - h''(\widetilde{\nu}^{\pi_s}(1)) \leq 0$ proved in part (i) of the lemma, the fourth inequality from $h''(\nu(1)) \leq -\eta$, and the final inequality from $\delta < \frac{1-\beta}{\beta}\eta$. Therefore Assumption 1 holds. □

To apply this lemma, we first consider the Shorrocks entropy

$$H(\nu) = \frac{1 - \nu(1)^{2-\rho} - \nu(2)^{2-\rho}}{(\rho-1)(\rho-2)}, \quad \rho \neq 1, 2.$$

Then

$$h(\nu(1)) = \frac{1 - \nu(1)^{2-\rho} - (1-\nu(1))^{2-\rho}}{(\rho-1)(\rho-2)}, \quad \rho \neq 1, 2.$$

The symmetry $h(\nu(1)) = h(1-\nu(1))$ clearly holds. Moreover,

$$h''(\nu(1)) = -\nu(1)^{-\rho} - (1-\nu(1))^{-\rho}.$$

Then $h''(\nu(1)) \leq -\eta$ holds for some $\eta > 0$ and any $\nu(1) \in [0, 1]$. Moreover h'' is increasing on $[0, 1/2]$ for $\rho \notin (-1, 0)$ and decreasing on $[0, 1/2]$ for $\rho \in (-1, 0)$. Thus Lemma 4 holds for the Shorrocks entropy with $\rho \notin (-1, 0)$.

For the total information cost case,

$$H(\nu) = -\omega\nu(1) \ln \frac{\nu(1)}{\nu(2)} - \omega\nu(2) \ln \frac{\nu(2)}{\nu(1)}, \ \omega > 0.$$

Then

$$h(\nu(1)) = -\omega\nu(1) \ln \frac{\nu(1)}{1-\nu(1)} - \omega(1-\nu(1)) \ln \frac{1-\nu(1)}{\nu(1)}.$$

The symmetry $h(\nu(1)) = h(1-\nu(1))$ clearly holds. Moreover,

$$h^{''}(\nu(1)) = -\frac{2\omega}{\nu(1)} - \frac{2\omega}{1-\nu(1)} - \frac{\omega}{\nu(1)^2} - \frac{\omega}{(1-\nu(1))^2}.$$

Then $h^{''}(\nu(1)) \le -\eta$ holds for some $\eta > 0$ and any $\nu(1) \in [0,1]$, and $h^{''}$ is increasing on $[0, 1/2]$. Thus Lemma 4 holds for the total information cost function.

A.4.3 Additional cases

Now we provide two other sufficient conditions for Assumption 1 with many states.

Lemma 5 *(i) Suppose there exist constants* $\overline{C} \ge \underline{C} > 0$ *such that*

$$-\overline{C}\|\nu-\mu\|^2 \le H(\nu)-H(\mu)-\big(\nabla H(\mu)\big)^\top(\nu-\mu) \le -\underline{C}\|\nu-\mu\|^2, \quad \text{for any } \nu,\mu \in \Delta(X), \tag{A.36}$$

where ∇H *is the gradient vector of* H *and* $\|\cdot\|$ *is the Euclidean norm in* $\mathbb{R}^{|X|}$. *Then Assumption 1 is satisfied for any* $\beta \in (0, \beta_0]$ *with some* $\beta_0 \le 1$ *and any transition kernel* π.

(ii) Suppose that only the second inequality in (A.36) holds. Then Assumption 1 is satisfied for any $\beta \in (0, \beta_0]$ *with with some* $\beta_0 \le 1$ *and any transition kernel* π *satisfying* $\pi(x'|x,a) \ge \epsilon$ *for some* $\epsilon > 0$ *and for any* $a \in A$, $x, x' \in X$.

Proof (i) To simplify notation, we omit the superscript a for the function G^a defined in Assumption 1. For any $\alpha \in (0,1)$, $\nu_1, \nu_2 \in \Delta(X)$,

$$\begin{aligned} &G(\alpha\nu_1 + (1-\alpha)\nu_2) - \alpha G(\nu_1) - (1-\alpha)G(\nu_2) \\ &\quad = \big[H(\alpha\nu_1 + (1-\alpha)\nu_2) - \alpha H(\nu_1) - (1-\alpha)H(\nu_2)\big] \\ &\quad\quad - \beta\big[H(\alpha\tilde{\nu}_1 + (1-\alpha)\tilde{\nu}_2) - \alpha H(\tilde{\nu}_1) - (1-\alpha)H(\tilde{\nu}_2)\big], \end{aligned} \tag{A.37}$$

where $\tilde{\nu}_i(x) \equiv \sum_{x'} \pi(x|x',a)\nu_i(x')$, $i = 1, 2$. It follows from the second inequality in (A.36) that

$$\begin{aligned} &H(\nu_1) - H(\alpha\nu_1 + (1-\alpha)\nu_2) - \big(\nabla H(\alpha\nu_1 + (1-\alpha)\nu_2)\big)^\top(1-\alpha)(\nu_1-\nu_2) \\ &\quad \le -\underline{C}(1-\alpha)^2\|\nu_1-\nu_2\|^2, \\ &H(\nu_2) - H(\alpha\nu_1 + (1-\alpha)\nu_2) - \big(\nabla H(\alpha_1\nu_1 + (1-\alpha)\nu_2)\big)^\top\alpha(\nu_2-\nu_1) \\ &\quad \le -\underline{C}\alpha^2\|\nu_1-\nu_2\|^2. \end{aligned}$$

Multiplying the first inequality by α and the second inequality by $1-\alpha$, and summing the resulting inequalities, we obtain

$$\begin{aligned} &H(\alpha\nu_1 + (1-\alpha)\nu_2) - \alpha H(\nu_1) - (1-\alpha)H(\nu_2) \\ &\quad \ge \underline{C}\alpha(1-\alpha)\|\nu_1-\nu_2\|^2. \end{aligned} \tag{A.38}$$

Similarly, it follows from the first inequality in (A.36) that

$$H(\alpha\tilde{\nu}_1 + (1-\alpha)\tilde{\nu}_2) - \alpha H(\tilde{\nu}_1) - (1-\alpha)H(\tilde{\nu}_2) \leq \overline{C}\alpha(1-\alpha)\|\tilde{\nu}_1 - \tilde{\nu}_2\|^2. \quad \text{(A.39)}$$

Meanwhile,

$$\tilde{\nu}_1(x) - \tilde{\nu}_2(x) = \sum_{x'} \pi(x|x', a)\big(\nu_1(x') - \nu_2(x')\big)$$

implies

$$\big|\tilde{\nu}_1(x) - \tilde{\nu}_2(x)\big| \leq \sum_{x'} \pi(x|x', a)\big|\nu_1(x') - \nu_2(x')\big|.$$

Therefore there exists a sufficiently large constant C such that

$$\|\tilde{\nu}_1 - \tilde{\nu}_2\|^2 \leq C\|\nu_1 - \nu_2\|^2. \quad \text{(A.40)}$$

Plugging (A.38), (A.39), and (A.40) into (A.37), we obtain

$$G(\alpha\nu_1 + (1-\alpha)\nu_2) - \alpha G(\nu_1) - (1-\alpha)G(\nu_2) \geq \big[\underline{C} - \beta\overline{C}C\big]\alpha(1-\alpha)\|\nu_1 - \nu_2\|^2,$$

which is not smaller than zero with $\beta \leq \frac{\underline{C}}{\overline{C}C}$ for any α, ν_1 and ν_2. Therefore G is concave for $\beta \in (0, \beta_0]$ where $\beta_0 = \min\left\{\frac{\underline{C}}{\overline{C}C}, 1\right\}$.

(ii) If there exists $\epsilon > 0$ such that $\pi(x'|x, a) \geq \epsilon$ for any $a \in A$, $x, x' \in X$, then $\tilde{\nu}(x|a) = \sum_{x'} \pi(x|x', a)\nu(x') \geq \epsilon$ for any a and x. Therefore $\tilde{\nu}$ is uniformly bounded away from the boundary of $\Delta(X)$. In the closed subset $\Delta^{\epsilon}(X) = \{\mu \in \Delta(X) \,|\, \mu(i) \geq \epsilon, i = 1, \ldots, M\}$ of $\Delta(X)$, all second order partial derivatives of H are bounded uniformly from below. Then the first inequality of (A.36) follows from a second-order Taylor expansion with the Lagrange remainder. The rest of the proof follows from that for part (i). □

The second inequality in (A.36) means that H is strongly concave in $\Delta(X)$. It is satisfied for the Shorrocks entropy index with $\rho > 0$. This is because the Hessian matrix for H is $\text{Hess}(\nu) \equiv (H_{ij}(\nu))_{1\leq i,j\leq M}$ with $H_{ii}(\nu) = -\nu(i)^{-\rho}$ and $H_{ij}(\nu) = 0$ for $i \neq j$. Therefore, the matrix $\text{Hess}(\nu) + I_M$, where I_M is an $M \times M$ identity matrix, is negative definite for any $\nu \in [0, 1]^{|X|}$ and $\rho > 0$, implying that the second inequality of (A.36) is satisfied.

B Proofs for the main text

Proof Lemma 1 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, given a strategy (d, σ), we can construct the choice rule $\{p_t\}$ as in Sect. 2 and define a sequence of joint distributions $\mu_t\left(x^t, a^{t-1}\right)$ as follows

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

for $t \geq 1$, where $\mu_1\left(x^1, a^0\right) = \mu_1(x_1)$. We can then construct the sequences of posteriors $\left\{\mu_t\left(\cdot|a^t\right)\right\}$ and default rules $\left\{q_t\left(\cdot|a^{t-1}\right)\right\}$. The distribution induced by the strategy (d, 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 $\left\{\mu_{t+1}\left(x^{t+1}, a^t\right)\right\}$, $\sum_{t=1}^T \beta^{t-1}\mathcal{I}(\mathbf{x}_t; \mathbf{a}_t|\mathbf{a}^{t-1})$, is not larger than $\sum_{t=1}^T \beta^{t-1}\mathcal{I}(\mathbf{x}_t; \mathbf{s}_t|\mathbf{s}^{t-1})$ associated with (d, σ). These information costs can be computed using the posteriors and predictive distributions (priors) induced by the corresponding joint distributions.

Formally, by the definition of the discounted UPS cost, we compute

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

Rearranging the terms in the discounted UPS cost yields

$$
\begin{aligned}
&\sum_{t=1}^T \beta^{t-1}\mathcal{I}(\mathbf{x}_t; \mathbf{a}_t|\mathbf{a}^{t-1}) \\
&\quad = H(\mu_1) + \sum_{t=1}^{T-1} \sum_{a^t} \beta^{t-1} q_t(a^t)\left[\beta H\left(\mu_{t+1}(\cdot|a^t)\right) - H(\mu_t(\cdot|a^t))\right] \\
&\qquad - \beta^{T-1} \sum_{a^T} q_T(a^T) H(\mu_T(\cdot|a^T)) \\
&\quad = H(\mu_1) - \sum_{t=1}^{T-1} \sum_{a^t} \beta^{t-1} q_t(a^t)\, G^{a_t}(\mu_t(\cdot|a^t)) - \beta^{T-1} \sum_{a^T} q_T(a^T) H(\mu_T(\cdot|a^T)).
\end{aligned}
\tag{B.2}
$$

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

Now we prove that

$$
\sum_{a^t} q_t(a^t)\, G^{a_t}(\mu_t(\cdot|a^t)) \geq \sum_{s^t} q_t(s^t)\, G^{a_t}(\mu_t(\cdot|s^t)),
$$

where we use $q_t(a^t)$ and $q_t\left(s^t\right)$ to denote the marginal distributions of a^t and of s^t, respectively, by abuse of notation. 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^{a_t} is concave, it follows from Jensen's inequality that

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

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

$$\begin{aligned}&\sum_{a^t} q_t(a^t)\, G^{a_t}(\mu_t(\cdot|a^t))\\&\quad\geq \sum_{s^t}\sum_{a^t} \Pr(s^t|a^t) q_t(a^t)\, G^{a_t}(\mu_t(\cdot|s^t)) = \sum_{s^t} q_t(s^t)\, G^{\sigma_t(s^t)}(\mu_t(\cdot|s^t)),\end{aligned}$$

where $a_t = \sigma_t\left(s^t\right)$.

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

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

Applying the preceding two inequalities to the second and the third terms on the right-hand side of (B.2), we obtain

$$\sum_{t=1}^{T} \beta^{t-1} \mathcal{I}(\mathbf{x}_t; \mathbf{a}_t|\mathbf{a}^{t-1}) \leq \sum_{t=1}^{T} \beta^{t-1} \mathcal{I}(\mathbf{x}_t; \mathbf{s}_t|\mathbf{s}^{t-1}).$$

By the above analysis, we deduce that the optimal value from Problem 1 is not larger than that from Problem 2.

Conversely, given any sequences of posteriors $\left\{\mu_t\left(x_t|a^t\right)\right\}$ and default rules $\left\{q_t\left(a_t|a^{t-1}\right)\right\}$, we can use the Bayes rule

$$p_t\left(a_t|x_t, a^{t-1}\right) = \frac{\mu_t\left(x_t|a^t\right) q_t\left(a_t|a^{t-1}\right)}{\mu_t\left(x_t|a^{t-1}\right)},\ t \geq 1, \tag{B.3}$$

to construct the choice rule $\{p_t\}$, and follow the same argument as in Sect. 2 to construct a strategy (d, σ). The discounted expected payoff from this strategy is identical to the value of the objective function in Problem 2 given the sequences of posteriors and default rules. Because histories of actions are a subset of information signals, the optimal value from Problem 1 is not smaller than that from Problem 2. We then obtain the desired result. □

Proof of Proposition 1 Recall that the net utility is defined as

$$N_H^a\left(\mu\left(\cdot|a\right)\right) \equiv \sum_x \mu\left(x|a\right) u\left(x,a\right) + \lambda H\left(\mu\left(\cdot|a\right)\right). \tag{B.4}$$

To apply Lemma 3 of CD (2013), we compute

$$\frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(x|a)} = u(x,a) + \lambda H_x(\mu(\cdot|a)) - u(M,a) - \lambda H_M\left(\mu(\cdot|a)\right), \tag{B.5}$$

for any $x = 1, \ldots, M-1$, and

$$\begin{aligned}
&N_H^a(\mu(\cdot|a)) - \sum_{x=1}^{M-1} \frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(x|a)} \mu(x|a) \\
&= \sum_x \mu\left(x|a\right) u\left(x,a\right) + \lambda H\left(\mu\left(\cdot|a\right)\right) \\
&\quad - \sum_{x=1}^{M-1} \left[u(x,a) + \lambda H_x(\mu(\cdot|a)) - u(M,a) - \lambda H_M\left(\mu(\cdot|a)\right)\right] \mu(x|a) \\
&= u(M,a) + \lambda H\left(\mu\left(\cdot|a\right)\right) - \sum_{x=1}^{M-1} \left[\lambda H_x(\mu(\cdot|a)) - \lambda H_M\left(\mu(\cdot|a)\right)\right] \mu(x|a) \\
&= u(M,a) + \lambda H_M\left(\mu(\cdot|a)\right) + \lambda H\left(\mu\left(\cdot|a\right)\right) - \sum_{x=1}^{M} \lambda H_x(\mu(\cdot|a)) \mu(x|a).
\end{aligned} \tag{B.6}$$

We will show that conditions FOC-CA and FOC-UA are equivalent to conditions (ED), (CT), and (UB) in Lemma 3 of CD (2013), which is also equivalent to the Lagrangian lemma in CDL (2022) Lemma 1.

Step 1. We rewrite conditions (ED), (CT), and (UB) in explicit forms.

Condition (ED) is equivalent to

$$\begin{aligned}
&u(x,a) + \lambda H_x(\mu(\cdot|a)) - u(M,a) - \lambda H_M\left(\mu(\cdot|a)\right) \\
&\quad = u(x,b) + \lambda H_x(\mu(\cdot|b)) - u(M,b) - \lambda H_M\left(\mu(\cdot|b)\right),
\end{aligned} \tag{B.7}$$

for any chosen actions a, b, and any $x = 1, \ldots, M-1$.

Condition (CT) is equivalent to

$$u(M,a)+\lambda H_M\left(\mu(\cdot|a)\right)+\lambda H\left(\mu\left(\cdot|a\right)\right)-\sum_{x=1}^{M}\lambda H_x(\mu(\cdot|a))\mu(x|a)$$
$$=u(M,b)+\lambda H_M\left(\mu(\cdot|b)\right)+\lambda H\left(\mu\left(\cdot|b\right)\right)-\sum_{x=1}^{M}\lambda H_x(\mu(\cdot|b))\mu(x|b), \tag{B.8}$$

for any chosen actions a and b.

Condition (UB) is equivalent to

$$u\left(M,b\right)+\lambda H_M\left(\mu^b\right)+\lambda H\left(\mu^b\right)-\sum_{x=1}^{M}\lambda H_x(\mu^b)\mu^b(x)$$
$$\leq u(M,a)+\lambda H_M\left(\mu(\cdot|a)\right)+\lambda H\left(\mu\left(\cdot|a\right)\right)-\sum_{x=1}^{M}\lambda H_x(\mu(\cdot|a))\mu(x|a), \tag{B.9}$$

for any chosen action a, any unchosen action b, and $\mu^b\in\Delta(X)$ such that $\frac{\partial N_H^b(\mu^b(\cdot))}{\partial\mu^b(x)}=\frac{\partial N_H^a(\mu(\cdot|a))}{\partial\mu(x|a)}$ for any $x=1,\ldots,M-1$.

Step 2. We show that condition FOC-CA is equivalent to conditions (ED) and (CT).

Define function f as in (12). Then we can verify that conditions (B.7) and (B.8) imply (13). Conversely, let (13) hold. Then setting $x=M$, (13) implies condition (CT) in (B.8). It follows from (B.8) that

$$u(M,a)+\lambda H_M(\mu(\cdot|a))-[u(M,b)+\lambda H_M(\mu(\cdot|b))]=\lambda\left[f\left(\mu(\cdot|b)\right)-f\left(\mu(\cdot|a)\right)\right].$$

Condition (13) also implies that

$$u(x,a)+\lambda H_x(\mu(\cdot|a))-[u(x,b)+\lambda H_x(\mu(\cdot|b))]$$
$$=\lambda f\left(\mu\left(\cdot|b\right)\right)-\lambda f\left(\mu\left(\cdot|a\right)\right), \tag{B.10}$$

Combining the above two equations yield condition (B.7). Thus we have shown that condition FOC-CA in Proposition 1 is equivalent to conditions (CT) and (ED).

Step 3. We show that function $\widehat{V}(x)$ defined in Proposition 7 of Appendix A.1 is equal to the common value in (13).

For any chosen action a, the height of the hyperplane passing through $N_H^a(\mu(\cdot|a))$ at $\mu(\cdot)$ is given by

$$N_H^a(\mu(\cdot|a))-\sum_{x=1}^{M-1}\frac{\partial N_H^a(\mu(\cdot|a))}{\partial\mu(x|a)}(\mu(x|a)-\mu(x)).$$

By conditions (CT) and (ED) in Lemma 3 of CD (2013), the expression above is independent of any chosen action a. The height of this hyperplane at the prior with

$\mu(x) = 1$ and $\mu(x') = 0$ for $x' \neq x$ is

$$\widehat{V}(x) = N_H^a(\mu(\cdot|a)) - \sum_{y=1}^{M-1} \frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(y|a)} \mu(y|a) + \frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(x|a)}, \quad \text{(B.11)}$$

$$\widehat{V}(M) = N_H^a(\mu(\cdot|a)) - \sum_{y=1}^{M-1} \frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(y|a)} \mu(y|a), \quad \text{(B.12)}$$

for $x = 1, \dots, M-1$. By (12), (B.6), and (B.12), we have

$$\widehat{V}(M) = \lambda f(\mu(\cdot|a)) + u(M, a) + \lambda H_M(\mu(\cdot|a)). \quad \text{(B.13)}$$

By (B.5), (B.11), (B.12), and (B.13), we have

$$\begin{aligned} \widehat{V}(x) &= \lambda f(\mu(\cdot|a)) + u(M, a) + \lambda H_M(\mu(\cdot|a)) + \frac{\partial N_H^a(\mu(\cdot|a))}{\partial \mu(x|a)} \\ &= \lambda f(\mu(\cdot|a)) + u(x, a) + \lambda H_x(\mu(\cdot|a)), \end{aligned}$$

for $x = 1, ..., M-1$. Thus $\widehat{V}(x)$ is equal to the common value in (13).

Step 4. We show that condition (UB) in (B.9) is equivalent to FOC-UA.

By (B.5) and the definition of f, condition (UB) in (B.9) is equivalent to

$$u(M, b) + \lambda H_M(\mu^b) + \lambda f\left(\mu^b\right) \leq u(M, a) + \lambda H_M(\mu(\cdot|a)) + \lambda f(\mu(\cdot|a)), \quad \text{(B.14)}$$

where $q(a) > 0$ and $\mu^b \in \Delta(X)$ satisfies

$$\begin{aligned} &\left[u(x, b) + \lambda H_x(\mu^b)\right] - \left[u(M, b) + \lambda H_M(\mu^b)\right] \\ &= \left[u(x, a) + \lambda H_x(\mu(\cdot|a))\right] - \left[u(M, a) + \lambda H_M(\mu(\cdot|a))\right], \end{aligned} \quad \text{(B.15)}$$

for $x = 1, ..., M-1$. Notice that (B.14) and (B.15) imply

$$u(x, b) + \lambda H_x(\mu^b) + \lambda f\left(\mu^b\right) \leq u(x, a) + \lambda H_x(\mu(\cdot|a)) + \lambda f(\mu(\cdot|a)), \quad \text{(B.16)}$$

for $x = 1, 2, ..., M$. Conversely, suppose that (B.15) holds but (B.14) fails for some chosen action a and some action $b \in A$. Then we can check that (B.16) fails too. Thus we have shown that (B.16) is equivalent to condition (UB) in (B.9) given (B.15).

By (B.16), (13), and the definition of I_x, we obtain

$$\mu^b(x) \geq I_x\left(\frac{\widehat{V}(x)}{\lambda} - \frac{u(x, b)}{\lambda} - f(\mu^b); \mu^b\right), \; x = 1, \dots, M. \quad \text{(B.17)}$$

Since $\sum_x \mu^b(x) = 1$, this inequality implies (15). Here μ^b satisfies (B.15), which is equivalent to (14) using (13). Conversely, if (B.17) fails for some x, the previous argument using (B.15) shows that it also fails for all other x. Hence (15) fails as well. Therefore (UB) is equivalent to FOC-UA in Proposition 1. □

Proof of Proposition 2 We can easily verify the operator $\mathcal{T}$ satisfies the Blackwell sufficient condition. Thus it is a contraction mapping. Since $\Delta(X)$ is compact when endowed with the weak topology, continuous functions on this space are bounded. Thus $\mathbb{V}$ is a Banach space. We can verify that $\mathcal{T}$ maps a function in $\mathbb{V}$ into $\mathbb{V}$ by the theorem of the maximum. By the contraction mapping theorem, there is a unique fixed point $V \in \mathbb{V}$ such $V = \mathcal{T}V$. Moreover, $\lim_{s\to\infty} \mathcal{T}^s V^0 = V$ for any $V^0 \in \mathbb{V}$. Thus $\lim_{T\to\infty} V^T = V$. See Stokey, Lucas with Prescott (1989) for a reference of the cited theorems here. □

Proof of Theorem 1 By (17), the sequence of value functions satisfies the dynamic programming equations at any history a^{t-1} reached with positive probabilities:

$$\begin{aligned}
&V^{T-t+1}\left(\mu_t\left(\cdot|a^{t-1}\right)\right)\\
&\quad= \max_{\mu_t(\cdot|\cdot,a^{t-1}),q_t(\cdot|a^{t-1})} \sum_{x_t,a_t} q_t\left(a_t|a^{t-1}\right)\mu_t\left(x_t|a^t\right)u\left(x_t,a_t\right)\\
&\qquad-\lambda C_H\left(\mu_t\left(\cdot|a^{t-1}\right),\mu_t\left(\cdot|\cdot,a^{t-1}\right),q_t\left(\cdot|a^{t-1}\right)\right)\\
&\qquad+\beta\sum_{a_t} q_t\left(a_t|a^{t-1}\right)V^{T-t}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) \qquad \text{(B.18)}
\end{aligned}$$

subject to

$$\mu_t\left(x_t|a^{t-1}\right) = \sum_{a_t} q_t\left(a_t|a^{t-1}\right)\mu_t\left(x_t|a^t\right), \qquad \text{(B.19)}$$

and

$$\mu_{t+1}\left(x_{t+1}|a^t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t,a_t\right)\mu_t\left(x_t|a^t\right), \qquad \text{(B.20)}$$

for $t = 1, ..., T$. In the last period we have a terminal condition

$$V^0\left(\mu_{T+1}\left(\cdot|a^T\right)\right) = \sum_{x_{T+1}} \mu_{T+1}\left(x_{T+1}|a^T\right)U\left(x_{T+1}\right).$$

Starting from the last period T, we apply the analysis for the two-period problem in Appendix A.3 recursively by backward induction. We can then prove Theorem 1. Without repeating the arguments in Appendix A.3, here we only outline the key steps and omit the detailed derivation. Plugging equation

$$V^{T-t}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) = \sum_{x_{t+1}} \mu_{t+1}\left(x_{t+1}|a^t\right)\widehat{V}_{t+1}\left(x_{t+1}|a^t\right) - \lambda H\left(\mu_{t+1}\left(\cdot|a^t\right)\right)$$

into the above Bellman equation, we obtain

$$V^{T-t+1}\left(\mu_t\left(\cdot|a^{t-1}\right)\right)$$

$$= \max_{\mu_t(\cdot|\cdot,a^{t-1}),q_t(\cdot|a^{t-1})} \sum_{x_t,a_t} q_t\left(a_t|a^{t-1}\right) N_G^{a_t}\left(\mu_t\left(x_t|a^t\right)\right) - \lambda H\left(\mu_t\left(\cdot|a^{t-1}\right)\right)$$

where the net utility function is defined as

$$N_G^{a_t}\left(\mu_t\left(\cdot|a^t\right)\right) = \sum_{x_t} \mu_t\left(x_t|a^t\right) u\left(x_t, a_t\right) + \beta \overline{V}_{t+1}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) + \lambda G^{a_t}\left(\mu_t\left(\cdot|a^t\right)\right),$$

where

$$\overline{V}_{t+1}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) = \sum_{x_{t+1}} \mu_{t+1}\left(x_{t+1}|a^t\right) \widehat{V}_{t+1}\left(x_{t+1}|a^t\right).$$

By Assumption 1, $G^{a_t}\left(\mu_t\left(\cdot|a^t\right)\right)$ is concave in $\mu_t\left(\cdot|a^t\right)$. The concave envelope $\overline{V}_{t+1}\left(\mu_{t+1}\left(\cdot|a^t\right)\right)$ is concave in $\mu_{t+1}\left(\cdot|a^t\right)$ by Proposition 7 and hence in the posterior $\mu_t\left(\cdot|a^t\right)$ by (B.20). Thus the net utility function $N_G^{a_t}\left(\mu_t\left(\cdot|a^t\right)\right)$ is concave. We can then use Proposition 1 and Lemma 3 to characterize the solution. □

Proof of Theorem 2 By Definition 3, at history a^{t-1} reached with positive probabilities, $\mu_t\left(\cdot|a^{t-1}\right)$ takes the form $\mu_t\left(\cdot|a_{t-1}\right)$. Then the DM solves the following Bellman equation:

$$\begin{aligned} & V^{T-t+1}\left(\mu_t\left(\cdot|a_{t-1}\right)\right) \\ & = \max_{\mu_t(\cdot|\cdot,a^{t-1}),q_t(\cdot|a^{t-1})} \sum_{x_t,a_t} q_t\left(a_t|a^{t-1}\right) \mu_t\left(x_t|a^t\right) u\left(x_t, a_t\right) \\ & \quad - \lambda C_H\left(\mu_t\left(\cdot|a_{t-1}\right), \mu_t\left(\cdot|\cdot, a^{t-1}\right), q_t\left(\cdot|a^{t-1}\right)\right) \\ & \quad + \beta \sum_{a_t} q_t\left(a_t|a^{t-1}\right) V^{T-t}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) \end{aligned} \tag{B.21}$$

subject to

$$\mu_t\left(x_t|a_{t-1}\right) = \sum_{a_t} q_t\left(a_t|a^{t-1}\right) \mu_t\left(x_t|a^t\right), \tag{B.22}$$

$$\mu_{t+1}\left(x_{t+1}|a^t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t, a_t\right) \mu_t\left(x_t|a^t\right), \tag{B.23}$$

for $t = 1, ..., T$, with the terminal condition:

$$V^0\left(\mu_{T+1}\left(\cdot|a^T\right)\right) = \sum_x \mu_{T+1}\left(x|a^T\right) U(x).$$

As discussed in Sect. 4.3, the solution is a function of the prior/predictive distribution $\mu_t\left(x_t|a_{t-1}\right)$ independent of history a^{t-2}. Thus the optimal solution for $q_t\left(a_t|a^{t-1}\right)$ and $\mu_t\left(x_t|a^t\right)$ takes the form $q_t\left(a_t|a_{t-1}\right)$ and $\mu_t\left(x_t|a_{t-1}^t\right)$.

We can compute the state dependent choice probability:

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

for any $\mu_{t+1}(x_{t+1}|a^t) > 0$. □

Proof of Proposition 3 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 of CD (2013), $\mu_2(\cdot|a^2)$ is independent of $\mu_2(\cdot|a_1)$ and hence independent of a_1. Thus $\mu_2(\cdot|a^2)$ takes the form $\mu_2(\cdot|a_2)$. The predictive distribution in period $t = 3$ 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. Moreover, the optimal posterior $\mu_t\left(x_t|a^t\right)$ takes the form $\mu_t\left(x_t|a_t\right)$ for any $t \geq 1$. □

Proof of Proposition 4 We first check that the limiting solution from our algorithm satisfies Eq. (27) using Eqs. (28) and (29). If condition (i) is satisfied, then we have

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

where we simply write $\mu_t(x_t|a^t_{t-1}) = \mu_t(x_t|a_t)$ for any a_{t-1} with $q_t(a_t|a_{t-1}) > 0$. Therefore $\mu_{t+1}(x_{t+1}|a_t)$ defined in (29) satisfies (27). If condition (ii) is satisfied, then we have

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

for the unique a_{t-1} leading to a_t with probability 1. Therefore $q_t(a_t) = \sum_{a_{t-1}} q_t(a_t|a_{t-1})q_{t-1}(a_{t-1}) = q_{t-1}(a_{t-1})$ and (27) also holds.

We then prove that the limiting solution from our algorithm is Markovian. By Theorem 2, we only need to prove by induction that $\mu_t(x_t|a^{t-1})$ does not depend on the history a^{t-2} for any $t = 1, 2, \ldots, T$. This is trivially true for $t = 1$ and 2. When $t = 3$,

$$\mu_3(x_3|a_1^2) = \sum_{x_2} \pi(x_3|x_2, a_2)\mu_2(x_2|a_1^2) = \mu_3(x_3|a_2),$$

for any a_1 leading to a_2 with positive probabilities, due to (27) with $t = 2$. It then follows from (26) that the posterior $\mu_3(x_3|a_1^3)$ does not depend on a_1, i.e., it takes the form $\mu_3(x_3|a_2^3)$. As a result, using (27) with $t = 3$, we have

$$\mu_4(x_4|a_1^3) = \sum_{x_3} \pi(x_4|x_3, a_3)\mu_3(x_3|a_1^3) = \sum_{x_3} \pi(x_4|x_2, a_3)\mu_3(x_3|a_2^3) = \mu_4(x_4|a_3),$$

for any a_2 leading to a_3 with positive probabilities. Therefore $\mu_4(x_4|a_1^3)$ does not depend on the history a_1^2. By induction, this argument applies to any $t \leq T$. □

Proof of Proposition 5 When the transition kernel $\pi(\cdot|x_t, a_t) = \mu_1(\cdot)$ for any x_t and a_t, the prior beliefs given any a^t is the same as μ_1 because

$$\mu_{t+1}\left(x_{t+1}|a^t\right) = \sum_{x_t} \pi\left(x_{t+1}|x_t, a_t\right) \mu_t\left(x_t|a^t\right) = \mu_1\left(x_{t+1}\right), \tag{B.24}$$

for any $t \geq 1$ and $x_{t+1} \in X$. Let $T < \infty$. Then in period T, the solution is the static solution with prior μ_1. In any period $t \leq T$, the continuation value $V^{T-t}\left(\mu_{t+1}\left(\cdot|a^t\right)\right)$ given any history a^t reached with positive probabilities can be written as $V^{T-t}\left(\mu_1\right)$ independent of history a^t. Thus we have

$$\sum_{a_t} q_t\left(a_t|a^{t-1}\right) V^{T-t}\left(\mu_{t+1}\left(\cdot|a^t\right)\right) = V^{T-t}\left(\mu_1\right).$$

Consider the dynamic programming Eq. (B.18) given history a^{t-1}. The prior beliefs $\mu_t\left(\cdot|a^{t-1}\right) = \mu_1$. The solution for $\mu_t\left(\cdot|\cdot, a^{t-1}\right)$ and $q_t\left(\cdot|a^{t-1}\right)$ is the same as the static solution, independent of history a^{t-1} and the future continuation value $V^{T-t}\left(\mu_1\right)$. By backward induction until $t = 1$, we deduce that the solution in any period is the static solution. By Bayesian rule, the solution for the choice probabilities $p_t\left(a_t|x_t, a^{t-1}\right)$ is also the same as the static solution. For the infinite-horizon case, the result is obtained by taking limits as $T \to \infty$. □

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

$$\widetilde{V}_2(1,1) = \widetilde{V}_2(2,2) = \lambda \ln\left[z \exp(1/\lambda) + 1 - z\right],$$
$$\widetilde{V}_2(1,2) = \widetilde{V}_2(2,1) = \lambda \ln\left[(1-z)\exp(1/\lambda) + z\right],$$

$$\widetilde{v}_1(1,1) = \widetilde{v}_1(2,2) = 1 + \beta\alpha\widetilde{V}_2(1,1) + \beta(1-\alpha)\widetilde{V}_2(2,1),$$
$$\widetilde{v}_1(1,2) = \widetilde{v}_1(2,1) = \beta\alpha\widetilde{V}_2(2,2) + \beta(1-\alpha)\widetilde{V}_2(1,2).$$

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

$$\begin{aligned} V_1 &= \frac{1}{2}\lambda \ln \frac{1}{2}\left[\exp(\widetilde{v}_1(1,1)/\lambda) + \exp(\widetilde{v}_1(1,2)/\lambda)\right] \\ &\quad + \frac{1}{2}\lambda \ln \frac{1}{2}\left[\exp(\widetilde{v}_1(2,1)/\lambda) + \exp(\widetilde{v}_1(2,2)/\lambda)\right] \\ &= \lambda \ln \frac{1}{2}\left[\exp(\widetilde{v}_1(1,1)/\lambda) + \exp(\widetilde{v}_1(1,2)/\lambda)\right]. \end{aligned}$$

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 $\alpha \in \left[0, \alpha^{**}\right]$, 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. Suppose that it is optimal. Then we use $q_1(1) = 1$ and Corollary 1 to derive

$$V_1 = \frac{1}{2}\widetilde{v}_1(1,1) + \frac{1}{2}\widetilde{v}_1(2,1),$$

where $\widetilde{v}_1(1,1)$ and $\widetilde{v}_1(2,1)$ are given earlier. Since $\exp(x/\lambda)$ is a convex function of x, we obtain that

$$\frac{1}{2}\widetilde{v}_1(1,1) + \frac{1}{2}\widetilde{v}_1(2,1) < \lambda \ln \frac{1}{2}\left[\exp(\widetilde{v}_1(1,1)/\lambda) + \exp(\widetilde{v}_1(2,1)/\lambda)\right].$$

Since $\widetilde{v}_1(2,1) = \widetilde{v}_1(1,2)$ 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 $q_1(2) = 1$ is not optimal. $\square$

C Forward-backward algorithm

C.1 Static case

We first present an algorithm to solve the static RI Problem 3.[18] Our algorithm consists of seven steps:

1. Initialize $\overline{\mu}^a \in \Delta(A)$ and $q \in \Delta(A)$ with $q(a) > 0$ for all $a \in A$.
2. For any $a \in A$, compute

$$\overline{f}(a) = H\left(\overline{\mu}^a\right) - \sum_{x=1}^{M} H_x(\overline{\mu}^a)\overline{\mu}^a(x).$$

3. For any $x \in X$, compute $\widehat{V}(x)$ that satisfies the equation:

$$\mu(x) = \sum_a q(a) I_x\left(\frac{\widehat{V}(x)}{\lambda} - \frac{u(x,a)}{\lambda} - \overline{f}(a); \overline{\mu}^a\right). \tag{C.1}$$

4. For any $a \in A$ and $x \in X$, compute

$$\overline{\mu}^a_+(x) = I_x\left(\frac{\widehat{V}(x)}{\lambda} - \frac{u(x,a)}{\lambda} - \overline{f}(a); \overline{\mu}^a\right). \tag{C.2}$$

5. Update $\overline{\mu}^a$ and $q(a)$ by

$$\overline{\mu}^a_+(x) \to \overline{\mu}^a(x),$$

   and

$$q_+(a) = \sum_x \overline{\mu}^a_+(x)\, q(a) \to q(a). \tag{C.3}$$

6. Go back to step 2, until $\left(q_+, \overline{\mu}^a_+(x)\right)$ converges to $(q, \overline{\mu}^a(x))$.
7. Find $\mu^b \in \Delta(X)$ that satisfies (14). Check whether (15) is satisfied, where $\widehat{V}(x)$ is the converged value obtained in Step 6. If it is satisfied, then stop and a solution is found; otherwise, go to step 1 with a new guess.

It follows (C.2) and (C.3) that

$$q_+(a) = \sum_x I_x\left(\frac{\widehat{V}(x)}{\lambda} - \frac{u(x,a)}{\lambda} - \overline{f}(a); \overline{\mu}^a\right) q(a). \tag{C.4}$$

This equation shows that there are two types of limits: Either $q_+(a) = q(a) \in (0, 1]$ when condition (15) holds as equality or $q_+(a) = q(a) = 0$ when it holds as inequality. In the first case, action a is chosen and hence $\sum_x \overline{\mu}^a(x) = 1$ by (C.3). That is, $\overline{\mu}^a$

[18] Our algorithm and the Arimoto–Blahut algorithm are related to the general block coordinate descent method in the mathematics literature. See Bertsekas (2016) for a convergence analysis. It is beyond the scope of this paper to provide a convergence proof.

is the optimal posterior distribution $\mu(\cdot|a)$. In the second case, action a is not chosen. Condition (15) ensures the iteration in (C.3) to converge.

Step 7 checks the sufficient condition (15) in Proposition 1. To implement this step, we use (C.2) to substitute $\widehat{V}$ in (14) and derive

$$H_x\left(\mu^b\right) - H_M\left(\mu^b\right) = H_x\left(\overline{\mu}^b\right) - H_M\left(\overline{\mu}^b\right), x = 1, ..., M-1.$$

We use these $M-1$ equations together with $\sum_{x=1}^{M} \mu^b(x) = 1$ to numerically solve for $\mu^b \in \Delta(X)$. For the Shannon entropy case, we can derive a closed-form solution: $\mu^b(x) = \overline{\mu}^b(x)/\left[\sum_{x'} \overline{\mu}^b(x')\right]$.

C.2 Dynamic case

Now we present an algorithm to compute a Markovian solution to the dynamic RI Problem 2. It consists of the following steps:

1. Initialize $f_t^{(0)}(a_t|a_{t-1}) \in \mathbb{R}$, $q_t^{(0)}(\cdot|a_{t-1}) \in \Delta(A)$, and $\mu_t^{(0)}(\cdot|a_{t-1}) \in \Delta(X)$, with $q_t^{(0)}(a_t|a_{t-1}) > 0$ and $\mu_t^{(0)}(x_t|a_{t-1}) > 0$ for any $x_t \in X$, $a_t, a_{t-1} \in A$, and $t = 1, \ldots, T$. Set $\mu_1^{(0)}(\cdot|a_0) = \mu_1(\cdot)$.
2. Choose a large integer K. For $k = 1, 2, \ldots, K$ until $q_t^{(k)}(\cdot|\cdot)$ and $\mu_t^{(k)}(\cdot|\cdot)$ get sufficiently close to $q_t^{(k-1)}(\cdot|\cdot)$ and $\mu_t^{(k-1)}(\cdot|\cdot)$ for all $t = 1, \ldots, T$, do the following: **Backward path:** Initialize

$$v_T^{(k)}(x_T, a_T) = u(x_T, a_T) + \beta \sum_{x_{T+1}} \pi(x_{T+1}|x_T, a_T) U(x_{T+1}).$$

For $t = T, T-1, \ldots, 1$ do:

- For each $a_{t-1} \in A$, take $v_t^{(k)}$, $f_t^{(k-1)}(\cdot|a_{t-1})$, $q_t^{(k-1)}(\cdot|a_{t-1})$, and $\mu_t^{(k-1)}(\cdot|a_{t-1})$ as the input u, $f(\cdot)$, $q(\cdot)$, μ for the algorithm in the static case and use the Markovian version of Eqs. (26), (22), (21), and (20) to compute the update $f_t^{(k)}(\cdot|a_{t-1})$, $q_t^{(k)}(\cdot|a_{t-1})$, $\widehat{V}_t^{(k)}(\cdot|a_{t-1})$, and $\mu_t^{(k)}(\cdot|\cdot, a_{t-1})$.
- If $t \geq 2$, compute

$$\begin{aligned} v_{t-1}^{(k)}(x_{t-1}, a_{t-1}) &= u(x_{t-1}, a_{t-1}) \\ &+ \beta \sum_{x_t} \pi(x_t|x_{t-1}, a_{t-1}) \left[\widehat{V}_t^{(k)}(x_t|a_{t-1}) - \lambda H_{x_t}(\mu_t^{(k-1)}(\cdot|a_{t-1}))\right]. \end{aligned}$$

Forward path: For $t = 1, 2, \ldots, T-1$ do:

- If $t = 1$, compute

$$\mu_2^{(k)}(x_2|a_1) = \sum_{x_1} \pi(x_2|x_1, a_1) \mu_1^{(k)}(x_1|a_1).$$

- If $t \geq 2$, compute

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

$$\mu_{t+1}^{(k)}(x_{t+1}|a_t) = \frac{\mu_{t+1}^{(k)}(x_{t+1}, a_t)}{\mu_{t+1}^{(k)}(a_t)}, \; \mu_{t+1}^{(k)}(a_t) = \sum_{x_{t+1}} \mu_{t+1}^{(k)}(x_{t+1}, a_t) > 0,$$

where we set $\mu_2^{(k)}(a_1) = q_1^{(k)}(a_1)$.

3. Return $q_t^{(K)}(\cdot|\cdot)$, $\mu_t^{(K)}(\cdot|\cdot,\cdot)$ and $\widehat{V}_t^{(K)}(\cdot|\cdot)$.
4. Check whether the converged solution satisfies (27).

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

D Markovian versus history-dependent solutions

In this appendix we first characterize a Markovian solution and then provide two numerical examples to illustrate Markovian and history-dependent solutions for the Shannon entropy case.

Due to Theorem 2, Definition 2 is reduced to the following form for a Markovian solution.

Definition 4 The sequences of $\{\mu_t(\cdot|a_{t-1}^t)\}_{t=1}^T$ and $\{q_t(\cdot|a_{t-1})\}_{t=1}^T$ satisfy

(i) MFOC$_t$-CA, if for any chosen action $a_t \in A$ with $q_t(a_t|a_{t-1}) > 0$ and for any $x_t \in X$,

$$\widehat{V}_t(x_t|a_{t-1}) = v_t(x_t, a_t) + \lambda H_{x_t}(\mu_t(\cdot|a_{t-1}^t)) + \lambda f_t(\mu_t(\cdot|a_{t-1}^t), a_t), \tag{D.1}$$

where

$$v_t(x_t, a_t) = u(x_t, a_t) + \beta \sum_{x_{t+1}} \pi(x_{t+1}|x_t, a_t)\left[\widehat{V}_{t+1}(x_{t+1}|a_t) - \lambda H_{x_{t+1}}(\mu_{t+1}(\cdot|a_t))\mathbf{1}_{\{t<T\}}\right],$$

(ii) MFOC$_t$-UA, if for any unchosen action $a_t \in A$ with $q_t\,(a_t|a_{t-1}) = 0$ and $\mu_t^{a_t} \in \Delta\,(X)$ such that

$$\left[v_t(x_t, a^t) + \lambda H_{x_t}(\mu_t^{a_t})\right] - \left[v_t(M, a_t) + \lambda H_M(\mu_t^{a_t})\right] = \widehat{V}_t\,(x_t|a_{t-1}) - \widehat{V}_t\,(M|a_{t-1}), \tag{D.2}$$

for any $x_t \in X$, we have

$$\sum_{x_t} I_{x_t}\left(\widehat{V}_t(x_t|a_{t-1})/\lambda - v_t(x_t, a_t)/\lambda - f_t(\mu_t^{a_t}, a_t); \mu_t^{a_t}\right) \leq 1. \tag{D.3}$$

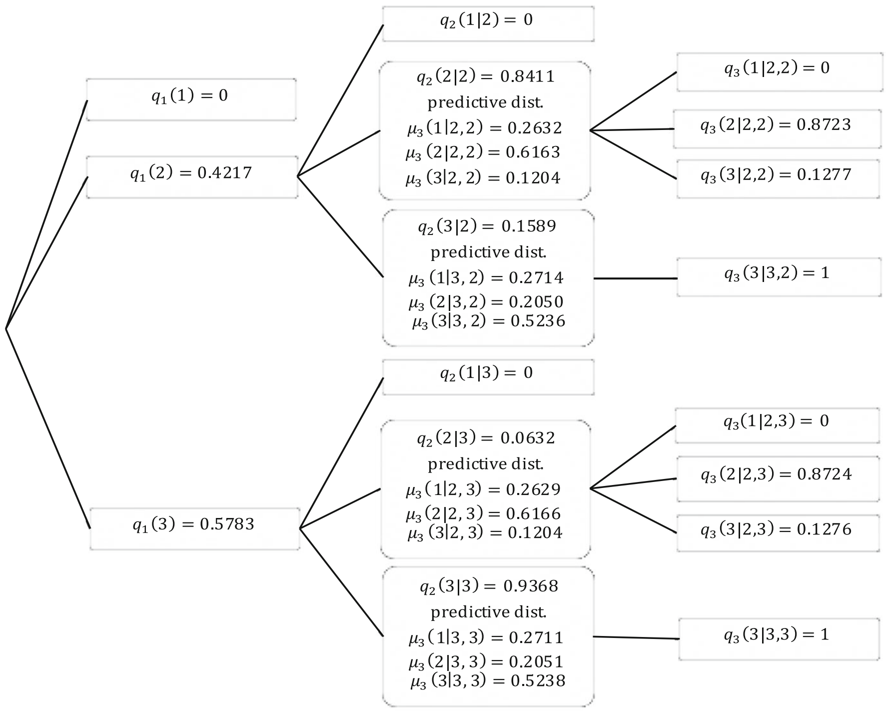

Fig. 5 Markovian solution

Combining Theorems 1 and 2, we obtain the following characterization of a Markovian solution.

Proposition 8 *Suppose that Assumptions 1 holds. Then the sequences of* $\left\{\mu_t\left(x_t|a_{t-1}^t\right)\right\}_{t=1}^T$ *and* $\{q_t\,(a_t|a_{t-1})\}_{t=1}^T$ *are a Markovian solution to dynamic RI Problem 2 if and only if they satisfy: (i) the Eqs. (26) and (27); (ii) MFOC$_t$-CA and MFOC$_t$-UA for* $t = 1, \ldots, T$. *The value function satisfies*

$$V^{T+1-t}(\mu_t(\cdot|a_{t-1})) = \sum_{x_t} \mu_t(x_t|a_{t-1})\widehat{V}_t(x_t|a_{t-1}) - \lambda H(\mu_t(\cdot|a_{t-1})),\ t = 1, ..., T. \tag{D.4}$$

Now we use both the fully history-dependent forward-backward Arimoto–Blahut algorithm and the Markovian version described in Appendix C 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 $\pi\,(x_{t+1}|x_t, a_t) = 1 - \gamma$ if $x_{t+1} = x_t$; $\pi\,(x_{t+1}|x_t, a_t) = \gamma/2$ 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 5 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

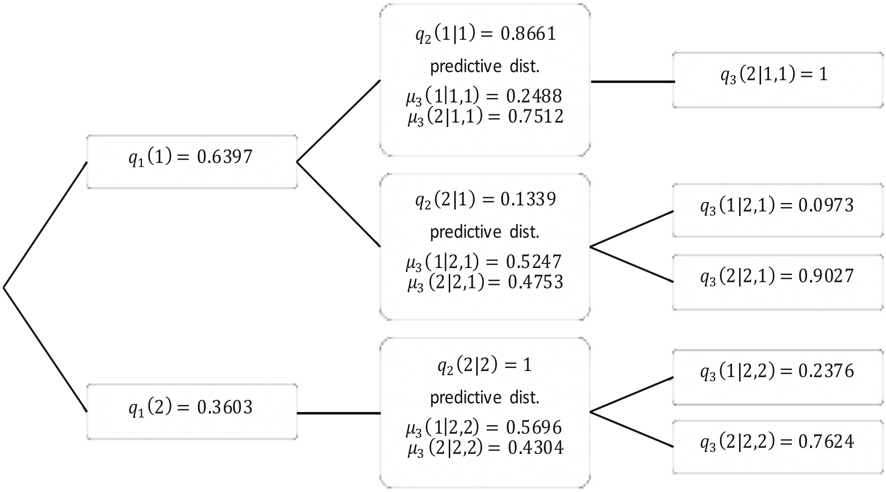

Fig. 6 History-dependent solution

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 C gives an almost identical numerical solution. Notice that this solution is not interior, a case not covered by SSM (2017).

For the second example, let $T = 3$, $u_{T+1} = 0$, $X = A = \{1, 2\}$, and the transition kernel satisfy $\pi_t(x_{t+1}|x_t, a_t) = 1-\gamma_t$ if $x_{t+1} = x_t$; $\pi_t(x_{t+1}|x_t, a_t) = \gamma_t$ 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. Figure 6presents 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 C 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.

References

Arimoto, S.: An algorithm for computing the capacity of arbitrary discrete memoryless channels. IEEE Trans. Inf. Theory **18**, 14–20 (1972)

Bertsekas, D.: Nonlinear Programming. Athena Scientific, Nashua (2016)

Blahut, R.E.: Computation of channel capacity and rate-distortion functions. IEEE Trans. Inf. Theory **18**, 460–473 (1972)

Bloedel, A., Zhong, W.: The cost of optimally-acquired information. Working Paper. Stanford University (2021)

Caplin, A., Dean, M.: Behavioral Implications of Rational Inattention with Shannon Entropy. NBER Working Paper 19318 (2013)

Caplin, A., Dean, M.: Revealed preference, rational inattention, and costly information acquisition. Am. Econ. Rev. **105**, 2183–2203 (2015)

Caplin, A., Dean, M., Leahy, J.: Rational inattention, optimal consideration sets, and stochastic choice. Rev. Econ. Stud. **86**, 1061–1094 (2019)
Caplin, A., Dean, M., Leahy, J.: Rationally inattentive behavior: characterizing and generalizing Shannon entropy. J. Polit. Econ. **130**, 1676–1715 (2022)
Catonini, E., Mayskaya, T.: Talking with an extremist. Econ. Theory (2023). https://doi.org/10.1007/s00199-023-01500-5
Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, Hoboken (2006)
Dean, M., Neligh, N.: Experimental tests of rational inattention. J. Polit. Econ. (2019) (forthcoming)
Denti, T.: Posterior separable cost of information. Am. Econ. Rev. **112**, 3215–3259 (2022)
Denti, T., Marinacci, M., Rustichini, A.: Experimental cost of information. Am. Econ. Rev. **112**, 3106–3123 (2022)
Dewan, A., Neligh, N.: Estimating information cost functions in models of rational inattention. J. Econ. Theory **187**, 1–32 (2020)
Fudenberg, D., Strzalecki, T.: Dynamic logit with choice aversion. Econometrica **83**, 651–691 (2015)
Hébert, B., Woodford, M.: Rational Inattention in Continuous Time. Working Paper. Columbia University (2018)
Hébert, B., Woodford, M.: Neighborhood-based information costs. Am. Econ. Rev. **111**, 3225–3255 (2021)
Jung, J., Kim, J.-H., Matějka, F., Sims, C.A.: Discrete actions in information-constrained decision problems. Rev. Econ. Stud. **86**, 2643–2667 (2019)
Kamenica, E.: Bayesian persuasion and information design. Annu. Rev. Econ. **11**, 249–272 (2019)
Luo, Y.: Consumption dynamics under information processing constraints. Rev. Econ. Dyn. **11**, 366–385 (2008)
Maćkowiak, B., Wiederholt, M.: Optimal sticky prices under rational inattention. Am. Econ. Rev. **99**, 769–803 (2009)
Maćkowiak, B., Matějka, F., Wiederholt, M.: Rational inattention: a review. J. Econ. Lit. (2018) (forthcoming)
Matějka, F., McKay, A.: Rational inattention to discrete choices: a new foundation for the multinomial logit model. Am. Econ. Rev. **105**, 272–298 (2015)
Mattsson, L.-G., Weibull, J.W.: Probabilistic choice and procedurally bounded rationality. Games Econom. Behav. **41**, 61–78 (2002)
Miao, J.: Multivariate LQG Control under Rational Inattention in Continuous Time. Working Paper. Boston University (2019)
Miao, J., Su, D.: Asset market equilibrium under rational inattention. Econ. Theor. **75**(1), 1–30 (2023). https://doi.org/10.1007/s00199-021-01396-z
Miao, J., Wu, J., Young, E.: Multivariate rational inattention. Econometrica **90**, 907–945 (2022)
Mondria, J.: Portfolio choice, attention allocation, and price comovement. J. Econ. Theory **145**, 1837–1864 (2010)
Morris, S., Strack, P.: The Wald problem and the equivalence of sequential sampling and static information costs. Working Paper. MIT (2019)
Nimark, K.P., Sundaresan, S.: Inattention and belief polarization. J. Econ. Theory **180**, 203–228 (2019)
Peng, L., Xiong, W.: Investor attention, overconfidence and category learning. J. Financ. Econ. **80**, 563–602 (2006)
Pomatto, L., Strack, P., Tamuz, O.: The cost of information. Am. Econ. Rev. **113**, 1360–1393 (2023)
Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, Hoboken (2005)
Ravid, D.: Ultimatum bargaining with rational inattention. Am. Econ. Rev. **110**, 2948–2963 (2020)
Rust, J.: Structural estimation of Markov decision processes. In: Engle, R., McFadden, D. (eds.) Handbook of Econometrics, vol. 4, pp. 3082–3139. North Holland, Amsterdam (1994)
Shannon, C.E.: A mathematical theory of communication. Bell Syst. Tech. J. **27**, 379–423 (1948)
Shorrocks, A.F.: The class of additively decomposable inequality measures. Econometrica **48**, 613–25 (1980)
Sims, C.A.: Stickiness. Carn.-Roch. Conf. Ser. Public Policy **49**, 317–356 (1998)
Sims, C.A.: Implications of rational inattention. J. Monet. Econ. **50**(3), 665–690 (2003)
Sims, C.A.: Rational inattention and monetary economics. In: Friedman, B.M., Woodford, M. (eds.) Handbook of Monetary Economics, vol. 3A. North-Holland, Amsterdam (2011)
Steiner, J., Stewart, C., Matějka, F.: Rational inattention dynamics: inertia and delay in decision-making. Econometrica **85**, 521–553 (2017)

Stokey, N.L., Lucas, R.E., Prescott, E.C.: Recursive Methods in Economic Dynamics. Harvard University Press, Cambridge (1989)

Tanaka, T., Sandberg, H., Skoglund, M.: Transfer-entropy-regularized Markov decision processes. IEEE Trans. Autom. Control **67**, 1944–1951 (2022)

Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. J. Stat. Phys. **52**, 479–487 (1988)

Van Nieuwerburgh, S., Veldkamp, L.: Information acquisition and under-diversification. Rev. Econ. Stud. **77**, 571–626 (2010)

Woodford, M.: Information constrained state dependent pricing. J. Monet. Econ. **56**(S), S100–S124 (2009)

Woodford, M.: Inattentive Valuation and Reference-dependent Choice. Working Paper. Columbia University (2012)

Zhong, W.: Optimal dynamic information acquisition. Econometrica **90**, 1537–1582 (2022)