Rational Inattention when Decisions Take Time*

Benjamin Hébert † Michael Woodford ‡
Stanford University Columbia University

October 21, 2019

Abstract

Decisions take time, and the time taken to reach a decision is likely to be informative about the cost of more precise judgments. We formalize this insight in the context of a dynamic rational inattention (RI) model. Under standard conditions on the flow cost of information in our discrete-time model, we obtain a tractable model in the continuous-time limit. We next provide conditions under which the resulting belief dynamics resemble either diffusion processes or processes with large jumps. We then demonstrate that the state-contingent choice probabilities predicted by our model are identical to those predicted by a static RI model, providing a micro-foundation for such models. In the diffusion case, our model provides a normative foundation for a variant of the DDM models studied in mathematical psychology.

*The authors would like to thank Mark Dean, Sebastian Di Tella, Mira Frick, Xavier Gabaix, Matthew Gentzkow, Emir Kamenica, Divya Kirti, Jacob Leshno, Stephen Morris, Pietro Ortoleva, José Scheinkman, Ilya Segal, Ran Shorrer, Joel Sobel, Miguel Villas-Boas, Ming Yang, and participants at the Cowles Theory conference, 16th SAET Conference, Barcelona GSE Summer Conference on Stochastic Choice, Stanford GSB research lunch, 2018 ASSA meetings, UC Berkeley Theory Seminar, and UC San Diego Theory for helpful discussions on this topic, and the NSF for research support. We would particularly like to thank Philipp Strack and Doron Ravid for discussing an earlier version of the paper. Portions of this paper circulated previously as the working papers "Rational Inattention with Sequential Information Sampling," "Rational Inattention in Continuous Time," and "Information Costs and Sequential Information Sampling," and appeared in Benjamin Hébert's Ph.D. dissertation at Harvard University. All remaining errors are our own.

†Hébert: Stanford University. Email: bhebert@stanford.edu.
‡Woodford: Columbia University. Email: mw2230@columbia.edu.

1 Introduction

It is common in economic modeling to assume that, when presented with a choice set, a decision maker (DM) will choose the option that is ranked highest according to a coherent preference ordering. However, observed choices in experimental settings often appear to be random, and while this could reflect random variation in preferences, it is often more sensible to view choice as imprecise. Models of rational inattention (such as Matêjka et al. [2015]) formalize this idea by assuming that the DM chooses her action based on a signal (a subjective assessment of the situation) that provides only an imperfect indication of the true state. The information structure that generates this signal is optimal, in the sense of allowing the best possible joint distribution of states and actions, net of a cost of information. In the terminology of Caplin and Dean [2015], models of rational inattention make predictions about patterns of state-dependent stochastic choice. These predictions will depend in part on the nature of the information cost, and several recent papers have attempted to recover information costs from observed behavior in laboratory experiments (Caplin and Dean [2015], Dean and Neligh [2019]).

However, in both laboratory experiments and real-world economic settings, decisions take time, and the time required to make a decision is likely to be informative about the nature of information costs. In this paper, we develop a framework to study rational inattention problems in which decisions take time, providing a means of connecting decision times to information costs and state-dependent stochastic choice.

There is an extensive literature in mathematical psychology that focuses on these issues. Variants of the drift-diffusion model (DDM, Ratcliff [1985], Ratcliff and Rouder [1998]) also make predictions about stopping times and state-dependent stochastic choice.[1] In particular, these models are designed to match the empirical observation that hasty decisions are likely to be of lower quality.[2] However, these models are not based on optimizing behavior, and this raises a question as to the extent to which they can be regarded as structural; it is unclear how the parameters of the DDM model should be expected to change when incentives or the costs of delay change, and this limits the use of the model for making counter-factual predictions. The framework we develop includes as a special case variants of the DDM model, while at the same time making predictions about state-dependent stochastic choice that match those of a static rational inattention model. Consequently, our framework is able to both speak to the relationship between stopping times and state-dependent stochastic choice (unlike standard rational inattention models) and make counter-factual predictions (unlike standard DDM models).

Specifically, we develop a class of rational inattention models in which the DM's imprecise perception of the decision problem is not merely random, but evolves with the passage of time, and a constrained optimization problem determines a joint probability distribution over both stopping

[1] DDM models were originally developed to explain imprecise perceptual classifications. See Woodford [2019] for a more general discussion of the usefulness of the analogy between perceptual classification errors and imprecision in economic decisions.

[2] The existence of a speed-accuracy trade-off is well-documented in perceptual classification experiments (e.g., Schouten and Bekker [1967]). Variants of the DDM that have been fit to stochastic choice data include Busemeyer and Townsend [1993] and more recently Krajbich et al. [2014] and Clithero [2018]; see Fehr and Rangel [2011] for a review of other early work. Shadlen and Shohamy [2016] provide a neural-process interpretation of sequential-sampling models of choice.

times and choices. We then demonstrate that the resulting state-dependent stochastic choice probabilities of our continuous-time model are equivalent to those of a static rational inattention model. Any cost function for a static rational inattention model in the uniformly posterior-separable family (in the terminology of Caplin et al. [2019]) can be justified in our framework. This result offers both a justification for using such cost functions in static rational inattention problem and provides a means of connecting those cost functions to dynamic processes for beliefs. Our results also provide a relationship between information costs in the static problem and decision times in the dynamic problem, and hence offer a new perspective on how the costs of information for rational inattention models might be uncovered from observable behavior.

Moreover, we provide conditions under which the process for beliefs in our model resembles a diffusion process, allowing our model to be interpreted as a variant of the DDM model. Our results therefore contribute to the literature on DDM-style models by offering an optimizing model that makes predictions about how decision boundaries and choice probabilities should change in response to changes in incentives, in particular by showing that these changes can be calculated using the comparative statics of a static rational inattention problem.

We begin by considering a discrete-time dynamic model of optimal information sampling, in which at each of the discrete time steps a random signal is obtained. We then develop our continuous-time framework as the limit of our discrete-time model when the length of the time steps between successive signals becomes small. In our discrete-time model, we assume that the information sampling process is optimally chosen from within a very flexible class of possibilities, in the spirit of models of rational inattention (Sims [2010]), with a flow information cost function that satisfies conditions commonly assumed in static models of rational inattention. The key step of our convergence proof is our demonstration that all such flow information cost functions can be approximated, for signal structures sufficiently close to uninformative, by a posterior-separable (in the terminology of Caplin et al. [2019]) flow information cost function.

Standard rational inattention models consider a static problem, in which a decision is made after a single noisy signal is obtained by the DM. This allows the set of possible signals to be identified with the set of possible decisions, which is no longer true in our dynamic setting. Steiner et al. [2017] also discuss a dynamic model of rational inattention, but their model is one in which, because of the kind of information cost assumed, it is never optimal to acquire information other than what is required for the current action. As a result, in each period of their discrete-time model, the set of possible signals can again be identified with the possible actions at that time. We instead consider situations in which evidence is accumulated over time before any action is taken, as in the DDM; this requires us to model the stochastic evolution of a belief state that is not simply an element of the set of possible actions.[3] Our central concerns are to study the conditions under which the resulting continuous-time model of optimal information sampling gives rise to belief dynamics and stochastic choices similar to those implied by a DDM-like model, and to study how variations in the

[3]The model differs from the one analyzed by Steiner et al. [2017] in several key respects. First, as just noted, we study a setting in which the DM can take an action only once, and chooses when to stop and take an action endogenously. Second, we consider a much more general class of flow cost functions than the mutual information cost assumed in their paper. And third, we assume that the DM has a motive to smooth her information gathering over time, rather than learn all of the relevant information at a single point in time.

opportunity cost of time or the payoffs of actions should affect stochastic choice.

A number of prior papers have also sought to endogenize aspects of a DDM-like process. Moscarini and Smith [2001] consider both the optimal intensity of information sampling per unit of time and the optimal stopping problem, when the only possible kind of information is given by the sample path of a Brownian motion with a drift that depends on the unknown state, as assumed in the DDM.[4] Fudenberg et al. [2018] consider a variant of this problem with a continuum of possible states, and an exogenously fixed sampling intensity.[5] Woodford [2014] instead takes as given the kind of stopping rule posited by the DDM, but allows a very flexible choice of the information sampling process, as in theories of rational inattention. Our approach differs from these earlier efforts in seeking to endogenize *both* the nature of the information that is sampled at each stage of the evidence accumulation process and the stopping rule that determines how much evidence is collected before a decision is made.[6]

We place minimal restrictions on the nature of the flow information costs, and yet obtain relatively sharp conclusions about the nature of the continuous-time limit. While we allow in general for decisions about both stopping and information sampling to be arbitrary functions of the compete history of information sampled to that point, we show that it is possible to characterize the dynamics in terms of the evolution of the belief state in a finite-dimensional space, with stopping if and only if one of the stopping regions associated with the different possible actions is reached. We further show that under relatively weak conditions the dimensionality of the space in which beliefs move will be one less than the number of actions (thus, for example, a line in the case of a binary decision problem, as assumed in the DDM). Moreover, we show in general that the dynamics of the belief state prior to stopping can be described by a jump-diffusion process, and we give conditions under which it will be either a pure diffusion (as assumed in the DDM) or a pure jump process (as in the models of Che and Mierendorff [2019] and Zhong [2019]).

We also offer, under relative general conditions, a characterization of the boundaries of the stopping regions and the predicted ex ante probabilities of different actions, as functions of model parameters including the opportunity cost of time. The key to this characterization is a demonstration that in a broad class of cases, both the stopping regions and the ex ante choice probabilities for any given initial prior are the same as in a static RI problem with an appropriately chosen static information cost function. Thus in addition to providing foundations for interest in DDM-like models of the decision process, our paper provides novel foundations for interest in static RI problems of particular types. For example, we provide conditions under which the predictions of our model will be equivalent to those of a static RI model with the mutual-information cost function proposed by Sims [2010]) — and thus equivalent to the model of stochastic choice analyzed by Matêjka et al. [2015] — but the foundations that we provide for this model do not rely on an analogy with rate-distortion

[4] Moscarini and Smith [2001] allow the instantaneous variance of the observation process to be freely chosen (subject to a cost), but this is equivalent to changing how much of the sample path of a given Brownian motion can be observed by the DM within a given amount of clock time.

[5] See also Tajima et al. [2016] for analysis of a related class of models, and Tajima et al. [2019] for an extension to the case of more than two alternatives.

[6] Both Morris and Strack [2019] and Zhong [2019] adopt our approach, and obtain special cases of the relationship between static and dynamic models of optimal information choice that we present below. Other recent dynamic models of optimal evidence accumulation include Che and Mierendorff [2019] and Zhong [2019], which differ from our treatment in not considering conditions under which beliefs will evolve as a diffusion process.

theory in communications engineering (the original motivation for the proposal of Sims).

More generally, we show that any cost function for a static RI model in the uniformly posterior-separable family studied by Caplin et al. [2019] can be justified by the process of sequential evidence accumulation that we describe. This includes the neighborhood-based cost functions discussed in Hébert and Woodford [2018], that lead to predictions that differ from those of the mutual-information cost function in ways that arguably better resemble the behavior observed in experiments such as those of Dean and Neligh [2019]. Our result provides both a justification for using such cost functions in static RI problems, and an answer (not given by static RI theory alone) to the question of how the cost function in an equivalent static RI model should change in the case of a change in the opportunity cost of time.

The connection that we establish between the choice probabilities implied by a dynamic model of optimal evidence accumulation and those implied by an equivalent static RI model holds both in the case that the belief dynamics in the dynamic model are described by a pure diffusion process and in the case that they are described by a jump process; thus we also show that with regard to these particular predictions, these two types of dynamic models are equivalent. However, this does not mean that no observations would make it possible to distinguish between them. We show that the predictions of the two types of model with regard to the distribution of decision times can be different, so that it should be possible in principle to use empirical evidence to determine which better describes actual decision making. And to the extent that we are interested in the predictions of our model for decision times and not solely for choices (as a number of authors have argued we should be), it matters which specification is the more realistic one.

We begin in section 2 by defining static RI problems, and introducing the discrete-time dynamic problem that we study and the continuous-time problem that is its limit. In section 3, we state the conditions that we impose on the flow information cost function, and present our result that all cost functions satisfying these conditions resemble posterior-separable cost functions in a local approximation. Section 4 that demonstrates that the continuous-time model described in section 2 represents a limiting case of our discrete-time problem. In section 4 we also describe conditions under which beliefs follow a diffusion or a diffusion-like process, and conditions under which the belief dynamics instead involve large jumps. In section 5 we demonstrate that the state-dependent choice probabilities predicted by the continuous time models (in both the diffusion case and the jump case) are equivalent to those predicted by a static rational inattention model with a uniformly posterior-separable cost function. In section 6 we discuss how models involving jumps in beliefs and models involving diffusions can in principle be distinguished using stopping time data. In section 7 we conclude.

2 Static and Dynamic Models of Rational Inattention

We begin by describing the class of static rational inattention models surveyed by Sims [2010], and then describe the discrete and continuous time dynamic models we study.

Notation: given a set X, we define $\mathcal{P}(X)$ as the probability simplex associated with that set. We describe an element of the simplex $r \in \mathcal{P}(X)$ by a vector in $\mathbb{R}_+^{|X|}$ whose elements sum to one, each of which corresponds to the likelihood of a particular element of $x \in X$. Except when necessary, we will call this vector r as well, and will not distinguish between the element of the simplex and its coordinate representation. We use the notation r_x to refer to the probability under r that $x \in X$ occurs. We discuss additional notation that appears in our proofs in the appendix, Section §B.

2.1 Static Models of Rational Inattention

Let $x \in X$ be the underlying state of nature, and let $s \in S$ be a signal the decision maker (DM) can receive, which might convey information about the state. We assume that X and S are finite sets. Let q denote the DM's prior belief (before receiving a signal) about the probability of state x; that is, q is element of the probability simplex $\mathcal{P}(X)$. Define $p_{s,x}$ as the probability of receiving signal s in state x, let $p_x \in \mathcal{P}(S)$ be the associated conditional probability distribution of the signals given state x, and let p be the $|S| \times |X|$ matrix whose elements are $p_{s,x}$. The matrix p, which is a set of conditional probability distributions for each state of nature, $\{p_x\}_{x \in X}$, defines an "information structure."

By Bayes' rule, the DM believes under her prior that the unconditional probability of receive a signal $s \in S$ is $\pi_s(p,q) = \sum_{x \in X} p_{s,x} q_x$. After receiving signal s, the DM will hold a posterior, $q_s(p,q) \in \mathcal{P}(X)$, which is defined by

$$q_{s,x}(p,q) = \frac{p_{s,x} q_x}{\pi_s(p,q)}$$

for all $s \in S$ such that $\pi_s(p,q) > 0$. We adopt the convention that $q_s(p,q) = q$ for all $s \in S$ such that $\pi_s(p,q) = 0$.

Let $a \in A$ be the action taken by the decision maker (DM). For simplicity, A is also a finite set, and we assume that the number of states is weakly larger than the number of actions, $|X| \geq |A|$. The DM's utility from taking action a in state x is $u_{a,x}$. We assume that $u_{a,x}$ is strictly positive and bounded above by a positive constant, $\bar{u}$.

The maximum achievable expected payoff, given an information structure p and prior q, can be written as

$$u(p,q) \equiv \max_{\{a(s)\}} \sum_{x \in X} \sum_{s \in S} q_x p_{s,x} u_{a(s),x}. \tag{1}$$

The standard static rational inattention problem, given the signal alphabet S,[7] is then

$$\max_{\{p_x \in \mathcal{P}(S)\}_{x \in X}} u(p,q) - C(p,q;S), \tag{2}$$

where

$$C(\cdot,\cdot;S) : \mathcal{P}(S)^{|X|} \times \mathcal{P}(X) \to \mathbb{R} \tag{3}$$

is a cost function for information structures.

In the classic formulation of Sims, a problem of the form equation (2) is considered, in which the

[7] The full problem includes a choice over the signal alphabet S. A standard result, which will hold for all of the cost functions we study, is that $|S| = |A|$ is sufficient.

cost function $C(p, q; S)$ is proportional to the Shannon mutual information between the signal and the state. Mutual information can be defined using Shannon's entropy,

$$H^{Shannon}(q) \equiv -\sum_{x \in X} q_x \ln(q_x). \tag{4}$$

Shannon's entropy can in turn be used to define a measure of the degree to which the posterior $q' \in \mathcal{P}(X)$ associated with any signal differs from the prior q, the Kullback-Leibler (KL) divergence,

$$D_{KL}(q'||q) \equiv H^{Shannon}(q) - H^{Shannon}(q') + (q' - q)^T \cdot H_q^{Shannon}(q), \tag{5}$$

where $H_q^{Shannon}(q)$ denotes the gradient of Shannon's entropy evaluated at beliefs q. Mutual information is then the expected value of the KL divergence over possible signals, and the cost of an information structure is assumed to be proportional to the mutual information,

$$C^{MI}(p, q; S) = \theta \sum_{s \in S} \pi_s(p, q) D_{KL}(q_s(p, q)||q), \tag{6}$$

where $\theta > 0$ is a positive constant. Mutual information is a measure of the informativeness of the signal, in that it provides a measure of the degree to which the signal changes what one should believe about the state, on average. However, Shannon's mutual information is not the only possible measure of the informativeness of an information structure, or the only plausible cost function for a static rational inattention problem. We discuss two large classes of cost functions, both of which will feature prominently in our analysis, below. For specific alternative proposals, see Caplin et al. [2019], Hébert and Woodford [2018], Pomatto et al. [2018].

Caplin et al. [2019] define a large class of cost functions, the "posterior-separable" class, which can written as

$$C^{PS}(p, q; S) = \sum_{s \in S} \pi_s(p, q) D(q_s(p, q)||q), \tag{7}$$

where $D(\cdot||\cdot)$ is a divergence. Recall that a divergence is defined as a function $D : \mathcal{P}(X) \times \mathcal{P}(X) \to \mathbb{R}_+$, with $D(q'||q) = 0$ if and only if $q' = q$. Divergences can be thought of as distances between probability distributions, although they are not necessarily symmetric and do not necessarily satisfy the triangle inequality. Mutual information is a posterior-separable cost function, whose associated divergence is (proportional to) the KL divergence.

Within this class, a cost function is said to be "uniformly posterior-separable" if the divergence D is a Bregman divergence, meaning that

$$D(q'||q) = H(q') - H(q) - (q' - q)^T \cdot H_q(q) \tag{8}$$

for some convex function $H : \mathcal{P}(X) \to \mathbb{R}$. The KL divergence is a Bregman divergence, whose associated convex function is the negative of Shannon's entropy. These two classes, the posterior-separable cost functions and the uniformly posterior-separable cost functions, will play a key role in our analysis.

In our derivation of the continuous time limit of our discrete time model, we adopt a more general approach (following De Oliveira et al. [2017] and Caplin and Dean [2015]) and study the entire class of costs functions satisfying certain conditions. We introduce these conditions after describing our discrete and continuous time dynamic models.

2.2 Dynamic Discrete Time Models of Rational Inattention

We next extend the static rational inattention model just described to a discrete time dynamic setting, in which the DM has numerous opportunities to gather information before taking an action. Our extension is similar in spirit to the one described by Steiner et al. [2017], with several significant differences. First, unlike those authors, we study a setting in which the DM can take an action only once, and chooses when to stop and take an action endogenously. Second, we are interested in general cost functions of the form described by equation (3), not just mutual information. Third, we assume the DM has a motive to smooth her information gathering over time, rather than learn all of the relevant information in a single period.

As in the static model, there is a state of the world, $x \in X$, that remains constant over time. At each time t, the DM can either stop and take an action $a \in A$, or continue and receive a signal drawn from the information structure $\{p_{t,x} \in \mathcal{P}(S)\}_{x \in X}$, for some signal alphabet S. We assume that the signal alphabet S is finite and fixed over time, with $|S| \geq 2|X|+2$. However, the information structure $\{p_{t,x}\}_{x \in X}$ is a choice variable that can depend on the past history of signal realizations. Fixing the signal alphabet S has no economic meaning, because the information content of receiving a particular signal $s \in S$ can change over time; our assumption that S is finite allows us to avoid technicalities associated with infinite-dimensional signal structures. We also assume, as a technical device, that S contains one signal, $\bar{s}$, that is required to be uninformative. This assumption ensures that the DM can choose to mix any arbitrary signal structure with an uninformative one, even if she has already used up her "useful" signals.

The DM's prior beliefs at time t, before receiving the signal, are denoted q_t. Each time period has a length Δ. Let τ denote the time at which the DM stops and makes a decision, with $\tau = 0$ corresponding to making a decision without acquiring any information. At this time, the DM receives utility $u_{a,x} - \kappa\tau$ if she takes action a at time τ and the true state of the world is x. Let $\hat{u}(q_\tau)$ be the utility (not including the penalty for delay) associated with taking an optimal action under beliefs q_τ:

$$\hat{u}(q_\tau) = \max_{a \in A} \sum_{x \in X} q_{\tau,x} u_{a,x}.$$

We assume, as in the static model, that $u_{a,x}$ is strictly positive and bounded above by the constant $\bar{u}$, and that $|A| \leq |X|$.

Each time period has length Δ, and the DM discounts the future exponentially, with discount factor β^{Δ}, for some $\beta \in (0, 1]$. The parameter κ and discount factor β together govern the size of the penalty the DM faces from delaying his decision. The DM does not always make a decision immediately because she is able to gather information over time before making a more-informed decision.

The DM can choose an information structure that depends on the current time and past history of the signals received. As we will see, the problem has a Markov structure, and the current time's "prior," q_t, summarizes all of the relevant information that the DM needs to design the information structure. The DM is constrained to satisfy

$$E_0[\frac{\Delta}{\rho}\sum_{j=0}^{\tau\Delta^{-1}-1}\beta^{j\Delta}C(p_{j\Delta},q_{j\Delta};S)^{\rho}]^{\frac{1}{\rho}} \leq \Delta c E_0[\Delta\sum_{j=0}^{\tau\Delta^{-1}-1}\beta^{j\Delta}]^{\frac{1}{\rho}}, \tag{9}$$

if the DM choose to acquire any information at all ($\tau > 0$ always in this case).[8] In words, the L^{ρ}-norm of the flow information cost function $C(\cdot)$ over time and possible histories must be less that the constant c per unit time. In the particular case in which information gathering is constant ($C(p_{j\Delta},q_{j\Delta};S) = \bar{C}$), this constraint simplifies to

$$\frac{\bar{C}}{\Delta} \leq \rho^{\frac{1}{\rho}}c,$$

emphasizing that the constant c can be thought of as a limit on the flow rate of information acquisition.

The parameter ρ governs the substitutability of information acquisition across time. In the limit as $\rho \to \infty$, the L^{ρ} norm becomes the essential supremum, and the constraint approaches a per-period constraint on the amount of information the DM can obtain. For finite values of ρ, the DM can allocate more information gathering to states and times in which it is more advantageous to gather more information. We assume, however, that $\rho > 1$, to ensure that it is optimal for the DM to gather information gradually, rather than all at once. Our assumption of $\rho > 1$ is similar to the convex cost of the rate of experimentation assumed by Moscarini and Smith [2001]. It is a critical assumption that separates our model from the model of Steiner et al. [2017], in which the DM will (under some circumstances) learn a large amount of information in a single period. (Further assumptions about the flow cost function $C(\cdot)$ are introduced in the next section.)

Let $V(q_0;\Delta)$ denote the value obtained in the sequence problem for a DM with prior beliefs q_0, and let q_{τ} denote the DM's beliefs when stopping to act. The DM's problem is

$$V(q_0;\Delta) = \max_{\{p_{j\Delta}\},\tau} E_0[\beta^{\tau}\hat{u}(q_{\tau}) - \Delta\kappa\frac{1-\beta^{\tau}}{1-\beta^{\Delta}}], \tag{10}$$

subject to the information-cost constraint (9).

Note that the constraint (9) can equivalently be written in a form that takes each side of the inequality to the power ρ; we then obtain a constraint that is additively separable over time and across possible histories. This allows us to write a Lagrangian for the DM's problem in the additively

[8]The factor Δ inside the operator $\mathrm{E}_0[\cdot]$ on both sides of this inequality might seem redundant. We include it so that the expression inside the operator on the left-hand side approaches a time integral of discounted information costs in the limit as Δ is made arbitrarily small.

separable form

$$W(q_0,\lambda;\Delta)=\max_{\{p_{j\Delta}\},\tau} E_0[\beta^\tau \hat{u}(q_\tau)-\Delta\kappa\frac{1-\beta^\tau}{1-\beta^\Delta}]-$$
$$\lambda E_0[\Delta^{1-\rho}\sum_{j=0}^{\tau\Delta^{-1}-1}\beta^{j\Delta}\{\frac{1}{\rho}C(p_{j\Delta},q_{j\Delta};S)^\rho-\Delta^\rho c^\rho\}], \tag{11}$$

where $\lambda \geq 0$ is a Lagrange multiplier associated with the information-cost constraint. With this definition, the value function defined in (10) corresponds to $V(q_0;\Delta)=\min_{\lambda\geq 0} W(q_0,\lambda;\Delta)$. An optimal information sampling policy then necessarily maximizes $W(q_0,\lambda;\Delta)$ for some value of $\lambda \geq 0$. This alternative formulation of the problem can be thought of as the value function of a different problem, in which there is a cost of gathering information proportional to $\lambda\frac{1}{\rho}C(\cdot)^\rho$.

We will next analyze the continuous time limits of these functions,

$$W(q_0,\lambda)=\lim_{\Delta\to 0^+} W(q_0,\lambda;\Delta)$$

and

$$V(q_0)=\lim_{\Delta\to 0^+} V(q_0;\Delta).$$

We will first present the limiting continuous time problem, and then state assumptions that ensure this continuous time problem is the limit of the discrete time problem. Our approach is to assume standard conditions on the flow cost function $C(\cdot)$, which we describe in the next section, and then characterize the cost of a small amount of information under these conditions. We demonstrate that, to an approximation of first order in Δ, every cost function satisfying our conditions is equivalent to a posterior-separable cost function. Because of our assumption that $\rho > 1$, it will be optimal, as $\Delta \to 0^+$, for the DM to acquire a small amount of information each period, and consequently our results about the cost of a small amount of information will determine the form of the continuous time limit.

2.3 Dynamic Continuous Time Models of Rational Inattention

Here we introduce the continuous time models that are the limits of our discrete time model. We will discuss two cases: a model with exponential discounting ($\beta < 1$) and a model with no discounting ($\beta = 1$). In both of these models, beliefs follow a jump-diffusion process. In our analysis of the models (section §4), we will discuss conditions under which optimal information sampling implies that beliefs follow pure jump or pure diffusion processes.

Our continuous time models are defined in terms of a divergence, $D^*(r||q)$. This divergence is continuously twice-differentiable in r for r sufficiently close to q, and we define an $|X|\times|X|$ positive semi-definite matrix-valued function $k(q)$ by

$$\frac{\partial^2 D^*(r||q)}{\partial r^i \partial r^j}|_{r=q} \;=\; Diag(q)k(q)Diag(q), \tag{12}$$

where $Diag(q)$ is an $|X|\times|X|$ diagonal matrix whose diagonal is the vector q. We call this $k(q)$

function the "information cost matrix" and we prove it satisfies certain properties as part of our derivation, which we describe in section §4.

The key step in our derivation of the continuous time model is Theorem 1, which demonstrates that, up to first-order in Δ, every cost function satisfying our conditions is posterior-separable:

$$C(p,q;S) = \Delta \sum_{s\in S} \pi_s(p,q) D^*(q_s(p,q)||q) + o(\Delta) \tag{13}$$

for all sufficiently uninformative p (for a more formal statement, see Theorem 1 below). Loosely, there are two ways for a signal structure to be close to uninformative. First, it might have signals that occur frequently but whose associated posteriors are close to the value of the prior. Second, it might have signals that occur rarely but whose posteriors are far from the prior. These two kinds of signals correspond in the continuous time limit, respectively, to a diffusion component and a jump component of the stochastic process for beliefs.[9]

We prove in the limit as $\Delta \to 0^+$ that the process q_t converges to a jump-diffusion:

$$dq_t = -\bar{\psi}_t z_t dt + z_t dJ_t + Diag(q_t)\sigma_t dB_t,$$

where dJ_t is a poisson process with intensity $\bar{\psi}_t$, z_t is change in beliefs conditional on a jump, dB_t is an $|X|$-dimensional Brownian motion, and σ_t is an $|X| \times |X|$ matrix that controls the diffusion component of beliefs. Because beliefs must remain in the simplex, $q_{t^-} + z_t$ is absolutely continuous with respect to q_{t^-}, and $q_t^T \sigma_t = \vec{0}$.

The flow information cost in this limit is

$$C(\sigma_t, \bar{\psi}_t, z_t; q_t)dt = \frac{1}{2} tr[\sigma_t \sigma_t^T k(q_t)]dt + \bar{\psi}_t D^*(q_t + z_t||q_t)dt,$$

which the continuous time analog of equation (13).

We now introduce the continuous time analog of the (dual) dynamic discrete time problem (equation (11)), which we will refer to as $W^+(q_t, \lambda)$.

Definition 1. The dual continuous time problem with discounting ($\beta < 1$) is

$$\begin{aligned} W(q_t,\lambda) = & \sup_{\{\sigma_s \in \mathbb{R}^{|X|\times|X|}, \bar{\psi}_s \in \mathbb{R}_+, z_s \in \mathbb{R}^{|X|}\}, \tau \in \mathbb{R}_+} E_t[\beta^{(\tau-t)} \hat{u}(q_\tau) - \frac{1}{-\ln(\beta)}(1-\beta^{(\tau-t)})(\kappa - \lambda c^\rho)] - \\ & - \frac{\lambda}{\rho} E_t[\int_t^\tau \beta^{(s-t)} \{\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_s)] + \bar{\psi}_s D^*(q_{s^-} + x_s||q_{s^-})\}^\rho ds], \end{aligned}$$

subject to the evolution of beliefs,

$$dq_t = -\bar{\psi}_t x_t dt + x_t dJ_t + Diag(q_t)\sigma_t dB_t,$$

[9] Signals whose frequency converges to zero and whose posterior converges to the prior can in fact be informative, if they converge at slow enough rates. When the frequency does not converge to zero, the signal is a diffusion in the continuous time limit. When the posterior does not converge to a prior, the signal is a jump in the continuous time limit. When both converge but sufficiently slowly, the signal becomes a non-jump-diffusion semi-martingale in the continuous time limit. However, part of our convergence proof demonstrates that it is without loss of generality to assume beliefs follow a jump-diffusion process.

where dJ_t is a Poisson process with intensity $\bar{\psi}_t$ and dB_t is an $|X|$-dimensional Brownian motion, and the constraints that $q_{t^-} + x_t \ll q_{t^-}$ and $q_t^T \sigma_t = \vec{0}$, and the constraint that, for all stopping times T measurable with respect filtration generated by q_s^* ,

$$E_t[\int_t^T \beta^{(s-t)}\{\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_t)] + \bar{\psi}_s D^*(q_{s^-} + x||q_{s^-})\}ds] \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} E_t[\frac{1-\beta^{(T-t)}}{-\ln(\beta)}],$$

where θ is a positive constant.

For the problem with discounting, there exists some λ^* such that the limit of the original sequence problem, $\lim_{n\to\infty} V(q_0;\Delta_n) = V(q_0)$, is equal to $W^+(q_t,\lambda^*)$.

In the particular case where the DM has no exponential discounting ($\beta = 1$), we demonstrate that the amount of information acquired at each moment ($C(\sigma_t, \psi_t, x_t; q_t)$ above) is constant. This property comes from the fact that the cost of delay is constant. In the case with discounting ($\beta < 1$), the cost of delay depends in part on the current level of the value function, which generates variation and leads to a non-constant flow cost function. In the $\beta = 1$ case, using the fact that the quantity of information acquired is constant, the problem can be written in a simpler form. We introduce this simpler form below, and explicitly characterize both the dual continuous time problem and the original sequence problem.

Definition 2. The dual continuous time problem without discounting ($\beta = 1$) is, for all $\lambda \in (0, \kappa c^{-\rho})$,

$$W(q_t,\lambda) = \sup_{\{\sigma_s\in\mathbb{R}^{|X|\times|X|},\bar{\psi}_s\in\mathbb{R}_+,z_s\in\mathbb{R}^{|X|}\},\tau\in\mathbb{R}_+} E_t[\hat{u}(q_{\tau^*}) - (\tau-t)\frac{\rho}{\rho-1}(\kappa - \lambda c^\rho)]$$

subject to the constraints that $q_{t^-} + z_t \ll q_{t^-}$, $q_t^T\sigma_t = \vec{0}$, and

$$\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_t)] + \bar{\psi}_s D^*(q_{s^-} + z_s||q_{s^-}) \leq \chi(\lambda),$$

for all times $s \in [t,\tau)$ and some constant $\chi(\lambda) > 0$.

The limit of the original sequence problem, $\lim_{n\to\infty} V(q_0;\Delta_n) = W(q_t,\lambda^*) = V(q_0)$, is

$$V(q_t) = \sup_{\{\sigma_s\in\mathbb{R}^{|X|\times|X|},\bar{\psi}_s\in\mathbb{R}_+,z_s\in\mathbb{R}^{|X|}\},\tau\in\mathbb{R}_+} E_t[\hat{u}(q_{\tau^*}) - (\tau-t)\kappa]$$

subject to the constraints that $q_{t^-} + z_t \ll q_{t^-}$, $q_t^T\sigma_t = \vec{0}$, and

$$\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_t)] + \bar{\psi}_s D^*(q_{s^-} + z_s||q_{s^-}) \leq \rho^{\frac{1}{\rho}} c.$$

These continuous time problems are described using the "unconditional" dynamics for the beliefs q_t, meaning that beliefs are martingales. Using Bayes' rule, we can also define the conditional dynamics for beliefs. Suppose that $x_{true} \in X$ is the true state of nature. The beliefs q_t follow the dynamics process

$$dq_t = Diag(q_t)\sigma_t\sigma_t^T e_{x_{true}} - \bar{\psi}_t z_t dt + x_t dJ_t^{x_{true}} + Diag(q_t)\sigma_t dB_t, \tag{14}$$

where $e_{x_{true}}$ is a vector equal to one in the state corresponding to x_{true} and zero otherwise, and $dJ_t^{x_{true}}$ is a Poisson process with arrival rate $\bar{\psi}_t(1 + \frac{e_{x_{true}}^T z_t}{e_{x_{true}}^T q_t})$. Intuitively, beliefs will tend to move towards x_{true}. Consider in particular the value $e_{x_{true}}^T dq_t$. There is a positive (by the positive definiteness of $\sigma_t\sigma_t^T$) drift term associated with the diffusion of beliefs, and jumps are more likely if $e_{x_{true}}^T z_t$ is positive and less likely if it is negative. Certain other models in the literature (e.g. DDM models, see Fudenberg et al. [2018]) are usually expressed in terms of the conditional dynamics of beliefs, and equation (14) allows us to relate our continuous time models to these other models.

Before proceeding to our derivation, we briefly discuss the differences between the $\beta < 1$ and $\beta = 1$ cases, and of the role of jumps vs. diffusion. Beginning with the issue of jumps vs. diffusions, observe that there is a kind of continuity between the two. Because the cost of both jumps and diffusion in beliefs are determined by the same divergence, D^*, there exists a limit in which the jump becomes very likely ($\bar{\psi}_t \to \infty$) and very small ($z_t \to 0$) and for which the stochastic process of beliefs and the cost of the information converge to the stochastic process and cost of a diffusion process. That is, it is without loss of generality in both our continuous time problems to assume that $\sigma_t = 0$, because the supremum of a sequence of jump controls $(\bar{\psi}_t, z_t)$ can perfectly replicate a diffusion process. Relatedly, although both of these problems appear to allow the DM to have jumps in only a single direction z_t, nothing requires that the DM choose controls $(\bar{\psi}_t, z_t)$ that vary smoothly in the state q_t. Consequently, the DM is perfectly able to replicate a process with jumps in multiple directions. These statements are proven as part of the proof of our convergence result, Theorem 2 below.

Turning now to the issue of $\beta < 1$ vs. $\beta = 1$, observe first that many decisions are made over short periods of time (seconds or minutes). With conventional rates of time preference, β should be extremely close to one. As we will demonstrate, in the $\beta = 1$ case, the model is tractable and we are able (under certain additional assumptions) to characterize the value function. Consequently, provided that behavior is continuous in the limit as β approaches one (and we will show that it is), we believe it is reasonable to focus on the $\beta = 1$ model.

Moreover, as noted by Fudenberg et al. [2018], when $\beta < 1$, the dynamics of beliefs are not invariant to non-action-contingent transformations of the utility function. That is, increasing the utility function will change the DM's behavior, because with $\beta < 1$ a higher value function creates a higher cost of delay. We view this as undesirable, and describe in Section §6 an experimentally feasible test of this implication of $\beta < 1$. This association between the cost of delay and the level of the value function is the key property that leads to the result of Zhong [2019] that, in a special case of our setting, jumps in beliefs and not diffusions are optimal. This result might lead one to conclude that there is a discontinuity between the $\beta < 1$ and $\beta = 1$ cases. However, we show below that continuity holds between the $\beta < 1$ and $\beta = 1$ cases. That is, in situations in which the $\beta = 1$ model leads to a diffusion process for beliefs, the $\beta < 1$ model has beliefs that are a sequence of tiny jumps, whose magnitude is on the order of $-\ln(\beta)$, and the belief process converges in the limit to a diffusion.

Having introduced static, discrete time dynamic, and continuous time dynamic models of rational inattention, we next outline the key conditions we impose on flow cost functions in the discrete time model that lead the continuous time model as the limit.

3 Flow Information Costs

At each date in the discrete-time sequential rational inattention problem introduced above, the DM chooses an information structure. Each information structure p has a flow cost $C(p, q; S)$, given by a function of the form of (3), where q is the DM's prior in this date (that is, the posterior beliefs following from observations prior to the current date of the dynamic problem), and S is the signal alphabet.[10] Our results depend only on assuming that this flow information-cost function satisfies a set of five general conditions, stated below.

All of these conditions are satisfied by the mutual-information cost function (6) proposed by Sims, but they are also satisfied by many other cost functions (for example, the neighborhood-based cost functions we describe in Hébert and Woodford [2018]). They are closely related to conditions that other authors (De Oliveira et al. [2017] and Caplin and Dean [2015]) have also proposed as attractive general properties to assume about information-cost function, in the context of static rational inattention models.

Condition 1. Information structures that convey no information ($p_x = p_{x'}$ for all x, x' in the support of q) have zero cost. All other information structures have a strictly positive cost.

This condition ensures that the least costly strategy for the DM in the standard static rational inattention problem is to acquire no information, and make her decision based on the prior. The requirement that gathering no information has zero cost is a normalization.

The next condition is called mixture feasibility by Caplin and Dean [2015]. Consider two information structures, $\{p_{1,x}\}_{x \in X}$, with signal alphabet S_1, and $\{p_{2,x}\}_{x \in X}$, with alphabet S_2. Given a parameter $\lambda \in (0,1)$, we define a mixed information structure, $\{p_{M,x}\}_{x \in X}$ over the signal alphabet $S_M = (S_1 \cup S_2) \times \{1,2\}$. For each $s = (s_1, 1)$ in the alphabet S_M, $p_{M,x}(s)$ is equal to $\lambda p_{1,x}(s)$ if $s_1 \in S_1$, and equal to 0 otherwise. Likewise, for each $s = (s_2, 2)$, $p_{M,x}(s)$ is equal to $(1-\lambda)p_{2,x}(s)$ if $s_2 \in S_2$, and equal to 0 otherwise.

That is, this information structure results, with probability λ, in a posterior associated with information structure p_1, and with probability $1 - \lambda$ in a posterior associated with information structure p_2. The distribution of posteriors under the mixed information structure is a convex combination of the distributions of posteriors under the two original information structures, as if the DM flipped a coin and used the result to chose one of the two information structures, and (crucially) remembered the result of the coin flip. The mixture feasibility condition requires that choosing a

[10] The information-cost functions that we study, like mutual information, are defined for all finite signal alphabets S. Note, however, that mutual information is also defined over alternative sets of states of nature X. We do not impose this requirement on our more general cost functions — all of our analysis takes the set of states of nature as given.

mixed information structure costs no more than the cost of randomizing over information structures (using a mixed strategy in the rational inattention problem).

Condition 2. Given two information structures, $\{p_{1,x}\}_{x\in X}$, with signal alphabet S_1, and $\{p_{2,x}\}_{x\in X}$, with alphabet S_2, the cost of the mixed information structure is weakly less than the weighted average of the cost of the separate information structures:

$$C(p_M, q; S_M) \leq \lambda C(p_1, q; S_1) + (1-\lambda)C(p_2, q; S_2).$$

The next condition uses Blackwell's ordering. Consider two signal structures, $\{p_x\}_{x\in X}$, with signal alphabet S, and $\{p'_x\}_{x\in X}$, with alphabet S'. The first information structure Blackwell dominates the second information structure if, for all utility functions $u_{a,x}$ and all priors $q \in \mathcal{P}(X)$,

$$u(p,q) \;\geq\; u(p',q),$$

where $u(p,q)$ is defined as in equation (1). In words, if one information structure Blackwell dominates another, it is weakly more useful for every decision maker, regardless of that decision maker's utility function and prior. In this sense, it conveys weakly more information. This ordering is incomplete; most information structures neither dominate nor are dominated by a given alternative information structure. However, when an information structure does Blackwell dominate another one, we assume that the dominant information structure is weakly more costly.

Condition 3. If the information structure $\{p_x\}_{x\in X}$ with signal alphabet S is more informative, in the Blackwell sense, than $\{p'_x\}_{x\in X}$, with signal alphabet S', then, for all $q \in \mathcal{P}(X)$,

$$C(p_x\}_{x\in X}, q; S) \geq C(p'_x\}_{x\in X}, q; S').$$

The first three conditions are, from a certain perspective, almost innocuous. For any joint distribution of actions and states that could have been generated by a DM solving a static rational inattention problem, with an arbitrary information cost function, there is a cost function consistent with these three conditions that also could have generated that data (Theorem 2 of Caplin and Dean [2015]). The result arises from the possibility of the DM pursuing mixed strategies over information structures and mixed strategies in the mapping between signals and actions. These conditions also characterize "canonical" rational inattention cost functions, in the terminology of De Oliveira et al. [2017].

We demonstrate that the mixture feasibility condition (Condition 2) and Blackwell monotonicity condition (Condition 3) are equivalent to requiring that the cost function be convex over information structures and Blackwell monotone.

Lemma 1. *Let p and p' be information structures with signal alphabet S. A cost function is convex in information structures if, for all $\lambda \in (0,1)$, all signal alphabets S, and all $q \in \mathcal{P}(X)$,*

$$C(\lambda p + (1-\lambda)p', q; S) \leq \lambda C(p,q;S) + (1-\lambda)C(p', q; S).$$

A cost function satisfies mixture feasibility and Blackwell monotonicity (Conditions 2 and 3) if and only if it is convex in information structures and satisfies Blackwell monotonicity.

Proof. See the appendix, section B.1. □

This result is useful because it allows us to restrict attention to cost functions that are convex in the signal structure, which is the choice variable at each date.

The fourth condition that we assume, which is not imposed by Caplin and Dean [2015], Caplin et al. [2019], or De Oliveira et al. [2017], is a differentiability condition that will allow us to characterize the local properties of our cost functions. Let $p^0(p,q)$ be an uninformative signal structure, with $p^0_{x,s}(p,q) = \pi_s(p,q)$ for all $x \in X$ and $s \in S$. That is, for each state $x \in X$, with the signal structure p^0, the likelihood of receiving each signal $s \in S$ is equal to the unconditional likelihood, and hence the signal is completely uninformative. Treating p and $p^0(p,q)$ as elements of $\mathbb{R}^{|X|\times|S|}$, let $||p - p^0(p,q)||$ be an arbitrary norm on the difference between two signal structures.

Condition 4. There exists an $\epsilon > 0$ such that, for all signal alphabets S and priors q and all information structures p sufficiently close to uninformative ($||p-p^0(p,q)|| < \epsilon$), the information cost function is continuously twice-differentiable with respect to p in all directions that do not change the support of the signal distribution, and directionally differentiable, with a continuous directional derivative, with respect to perturbations that increase the support of the signal distribution.

While this may seem a relatively innocuous regularity condition, it is not completely general; for example, it rules out the case in which the DM is constrained to use only signals in a parametric family of probability distributions, and the cost of other information structures is infinite. Thus it rules out information structures of the kind assumed in Fudenberg et al. [2018] or Morris and Strack [2019]. Condition 4 also rules out other proposed alternatives, such as the channel-capacity constraint suggested by Woodford [2012]. The "$\epsilon > 0$" part of the condition indicates that that this differentiability need only hold at nearly uninformative information structures; obviously, if differentiability holds everywhere, the condition will be satisfied.

We should also note that this condition imposes the assumption that the cost function is finite for posteriors on the boundary of the simplex. This feature of the condition is convenient but unnecessary for our results; we could weaken the condition to require directional differentiability with respect to perturbations that increase the support of the signal distribution only for perturbations for which the resulting posteriors are interior to the simplex defined by the support of q. We should also observe that, because the convex function is convex in p, the existence of a directional derivative is guaranteed. What the condition is imposing is that this directional derivative be continuous.

The next condition that we assume, which is also not imposed by Caplin and Dean [2015], Caplin et al. [2019], or De Oliveira et al. [2017], is a sort of local strong convexity. We will assume that the cost function exhibits strong convexity, in the neighborhood of an uninformative information structure, with respect to information structures that hold fixed the unconditional distribution of signals, uniformly over the set of possible priors.

Condition 5. There exists constants $m > 0$ and $B > 0$ such that, for all priors $q \in \mathcal{P}(X)$, and all information structures that are sufficiently close to uninformative ($C(p, q; S) < B$),

$$C(p, q; S) \geq \frac{m}{2} \sum_{s \in S} \pi_s(p, q) ||q_s(p, q) - q||_X^2,$$

where q_s is the posterior given by Bayes' rule and $|| \cdot ||_X$ is an arbitrary norm on the tangent space of $\mathcal{P}(X)$.

This condition is slightly stronger than Condition 1; it it essentially an assumption of "local strong convexity" instead of merely local strict convexity. It implies that all informative information structures have a non-trivial positive cost, and that (regardless of the DMs' current beliefs) there are no informative information structures that are "almost free." This condition allows us to assert that if the flow cost $C(p, q; S)$ is converging to zero, then either the posteriors must become close to the prior ($q_s(p, q)$ close to q) or the signals must become rare ($\pi_s(p, q)$ close to zero), or some combination thereof.

The mutual-information cost function (6) satisfies each of these five conditions. However, it is not the only cost function to do so. For example, we can construct a family of such cost functions, using the family of "f-divergences" (see, e.g., Ali and Silvey [1966] or Amari and Nagaoka [2007]), defined as

$$D_f(q'||q) = \sum_{x \in X} q_x f(\frac{q'_x}{q_x}),$$

where f is any strictly convex, twice-differentiable function with $f(1) = f'(1) = 0$ and $f''(1) = 1$. The KL divergence is a member of this family, corresponding to $f(u) = u \ln u - u + 1$.

The posterior-separable cost function associated with this divergence,

$$C^f(p, q;\, S) = \sum_{s \in S} \pi_s(p, q) D_f(q_s(p, q)||q), \tag{15}$$

satisfy all five of the conditions described above. Another example satisfying our conditions, as mentioned previously, are the neighborhood cost functions of described by Hébert and Woodford [2018]. These cost functions fall into the uniformly posterior-separable family of cost functions described by Caplin et al. [2019], and (under mild regularity assumptions) all such functions satisfy our conditions. We discuss these issues in more detail in the appendix, section §A.

We are now in a position to use these conditions to characterize the cost of a small amount of information.

Theorem 1. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that $\lim_{m \to \infty} \Delta_m = 0$. Given an prior $q \in \mathcal{P}(X)$ and signal alphabet S, let $p_m \in \mathcal{P}(S)^{|X|}$ be a sequence of signal structures, and suppose that there exists a $B > 0$ such that, for all $s \in S$ and all $m \in \mathbb{N}$,*

$$\pi_s(p_m, q)||q_s(p_m, q) - q||_X^2 \leq B\Delta_m.$$

Then any cost function satisfying Conditions 1-4 is posterior-separable up to first order in Δ:

$$C(p_m, q; S) = \sum_{s \in S} \pi_s(p_m, q) D^*(q_s(p_m, q)||q) + o(\Delta_m),$$

where D^* is a divergence.

The divergence D^* is finite and convex in its first argument, and continuous in both arguments. $D^*(r||q)$ is twice-differentiable in r for r sufficiently close to q, with

$$\frac{\partial^2 D^*(r||q)}{\partial r^i \partial r^j}|_{r=q} \; = \; Diag(q)k(q)Diag(q)$$

for some continuous matrix-valued function $k(q)$. For all q, $k(q)$ is positive semi-definite and symmetric, and satisfies $z^T k(q) z = 0$ for all $z \in \mathbb{R}^{|X|}$ that are constant in the support of q. If in addition the cost function satisfies Condition 5, then there exists a constant $m_g > 0$ such that

$$k(q) - m_g(Diag(q) - qq^T) \succeq 0.$$

Proof. See the appendix, section B.2. □

Going forward, we will use the notation

$$\bar{k}(q) = Diag(q)k(q)Diag(q)$$

to describe the Hessian of the divergence $D^*(r||q)$ evaluated at $r = q$, and observe that it is positive semi-definite.

Now that we have derived our approximation result, we turn to considering the continuous time limit of the dynamic rational inattention problem introduced in section §2.

4 Rational Inattention in Continuous Time

We next study the limit of the dynamic rational inattention problem described in section §2. We will derive these limits for several specific cases, depending whether the DM's has exponential discounting ($\beta < 1$ vs. $\beta = 1$) and on the nature of the flow information cost function $C(\cdot)$. In particular, we consider two cases for $C(\cdot)$: when there is a "preference for gradual learning" and when there is a "preference for discrete learning," terms we define below. These two classes of flow cost functions will lead, respectively, to beliefs that move in small increments and beliefs that move in large increments.

Our most general convergence result proves that the discrete time dual problem converges $W(q_0, \lambda; \Delta)$ (defined in equation (11)) converges to our the continuous time dual value function, $W^+(q_t, \lambda)$, described in Definition 1. We also prove, for some value of λ^*, the convergence of the original problem $V(q_0; \Delta_n)$ to $W^+(q_t, \lambda^*)$.

Theorem 2. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that $\lim_{m\to\infty} \Delta_m = 0$, let $V(q_0; \Delta_m)$ and $W(q_0, \lambda; \Delta_m)$ denote the discrete-time value functions defined in equations (10) and (11), and suppose that the flow cost function satisfies Conditions 1-5.*

The value function of the continuous time problem of Definition 1, $W(q_0, \lambda)$, is bounded, convex, and differentiable, and there exists a sub-sequence $n \in \mathbb{N}$ such that, for all λ if $\beta < 1$ and all $\lambda \in (0, \kappa c^{-\rho})$ if $\beta = 1$,

$$\lim_{n\to\infty} W(q_0, \lambda; \Delta_n) = W(q_0, \lambda).$$

There exists a λ^* such that

$$\lim_{n\to\infty} V(q_0; \Delta_n) = W(q_0, \lambda^*).$$

Moreover, it is without loss of generality to suppose that the diffusion terms (σ_s) of the optimal policy associated with $W(q_0, \lambda)$ are zero.

Proof. See the appendix, section B.6. □

Theorem 2 demonstrates the convergence of the original and dual problems, and shows (as part of the proof) that it is without loss of generality to assume there is no diffusion component. The intuition for the latter result, which we mentioned previously, is that it is possible to synthesize a "diffusion-like" process using the jump controls.

We next present the convergence result that is specific to the $\beta = 1$ case.

Corollary 1. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that $\lim_{m\to\infty} \Delta_m = 0$, let $V(q_0; \Delta_m)$ and $W(q_0, \lambda; \Delta_m)$ denote the discrete-time value functions defined in equations (10) and (11), and suppose that the flow cost function satisfies Conditions 1-5. If $\beta = 1$, then for all $\lambda \in (0, \kappa c^{-\rho})$, $\lim_{n\to\infty} W(q_0, \lambda; \Delta_n)$ is equal to the dual value function defined in Definition 2, and the limit $\lim_{n\to\infty} V(q_0; \Delta_n)$ is equal to the value function defined in Definition 2.*

Proof. See the appendix, section B.10. □

These results demonstrate that the value functions in our discrete time problems converge to the value functions of our continuous time problem. However, the results leave unanswered two key questions: first, what are the dynamics of beliefs in the continuous time problem, and second, are these dynamics the limit of the optimal policies in the discrete time problem? To address these questions, we present conditions under which the beliefs will evolve gradually and conditions under which they will evolve in large increments.

4.1 Gradual Learning

We begin by defining what we call a "preference for gradual learning." This condition describes the relative costs of learning via jumps in beliefs vs. continuously diffusing beliefs, which are governed by the properties of the divergence D^* that defines the cost function up to first order (Theorem 1).

Definition 3. The cost function $C(p, q; S)$ exhibits a "*preference for gradual learning*" if the associated divergence D^* and its Hessian $\bar{k}(q)$ satisfy, for all $q, q' \in \mathcal{P}(X)$ with $q' \ll q$,

$$D^*(q'||q) - (q' - q)^T \left(\int_0^1 (1-s)\bar{k}(sq' + (1-s)q)ds\right)(q' - q) \geq 0. \tag{16}$$

This preference is "strict" if the inequality is strict for all $q' \neq q$, and is "strong" if, for some $\delta > 0$ and some $m > 0$,

$$D^*(q'||q) - (q'-q)^T(\int_0^1 (1-s)\bar{k}(sq' + (1-s)q)ds)(q'-q) > m||q'-q||_X^{2+\delta}. \tag{17}$$

Recall that a divergence is a Bregman divergence if it can be written, using some convex function $H : \mathcal{P}(X) \to \mathbb{R}$, as

$$D_H(q'||q) = H(q') - H(q) - (q'-q)^T \cdot H_q(q).$$

To clarify the meaning of a preference for gradual learning, consider the "chain rule" (Cover and Thomas [2012]) that characterizes the Bregman divergences (including in particular the Kullback-Leibler divergence),

$$D_H(q'||q) + \sum_{s \in S} \pi_s D_H(q_s||q') = \sum_{s \in S} \pi_s D_H(q_s||q)$$

for all probability distributions π_s such that $\sum_{s \in S} \pi_s q_s = q'$. The following lemma demonstrates that, for any divergence D^* for which this chain rule is a "less-than-or-equal" inequality, there is a preference for gradual learning.

Lemma 2. *If the divergence D^* satisfies, for all $\pi_s \in \mathcal{P}(S)$ and $q, q', \{q_s\}_{s \in S} \in \mathcal{P}(X)$ such that $\sum_{s \in S} \pi_s q_s = q'$ and $q' \ll q$,*

$$D^*(q'||q) + \sum_{s \in S} \pi_s D^*(q_s||q') \leq \sum_{s \in S} \pi_s D^*(q_s||q),$$

and is twice-differentiable in its first argument, then the cost function $C(p, q; S)$ exhibits a preference for gradual learning.

Proof. See the appendix, section B.11. □

This chain-rule inequality has a simple interpretation– it is more costly to jump directly to the beliefs $\{q_s\}$ than to have beliefs travel first to q' and then on to $\{q_s\}$. It is straightforward to see why such an assumption leads directly to gradual learning, although it is worth noting that the preference for gradual learning is a weaker condition than this chain rule inequality. This chain rule inequality could also be called "super-additivity," as the opposite of the "sub-additivity" assumption discussed by Zhong [2017].[11] In the next subsection, building on the results of Zhong [2019] and Zhong [2017], we will show in our model that that opposite inequality leads to immediate decision-making.

We begin by stating our gradual learning result for the $\beta = 1$ case. With a preference for gradual learning, we show that the value function converges to a continuous time problem with only a diffusion control (that is, without jumps). If the preference is strict, then the limiting process for beliefs must be a diffusion. If not (i.e. if D^* is a Bregman divergence, as in Zhong [2019]), there may be sequences of optimal policies in the discrete time models that converge to jump-diffusions, or even pure jump processes.

[11] This is not the exact mirror of the sub-additivity assumption, because the inequality only applies to D^*, which describes the cost of a small amount of information, as opposed being an assumption on the cost function $C(\cdot)$ as in Zhong [2017].

Theorem 3. *Under the assumptions of Theorem 2, if $\beta = 1$ and the cost function satisfies a preference for gradual learning, there exists a sub-sequence, indexed by n, and a λ^* such that*

$$\lim_{n\to\infty} W(q_0, \lambda^*; \Delta_n) = \lim_{n\to\infty} V(q_0; \Delta_n) = W^+(q_0, \lambda^*) = V(q_0),$$

where

$$V(q_0) = \sup_{\{\sigma_t \in \mathbb{R}^{|X|\times|X|}\}, \tau \in \mathbb{R}_+} E_0[\hat{u}(q_\tau) - \kappa\tau)]$$

subject to the constraints that $q_t^T \sigma_t = \vec{0}$,

$$dq_t = Diag(q_t)\sigma_t dB_t$$

and

$$\frac{1}{2} tr[\sigma_t \sigma_t^T k(q_t)] \leq \rho^{\frac{1}{\rho}} c.$$

If the cost function exhibits a strict preference for gradual learning, every convergent sub-sequence of belief processes $q_{t,n}^*$ associated with optimal policies in the discrete-time model converges in law to a diffusion.

Proof. See the appendix, section B.12. □

The sequence problem $V(q_0)$ has a straightforward interpretation– the DM controls the diffusion coefficient of her beliefs subject to a maximum rate of information acquisition defined by the information cost matrix function $k(q_t)$.

Under an additional assumption (described in the next section), a strong preference for gradual learning is not only sufficient but necessary for beliefs to follow a diffusion process in the $\beta = 1$ case. In particular, we will demonstrate that if, for all utility functions, the belief process in the continuous time limit is a diffusion, then the divergence D^* must exhibit a preference for gradual learning. However, to make this statement, we must be able to characterize the belief dynamics, which we are able to do given an additional assumption. We therefore postpone our proof of necessity to the next section.

We next turn to the case with discounting ($\beta < 1$). A remarkable result by Zhong [2019] (theorem 5 of that paper) demonstrates that, in a model very much like our continuous time limit, specialized to the case of only two states and with no linear time costs, the DM will generically choose to have no diffusion in her process, only jumps. This result can be understood in two parts. First, as discussed previously, it is without loss of generality to write the DM's belief process as a pure jump process, even when the law of the supremum over such processes is equivalent to the law of a diffusion.

The second part of the result of Zhong [2019] can be thought of as a (generic) lower bound on the magnitude of the jumps. That is, not only is it without loss of generality to consider a pure jump process, but the optimal policy of the DM is in fact a jump process with non-infinitesimal jumps. What is remarkable about this result, from the perspective of our results for the $\beta = 1$ case, is that it applies even when the cost function exhibits a preference for gradual learning. Zhong [2019] also shows, in the particular case of indifference to gradual learning (equality in equation (16)), which is the setting for most of his results, that the beliefs jump all the way to stopping points, a result we will replicate below.

To understand how our $\beta = 1$ results are connected to the results of Zhong [2019], for the $\beta < 1$ case, we suppose that the cost function exhibits a "strong" preference for gradual learning, as defined in Definition 3 and equation (17). Under this assumption, we prove an upper bound on the size of the jumps, as a function of β, and show that in the limit as $\beta \to 1$, the bound converges to zero. In other words, there is no discontinuity– with discounting and a strong preference for gradual learning, the "pure jump process" for beliefs will become increasingly like a diffusion as the discount factor converges to unity. Recall that m and δ are part of the definition of a strong preference for gradual learning, and $\bar{u}$ is the upper bound on the flow utility.

Theorem 4. *In problem defined in Definition 1, if the cost function satisfies a strong preference for gradual learning (equation (17)), then the optimal jump directions z_t^* satisfy*

$$||z_t^*||_X \leq \left(\frac{-\bar{u}\ln(\beta)}{m}\right)^{\delta^{-1}},$$

and $\lim_{\beta \to 1^-} ||z_t^*||_X = 0$. *Moreover, jumps always strictly increase the value function,*

$$W(q_{t^-} + z_t^*, \lambda) > W(q_{t^-}, \lambda).$$

Proof. See the appendix, section B.13. □

The optimal policy for the discounting case features upward (in the sense of the value function) jumps and downward drift. The jumps become shorter and more frequent as the discount factor approaches one, eventually converging to a diffusion. The fact that jumps only increase and never decrease the value function is a consequence of the exponential discounting. Exponential discounting can be thought of as a penalty for delay that is increasing in the current level of the value function. For this reason, drifting upward and jumping downward are sub-optimal, because the former causes information to be acquired at a time when the cost of delay is high, and the latter acquires information at a time when the cost of delay is high rather than waiting for the cost of delay to decrease. We should note that the result that jumps always strictly increase the value function requires only a strict preference for gradual learning, rather than a strong preference, and such a result is without loss of generality (but not necessarily required) under a weak preference for gradual learning. Our result that jumps always increase the value function is reminiscent of a result in Che and Mierendorff [2019] and Zhong [2019], which is sometimes characterized as the DM seeking out confirmatory evidence. For reasons that we discuss below in our analysis of stopping times (and are illustrated by the one-sided learning example of Zhong [2019]), we prefer not to describe the result in this fashion.

Having analyzed the case of a preference for gradual learning, we next turn to the case of a preference for discrete learning,

4.2 Discrete Learning

Building on Zhong [2019] and Zhong [2017], to provide contrast to our results on discrete learning, we provide conditions under which, in the $\beta = 1$ case, the DM jumps immediately to stopping beliefs. We define what we call a "preference for discrete learning" if the divergence D^* satisfies the opposite of the "chain rule" inequality described previously (Lemma 2).

Definition 4. The cost function $C(p, q; S)$ exhibits a "*preference for discrete learning*" if divergence D^* associated with it satisfies, for all $\pi_s \in \mathcal{P}(S)$ and $q, q', \{q_s\}_{s\in S} \in \mathcal{P}(X)$ such that $\sum_{s\in S} \pi_s q_s = q'$ and $q' \ll q$,

$$D^*(q'||q) + \sum_{s\in S} \pi_s D^*(q_s||q') \geq \sum_{s\in S} \pi_s D^*(q_s||q). \tag{18}$$

The cost function $C(p, q; S)$ exhibits a "*strict preference for discrete learning*" if this inequality is strict for all $\pi_s, q, q', \{q_s\}_{s\in S}$ with $q' \neq q$ and $q_s \neq q'$ for some $s \in S$ such that $\pi_s > 0$.

If the cost function satisfies a preference for discrete learning, it its cheaper for the DM to jump to beliefs $\{q_s\}$ rather than visit the beliefs q'. Unsurprisingly, because this holds everywhere, it leads to optimal policies that stop immediately after jumping. Before discussing this result, we note that our definitions of a preference for discrete learning and a preference for gradual learning are not symmetric, although they are linked by Lemma 2. As mentioned previously, our definition of a preference for gradual learning includes cost functions that do not satisfy the chain rule inequality everywhere. The following lemma clarifies the relationship between a preference for gradual and discrete learning, showing that the strict versions of those preferences are incompatible.

Lemma 3. *Assume that the divergence D^* associated with the cost function $C(p, q; S)$ is twice-differentiable with respect to its first argument. If a cost function $C(p, q; S)$ exhibits a strict preference for gradual learning, it does not exhibit a preference for discrete learning. If a cost function $C(p, q; S)$ exhibits a strict preference for discrete learning, it does not exhibit a preference for gradual learning.*

Proof. See the appendix, section B.14. □

Note in particular that any twice-differentiable Bregman divergence (recall equation (8)) exhibits both a preference for gradual and discrete learning. Note also that there is no reason, in general, to expect either of the chain rule inequalities to hold over the entire simplex. That is, the cases we have an analyzed are not exhaustive, and there many be cost functions that lead to large jumps in certain regions of the parameter space and small jumps or diffusions in other regions.

We now turn to characterizing the behavior of beliefs given a preference for discrete learning.

Theorem 5. *In the problem described by Corollary 1 (that is, $\beta = 1$), if the cost function satisfies a preference for discrete learning (Definition 4), then it is without loss of generality to suppose that, for all q_{t^-} in the continuation region and any z_t^* that characterizes an optimal policy , $V(q_{t^-} + z_t^*) = \hat{u}(q_{t^-} + z_t^*)$. That is, without loss of generality, all jumps enter the stopping region.*

Proof. See the appendix, section B.15. □

The statement of Theorem 5 shows that is without loss of generality to assume that the DM stops immediately after a jump in beliefs. This result would be necessary, as opposed to without loss of generality, if the cost function exhibited a strict preference for discrete learning (although we will see below that such cost functions may not exist). In contrast, in the case of indifference (equation (18) holds with equality everywhere), both Theorem 3 and Theorem 5 hold. This happens if D^* is the KL divergence, or a Bregman divergence more generally. This observation implies that the two continuous time value functions must be identical, despite one being written as controlling a diffusion process and the other (without loss of generality) a pure jump process. We revisit this observation in the next section.

For completeness, we also provide an analog to our result about the magnitude and direction of jumps in the gradual learning case (Theorem 4) for the discounting case ($\beta < 1$). This result generalizes slightly a result in appendix A.3 of Zhong [2019]. With a preference for discrete learning, as with a preference for gradual learning, jumps will increase the value function. The intuition is essentially the same as the gradual learning case, and comes from the observation that with discounting, delay is particularly costly when the value function is high. However, unlike the gradual learning case, in which jumps are of bounded size, with a preference for discrete learning jumps are always immediately following by stopping.

Theorem 6. *In the problem defined in Definition 1, if $\beta < 1$ and the cost function satisfies a preference for discrete learning (Definition 4), then for all q_{t^-} in the continuation region and any z_t^* that characterizes an optimal policy , $V(q_{t^-} + z_t^*) = \hat{u}(q_{t^-} + z_t^*)$. That is, without loss of generality, all jumps enter the stopping region. Moreover, jumps always strictly increase the value function,*

$$W(q_{t^-} + z_t^*, \lambda) > W(q_{t^-}, \lambda).$$

Proof. See the appendix, section B.16. □

Moving beyond the results of Zhong [2019], we provide the "only-if" result: if a divergence always results in large jumps and immediate stopping, then it must satisfy a preference for discrete learning. The intuition is that if it is always optimal to jump outside the continuation region, it cannot possibly be less costly, in the sense of the divergence D^*, to jump to some intermediate point. Otherwise, there would be some utility function for which such behavior is optimal.

Theorem 7. *In the problem defined in Definition 1, if there exists a cost function $C(p, q; S)$ and associated divergence D^* such that, for all strictly positive utility functions $u_{a,x}$, there exists an optimal policy such that $V(q_{t^-} + z_t^*) = \hat{u}(q_{t^-} + z_t^*)$ for all q_{t^-} in the continuation region, then the cost function satisfies a preference for discrete learning.*

Proof. See the appendix, section B.17. □

We have proven the convergence of our discrete time problems to our continuous time rational inattention problems and shown that there are at least two different kinds of belief dynamics that might arise from the model, depending on whether there is a preference for gradual or discrete learning. We next provide an example of a family of cost functions that exhibit a preference for gradual learning, and demonstrate that, subject to a regularity condition, Bregman divergences are the only cost functions that exhibit a preference for gradual learning.

4.3 Examples with a Preference for Discrete and Gradual Learning

Here, we provide an intuitive example of families of cost functions that generate preferences for gradual and discrete learning. We provide the examples in the context of posterior-separable cost functions,

$$C(p, q; S) = \sum_{s \in S} \pi_s(p, q) D^*(q_s(p, q)||q).$$

Note that our derivation of the model (Theorem 1) imposes a number of conditions on $D^*(p||q)$, including continuity and convexity in its first argument.

Cost functions exhibiting a (strict or strong) preference for gradual learning can be easily constructed from Bregman divergences. Suppose that

$$D^*(q'||q) = f(H(q') - H(q) - (q' - q)^T \cdot H_q(q)),$$

where $f(\cdot)$ is a twice-differentiable, strictly increasing, convex function with $f(0) = 0$, and $H(\cdot)$ is a strictly convex function on the simplex. It is straight-forward to observe that the Hessian of D^* evaluated at $q' = q$ is $f'(0)H_{qq}(q)$, and by convexity

$$D^*(q'||q) \geq f'(0)(H(q') - H(q) - (q' - q)^T \cdot H_q(q)),$$

implying that the divergence D^* satisfies a preference for gradual learning. This preference will be strict if $f(\cdot)$ is strictly convex, and will be strong if both $f(\cdot)$ and $H(\cdot)$ are strongly convex.

Constructing examples of cost functions with a strict preference for discrete learning has proven challenging. This class of cost functions may seem like a large class; however, as the following lemma demonstrates, subject to mild regularity conditions it only contains uniformly posterior separable cost functions (the Bregman divergence case).

Lemma 4. *Suppose the cost function $C(p, q; S)$ exhibits a preference for discrete learning and that the associated divergence D^* is continuously differentiable in both its arguments. Then D^* is a Bregman divergence and the cost function $C(p, q; S)$ is a uniformly posterior-separable cost function.*

Proof. See the appendix, section B.18. The proof builds on Banerjee et al. [2005]. □

As noted above, if D^* is a Bregman divergence, it exhibits a (non-strict) preference for discrete learning, and also a (non-strict) preference for gradual learning. Consequently, under mild regularity conditions, no cost functions exhibit a strict preference for discrete learning. In contrast, many cost functions exhibit a strict or strong preference for gradual learning, and many others fall into neither category (i.e. they have a preference for gradual learning in some parts of the parameter space and discrete learning in others). Combining this result with Theorem 7, we have demonstrated that, subject to mild regularity conditions, the jump-and-immediately-stop result of Zhong [2019] holds for all utility functions if and only if D^* is a Bregman divergence.

The results of this section naturally lead to the question of whether the different belief dynamics that occur under a preference for gradual or discrete learning lead to different predictions about the DM's behavior. We explore this question in the next two sections.

5 The Equivalence of Static and Dynamic Models

In this section, we analyze the $\beta = 1$ continuous time model under a preference for gradual learning and a preference for discrete learning. The main result of this section is that, both with a preference for gradual learning and a preference for discrete learning (under an integrability assumption in the case of a preference for gradual learning, and assuming continuous differentiability in the case of a preference for discrete learning), the value function with $\beta = 1$ is equivalent to a static rational inattention problem with a uniformly posterior-separable cost function. Moreover, any twice-differentiable uniformly posterior-separable cost function can be justified through either of these routes.

This result has several implications. First, it demonstrates that both jump and diffusion-based models are tractable and that the value functions can be characterized without directly solving the associated partial differential equation. Second, it provides a micro-foundation for the uniformly posterior-separable cost functions that have been emphasized in the literature (see, e.g., Caplin et al. [2019]). Third, it proves that the two approaches are completely equivalent in terms of the predicted joint distribution of states $x \in X$ and actions $a \in A$. That is, any joint distribution of (x, a) that could be observed under discrete learning could be observed under gradual learning and vice versa.

One this last point, however, we do not mean to imply that the diffusion and jump processes are equivalent. Both of them endogenously will result in the same joint distribution of actions and states, but will have different predictions about the joint distribution of actions, states, and stopping times. As a consequence, considering stopping times can help differentiate the two models, and we take some steps toward this in the next section.

Finally, before presenting our results, we should emphasize that in the case of gradual learning, our results depend on additional assumptions beyond those implied by our derivation of the continuous time model. Our assumption for the gradual learning case is an integrability assumption, and will not hold generically. In contrast, in the discrete learning case, we make the mild additional assumption of continuous differentiability. Consequently, equivalence with static models holds for essentially all cost functions with a preference for discrete learning but only some cost functions for a preference for gradual learning, and essentially all cost functions with a preference for discrete learning generate the same joint distribution of actions and states as some cost function with a preference for gradual learning, but the reverse is not true.

5.1 Gradual Learning

To prove our equivalence result, we restrict our attention to information-cost matrix functions that are "integrable," in the sense described by the following assumption.

Assumption 1. *There exists a twice-differentiable function* $H : \mathbb{R}_{+}^{|X|} \to \mathbb{R}$ *such that, for all* q *in the interior of the simplex,*

$$Diag(q)k(q)Diag(q) = H_{qq}(q), \tag{19}$$

where $H_{qq}(q)$ *denotes the Hessian of* H *evaluated at* q.

This class includes a number of information-cost matrix functions of interest: for example, it includes the case in which $k(q_t)$ is the inverse Fisher information matrix, which we will show corresponds to the standard rational inattention model, and the case in which $k(q_t)$ is the "neighborhood-based" function that we introduce in Hébert and Woodford [2018]. It is, however, a restrictive assumption. It rules out, for example, constant $k(q)$ (a hypothetical H would have asymmetric third-derivative cross-partials). We shall refer to the function H as the "entropy function," for reasons that will become clear below. Note that $H(q)$ is convex, by the positive semi-definiteness of $k(q)$, and homogenous of degree one ($q^T \cdot H_{qq}(q) = \iota^T k(q) Diag(q)^{-1} = \vec{0}$).

For every convex function H, there is a Bregman divergence,

$$D_H(q_s||q) = H(q_s) - H(q) - (q_s - q)^T H_q(q),$$

and a corresponding uniformly posterior-separable cost function.

The problem we are analyzing is the result of Theorem 3, the problem with $\beta = 1$ and a diffusion process for beliefs. To analyze our continuous time problem, we begin by proving that the information constraint,

$$\frac{1}{2} tr[\sigma_t \sigma_t^T k(q_t)] \leq \rho^{\frac{1}{\rho}} c,$$

binds. Because the constraint binds, we can substitute the constraint into the HJB equation for the problem and obtain the following result:

Lemma 5. *At every point at which the value function* $V(q_t)$ *is twice-differentiable and the DM chooses not to stop, for all* σ *such that* $q^T \sigma = \vec{0}$,

$$tr[\sigma^T \{Diag(q_t) V_{qq}(q_t) Diag(q_t) - \theta k(q_t)\} \sigma] \leq 0, \tag{20}$$

where $\theta = \frac{\kappa}{\rho^{\frac{1}{\rho}} c}$, *with equality under the optimal policy.*

Proof. See the appendix, Section B.19. □

The parameter θ, introduced in the lemma, describes the race between information acquisition and time in this model. The larger the penalty for delay, and the tighter the information constraint, the larger the parameter θ. We now describe our equivalence result, and then outline the key step of its proof, which relies on this lemma.

Theorem 8. *Under Assumption 1, the value function that solves the continuous time problem with* $\beta = 1$ *(Definition 2) and with a preference for gradual learning is*

$$V(q_0) = \max_{\pi \in \mathcal{P}(A), \{q_a \in \mathcal{P}(X)\}_{a \in A}} \sum_{a \in A} \pi(a)(u_a^T \cdot q_a) - \theta \sum_{a \in A} \pi(a) D_H(q_a||q_0),$$

subject to the constraint that $\sum_{a\in A}\pi(a)q_a = q_0$*, where* D_H *is the Bregman divergence associated with the entropy function* H *that is defined by Assumption 1.*

There exist maximizers π^* *and* q_a^* *such that* π^* *is the unconditional probability, in the continuous time problem, of choosing a particular action, and* q_a^**, for all* a *such that* $\pi^*(a) > 0$*, is the unique belief the DM will hold when stopping and choosing that action.*

Proof. See the appendix, Section section B.20. □

The continuous time sequential evidence accumulation problem is equivalent to a static rational inattention problem, with a particular uniformly posterior-separable cost function,

$$C(p, q_0; S) \quad = \quad \sum_{s\in S}\pi_s(p, q_0)D_H(q_s(p, q_0)||q_0), \tag{21}$$

and with the signal space S identified with the set of possible actions A.

The mutual information cost function (6) proposed by Sims is one such cost function. In this case, the entropy function H is the negative of Shannon's entropy (4), the corresponding information-cost matrix function is the inverse Fisher information matrix $k(q) = Diag(q) - qq^T$, the Bregman divergence is the Kullback-Leibler divergence (5), and the information measure defined by (21) is mutual information (6). Thus Theorem 8 provides a foundation for the standard static rational inattention model, and hence for the same predictions regarding stochastic choice as are obtained by Matêjka et al. [2015]. On the other hand, Theorem 8 also implies that other cost functions can also be justified. Indeed, any (twice-differentiable) uniformly posterior-separable cost function (21) can be given such a justification, by choosing the information cost matrix function defined by equation (19).

However, not all information cost matrix functions are reasonable. We interpret the information cost matrix function $k(q)$ as a description of how hard it is to distinguish any pair of states. In many economic applications, there is a natural ordering or structure of the states, and we would like the information cost matrix function and the associated entropy function and Bregman divergence to reflect this structure. In Hébert and Woodford [2018], we propose such a cost function.

We next outline the key step of our proof. Our proof strategy is best described as "guess and verify," in that we start with the static value function described in Theorem 8 and then show that it is the value function of the continuous time model described in Theorem 3. The key step of the proof is to show that the static value function satisfies (20) in Lemma 5.[12] For expositional purposes, we will assume that the optimal policies of the static model, $\pi^*(a; q_0)$ and $q_a^*(q_0)$, are differentiable with respect to q_0 and strictly interior (we do not require these assumptions in the proof).

We begin by examining the first-order conditions with respect to q_a, and applying the envelope theorem. Let $\kappa(q_0)$ denote the vector of multipliers on the constraint that $\sum_{a\in A}\pi(a)q_a = q_0$. We

[12] Technical footnote: the "anywhere the value function is twice differentiable" caveat of Lemma 5 is relevant for our problem. The PDE described by equation (20) is "degenerate elliptic" and hence will not in general have a classical solution. Indeed, we do not prove that our static value function is twice differentiable everywhere, and suspect it is not at points where the "consideration set" (the set of actions with $\pi(a) > 0$, Caplin et al. [2018]) changes. In our proof, we establish that the static problem value function is convex and continuously differentiable, which is sufficient to invoke a generalized version of Ito's lemma for convex functions to verify that the static value function is the solution to the continuous time problem.

have the first-order condition (FOC) and envelope theorem condition (ET),

$$\text{FOC: } u_a - \kappa(q_0) - \theta H_q(q_a^*(q_0)) + \theta H_q(q_0) = 0,\ \forall a \in A, \tag{22}$$

$$\text{ET: } V_q(q_0) = \kappa(q_0) + \sum_{a\in A} \pi^*(a; q_0)(q_a^*(q_0) - q_0)^T \cdot H_{qq}(q_0) = \kappa(q_0).$$

Now consider a perturbation $q_0 \to q_0 + \epsilon z$, for some tangent vector z. Combining the FOC and ET, and then differentiating with respect to ϵ and evaluating at $\epsilon = 0$,

$$V_{qq}(q_0) \cdot z = \theta H_{qq}(q_0) \cdot z - \theta H_{qq}(q_a^*(q_0)) \cdot \frac{dq_a^*(q_0 + \epsilon z)}{d\epsilon}|_{\epsilon=0},\ \forall a \in A. \tag{23}$$

Observe that, due to the constraint, $\sum_{a\in A} \frac{d(\pi^*(a;q_0+\epsilon z)q_a^*(q_0+\epsilon z))}{d\epsilon}|_{\epsilon=0} = z$. Multiplying both sides of equation (23) by $\frac{d(\pi^*(a;q_0+\epsilon z)q_a^*(q_0+\epsilon z)^T)}{d\epsilon}|_{\epsilon=0}$, and then taking sums,

$$\begin{aligned} z^T \cdot V_{qq}(q_0) \cdot z = {} & \theta z^T \cdot H_{qq}(q_0) \cdot z \\ & - \theta \sum_{a\in A} \pi^*(a; q_0)\left(\frac{dq_a^*(q_0 + \epsilon z)}{d\epsilon}|_{\epsilon=0}\right)^T \cdot H_{qq}(q_a^*(q_0)) \cdot \frac{dq_a^*(q_0 + \epsilon z)}{d\epsilon}|_{\epsilon=0} \\ & - \theta \sum_{a\in A} \frac{d\pi^*(a; q_0 + \epsilon z)}{d\epsilon}|_{\epsilon=0}\ q_a^*(q_0)^T \cdot H_{qq}(q_a^*(q_0)) \cdot \frac{dq_a^*(q_0 + \epsilon z)}{d\epsilon}|_{\epsilon=0}. \end{aligned}$$

By Assumption 1, $q^T \cdot H_{qq}(q) = 0$, and hence the last line in this expression is zero. By the convexity of H, the summation on the second line is positive. Therefore, by Assumption 1,

$$z^T \cdot V_{qq}(q_0) \cdot z \leq z^T \cdot Diag(q_0)^{-1} k(q_0) Diag(q_0)^{-1} \cdot z,$$

establishing that (20) holds. To show that there is a direction z^* in which (20) holds with equality, it is sufficient to show that $\frac{dq_a^*(q_0+\epsilon z^*)}{d\epsilon}|_{\epsilon=0} = 0$ for all $a \in A$. In any direction z spanned by the initial $q_a^*(q_0) - q_0$, it is not optimal for the DM to change the $q_a^*(q_0)$, only the probabilities $\pi^*(a, q_0)$ (this property, "Locally Invariant Posteriors," was shown by Caplin et al. [2019]). Thus, any of these directions can serve as z^*.[13]

We conclude that all continuous time models with gradual learning that also satisfy our integrability condition are equivalent to a static model with a uniformly posterior-separable cost function, and that any such static model can be justified from some model with gradual learning. We next show that the same set of static models can be justified from a model with discrete learning. Before proceeding, however, we observe that this result allows us to demonstrate that a preference for gradual learning is necessary for beliefs to always result in a diffusion process, provided that Assumption 1 holds.

Corollary 2. *In the problem defined in Definition 2 (the $\beta = 1$ case), if a cost function $C(p, q; S)$, associated divergence D^*, and information cost matrix function $k(q)$ satisfying Assumption 1 are*

[13]That any such direction can serve as z^* indicates that there are (usually) many optimal policies in the continuous time problem that achieve the same value function. Intuitively, at each point, if the DM does not learn in some particular direction, she could always learn in that direction in the next instant.

such that, for all strictly positive utility functions $u_{a,x}$, there exists an optimal policy such that $\bar{\psi}_s = 0$ everywhere in the continuation region (meaning beliefs follow a diffusion process), then the cost function satisfies a preference for gradual learning.

Proof. See the appendix, section B.21. □

5.2 Discrete Learning

The result with a preference for discrete is an immediate corollary of Lemma 4 and the preceding Theorem 8 (the result with gradual learning).

Corollary 3. *Assume the cost function $C(p, q; S)$ exhibits a preference for discrete learning, and the associated divergence D^* is continuously differentiable. Then D^* is a Bregman divergence and value function that solves the continuous time problem with $\beta = 1$ (Definition 2) is identical to the static rational inattention problem described in Theorem 8 with D^* in the place of D_H.*

Proof. Immediate from Lemma 4 and Theorem 8. □

The requirement that D^* be continuously differentiable is likely unnecessary for the result. Our derivation of D^* requires that it be convex in its first argument and continuous in both arguments. Applying "mollification" techniques for convex functions (see for example Banerjee et al. [2005]) would likely allow us to extend the result to all D^* functions satisfying the conditions of Theorem 1.

Given any uniformly posterior-separable cost function in a static rational inattention model, by setting D^* equal to the Bregman divergence associated with that cost function, we can justify that static model as the result of a dynamic model with a preference for discrete learning. We therefore conclude that models with a preference for gradual learning satisfying our integrability condition and models with a preference for discrete learning are indistinguishable from the perspective of their predictions about the joint distribution of states and actions.[14] In the next section, we begin to explore how information about stopping times can be used to distinguish the models.

6 Implications for Response Times

Because our model is dynamic, the observable behavior of the DM includes not only the joint distribution of actions and states, but also information about the length of time taken to decide, including how this may vary depending on the action and the state. The psychophysics literature gives considerable emphasis to facts about response times as a window on the nature of decision processes (e.g., Ratcliff and Rouder [1998]). In economic contexts as well, response times provide important information that can be used to discriminate between models, even when response times themselves are not what the economic analyst cares about. For example, Clithero [2018] and Alos-Ferrer et al. [2018] argue that preferences can more accurately be recovered from stochastic choice data when data on response times are used alongside observed choice frequencies.

[14] In situations in which the static rational inattention problem does not itself have a unique solution, we have not ruled out the possibility that the models with discrete and gradual learning will make different predictions. However, we have no reason to believe this is the case.

Here we propose that data on response times can be used to discriminate between alternative information-cost specifications. We will show that cost functions that are equivalent in the sense of implying the same value function nevertheless make different predictions about the stopping time conditional on taking a particular action. Consequently, data on stopping times can inform us about whether there is a preference for gradual learning or for learning through discrete jumps. Interestingly, it is possible to distinguish between these two hypotheses even when (as in the problems considered here) actions are taken only infrequently.[15]

We illustrate how response time data can discriminate between the two hypotheses using a simple example. In this example, we consider the no-discounting limit, $\beta \to 1$. We focus on this limit for several reasons. First, with regards to decision-making experiments, the cumulative discount factor over the length of time required to make a decision (often seconds or minutes) should be small. Second, it is easier to characterize stopping times when beliefs are a diffusion process than when they are a sequence of small jumps. Third, as noted by Fudenberg et al. [2018], behavior in the $\beta < 1$ case is not invariant to transformations of the utility function. Fourth, and most significantly, our results from the previous section demonstrate that in this case the gradual-learning and jump-learning approaches are equivalent in terms of their predictions for the joint distribution of actions and states. We consider the limit $\beta \to 1$, rather than simply assuming no discounting, as this allows us to choose, in the case of a preference for discrete learning, between the multiple optimal policies for the $\beta = 1$ case.

We also restrict attention to the case of only two states. This allows us to ignore the integrability condition in the gradual learning case, as it is always satisfied. We can then compare cost functions with a preference for gradual and discrete learning that generate the exact same predictions for the joint distribution of actions and signals. Second, with $\beta = 1$, two states, and a preference for gradual learning, the optimal policies that generate the stochastic process for beliefs are uniquely determined by the results of Theorem 3. Similarly, in the limit as β approaches 1 from below, with two states and a preference for discrete learning, the optimal policies that generate the stochastic process for beliefs are uniquely determined by the results of Theorem 6.[16]

The difference that we illustrate concerns the way in which changes in the payoffs associated with particular actions is predicted to affect the stopping time dynamics. We will show that a preference for gradual learning implies an invariance result, in which a shift in the level of the utility for all actions in a given state does not change the joint distribution of actions, states, and stopping times. In contrast, we will show that such a shift can radically alter the distribution of stopping times under a preference for discrete learning.

Label the states $X = \{G, B\}$, let the two actions available be $A = \{L, R\}$, and suppose that the optimal action in state G is L, while the optimal action in state B is R. We assume that these actions are strictly optimal, meaning that $u(L, G) > u(R, G)$ and $u(R, B) > u(L, G)$ (without these assumptions, the DM should immediately stop and not gather any information). We also assume,

[15] It would obviously be easier to tell whether beliefs evolve continuously or in discrete jumps in a case where the DM is required to continuously adjust some response variable that can provide an indicator of her current state of belief.

[16] In the $\beta = 1$ case, with a preference for discrete learning, there are multiple optimal policies. However, only one of these is the limit of the optimal policies from the $\beta < 1$ case.

without loss of generality, that taking the optimal action in state G generates a weakly higher payoff than taking the optimal action in state B, $u(L,G) \geq u(R,B)$.

With two states, we can interpret the beliefs q as a scalar, and we adopt the convention that q is the probability of state G. We compare two different situations, one in which the value function $W^{+}(q,\lambda)$ is "u-shaped" and one in which the value function $W^{+}(q,\lambda)$ is upward-sloping. The "u-shaped" value function occurs when $u(R,B) > u(R,G)$; that is, when taking the correct action in state B generates a higher payoff than taking the wrong action in state G. The downward-sloping value function occurs when $u(R,B) \leq u(R,G)$. In the case, the state G is simply better for the DM than the state B, regardless of the action the DM takes. For both of these models, there will be unique beliefs $0 < q_R < q_L < 1$ associated with taking actions R and L, respectively.

The objects we will study are the cumulative probabilities of stopping by time t and taking action L or R, $F_L(t)$ and $F_R(t)$. We also study the conditional (on states G and B) versions of these processes, $F_{L,G}(t), F_{L,B}(t), F_{R,G}(t), F_{R,B}(t)$. We construct two different numerical examples, described in terms of the utility function.

Case 1. $u(L,G) = u(R,B) = 3$, $u(L,B) = u(R,G) = 2$

Case 2. $u(L,G) = 3, u(R,B) = u(R,G) = 2, u(L,B) = 1$

Note that in both cases, utility increases by one when making the correct decision relative to the wrong decision ($u(L,G) - u(R,G) = u(R,B) - u(L,B) = 1$). What differentiates the two cases is that in case 2 utilities are lower by one in state B. This is convenient because it implies that the stopping beliefs q_R and q_L are identical in the two cases. This result in fact applies regardless of the number of states and actions, as the following lemma describes.

Lemma 6. *In the problem described by Definition 2 (that is, $\beta = 1$), if $u_{a,x}(\epsilon) = u_{a,x} + \epsilon v_x$ for some $v \in \mathbb{R}^{|X|}$, then the set of joint distributions of stopping times, actions, and states under optimal policies is identical for all $\epsilon \in \mathbb{R}$.*

Proof. This follows immediately from the invariance of the value function to level shifts in the utility function. □

This conclusion depends only on the assumption that $\beta = 1$. It also provides a way of testing if $\beta = 1$ or not– does shifting all rewards up or down change the observed distribution of stopping times, actions, and states? Note also that this result holds regardless of whether the cost function exhibits a preference for gradual or discrete learning. However, the import of the result is different in the two cases.

With a strict preference for gradual learning, in the two-state case, there is a unique optimal policy involving a diffusion, which generates a unique joint distribution of stopping times, actions, and states, and this unique distribution will occur with both the case 1 and case 2 utility functions. Moreover, this unique optimal policy is the limit of the optimal policies in the $\beta < 1$ case. This result, for the case of a strict preference for gradual learning, does depend on the assumption that $\beta = 1$. That is, with $\beta < 1$, different policies will be optimal under case 1 and case 2 utility functions, although they will remain qualitatively similar and converge to an identical limit as β approaches one.

In contrast, with a preference for discrete learning, there are many optimal policies, all of which lead to the same joint distribution of actions and states, but which vary in terms of their joint distribution of actions, states, and stopping times. Moreover, in the two state case, only one policy from this set is the limit of the optimal policies from the $\beta < 1$ case (which are themselves unique), and which policy corresponds to this limit depends on whether the utility function is from case 1 or case 2.

Putting together Theorem 8, Corollary 3, and Lemma 6, we proceed by choosing a cost function that satisfies a preference for discrete learning, and construct the cost function with a strict preference for gradual learning that makes equivalent predictions about the joint distributions of actions and states. We then analyze the stopping time distributions $F_L(t)$ and $F_R(t)$ with gradual learning and for both cases with a preference for discrete learning.

Because it is a standard in the literature, we choose mutual information as our cost function that satisfies a preference for discrete learning. The associated divergence D^* is the Kullback-Leibler divergence, which is a Bregman divergence. To construct another cost function with a strict preference for gradual learning that makes identical predictions, we use a posterior-separable cost function with

$$D^*(p||q) = f(D_{KL}(p||q)), \tag{24}$$

where f is any strictly-increasing, strictly convex function with $f(0) = 0$ and $f'(0) = 1$ (all such functions will generate the same predicted behavior).

To complete the model specification, we set $\kappa = \rho^{\frac{1}{\rho}} c = 1$, and assume a prior $q_0 = \frac{3}{5}$. Using Theorem 8 and Corollary 3, it follows that the stopping boundaries q_L and q_R solve

$$\begin{aligned}\max_{\pi_L \in [0,1], q_L \in [0,1], q_R \in [0,1]} \; & \pi_L q_L u(L, G) + \pi_L (1 - q_L) u(L, B) + \\ & (1 - \pi_L) q_R u(R, G) + (1 - \pi_L)(1 - q_R) u(R, B) - \\ & \pi_L D_{KL}(q_L||q_0) - (1 - \pi_L) D_{KL}(q_R||q_0),\end{aligned}$$

subject to $\pi_L q_L + (1 - \pi_L) q_R = \frac{3}{5}$. The unique solution to this problem, for both case 1 and case 2, is $q_L = \frac{e}{1+e} \approx 0.73$ and $q_R = \frac{1}{1+e} \approx 0.27$. As mentioned previously, these stopping boundaries apply regardless of whether the cost function exhibits a preference for discrete learning or a strict preference for gradual learning.

We begin by discussing the case of a strict preference for gradual learning. This case is in one sense simple. Beliefs follow a diffusion process, eventually hitting one of two boundaries. Consequently, the cumulative distributions $F_L(t)$ and $F_R(t)$ are the first-passage times of a diffusion process. The diffusion coefficient of this process is determined by the information cost matrix function $k(q_t)$ (see Theorem 3). The resulting CDFs $F_L(t)$ and $F_R(t)$ will generally have "sigmoid" shapes. What makes this case complicated is that the diffusion coefficient is usually not constant, and as a result we cannot rely on analytical results about normal distributions.

Fortunately, in our example, the resulting diffusion process is well-studied. When D^* is defined as in equation (24), the associated Hessian matrix $\bar{k}$ is the Fisher information matrix. As a result, the dynamics of beliefs follow a diffusion that is studied in the mathematical genetics literature as the "Fisher-Wright" model.[17] The lemma below summarizes these results.

Lemma 7. *In the problem described by Corollary 1 (that is, $\beta = 1$), if the divergence D^* is defined by equation (24), the process for beliefs in the two-state model is a diffusion,*

$$dq_t = (2\rho^{\frac{1}{\rho}} c q_t (1 - q_t))^{\frac{1}{2}} dB_t.$$

Conditional on the true state being G, the process is

$$dq_t = 2\rho^{\frac{1}{\rho}} c(1 - q_t)dt + (2\rho^{\frac{1}{\rho}} c q_t (1 - q_t))^{\frac{1}{2}} dB_t,$$

and conditional on the true state being B the process is

$$dq_t = -2\rho^{\frac{1}{\rho}} c q_t dt + (2\rho^{\frac{1}{\rho}} c q_t (1 - q_t))^{\frac{1}{2}} dB_t.$$

Proof. See the appendix, section B.22. □

Armed with this diffusion process, we can solve for the functions $F_L(t)$ and $F_R(t)$. Let $\phi^L(q, t, \tau)$ be the probability of hitting q_L before time τ given $q_t = q$, and observe that $F_L(t) = \phi^L(q_0, 0, t)$. By the Markov property, $\phi^L(q, t, \tau) = \phi^L(q, 0, \tau - t)$, which we abbreviate as $\phi^L(q, s)$.

By the usual dynamic programming arguments, we have

$$\phi^L_s(q, s) = \rho^{\frac{1}{\rho}} c q (1 - q) \phi^L_{qq}(q, s),$$

with the boundary conditions $\phi^L(q_L, s) = 1$ for all $s \geq 0$, $\phi(q_R, s) = 0$ for all $s \geq 0$, and $\phi(q, 0) = 0$ for all $q \in (q_L, q_R)$. The same partial differential equation, but with different boundary conditions, characterizes $F_R(t)$. The conditional (on G or B) partial differential equations involve a first-order term for the drift, but are otherwise identical.

We numerically solve this PDE, and show in figure 1 that (intuitively) both $F_L(t)$ and $F_R(t)$ have the expected sigmoid shapes. Note, as discussed above, that these stopping time distributions apply in both case 1 and case 2. We show conditional (on G or B) CDFs in figure 2.

[17]The connection between these models, as indicated by R. A. Fisher's name appearing in both the Fisher-Wright model and the Fisher information matrix, is not a coincidence. One minor difference between our application and the Fisher-Wright model is the location of the boundaries. Morris and Strack [2019] study this process as a special case of our framework.

Figure 1: Stopping Times, Strict Preference for Gradual Learning

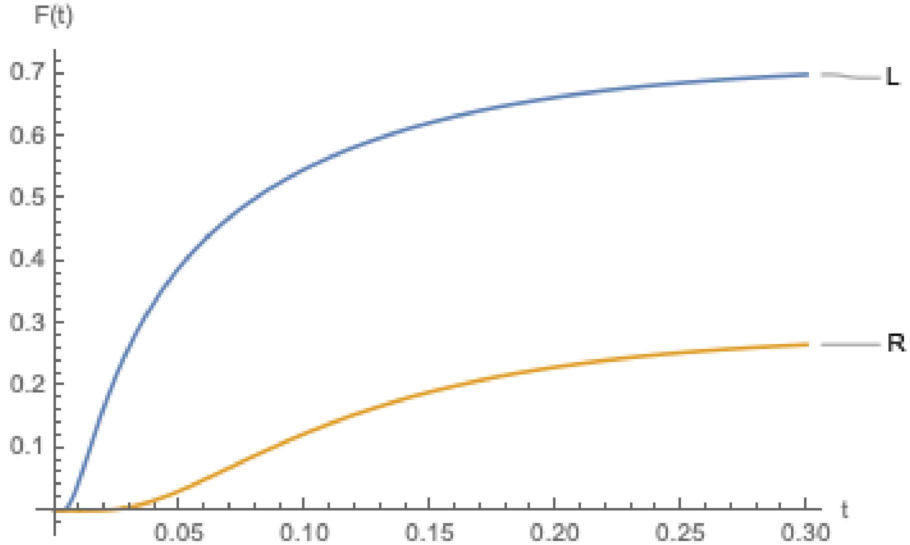
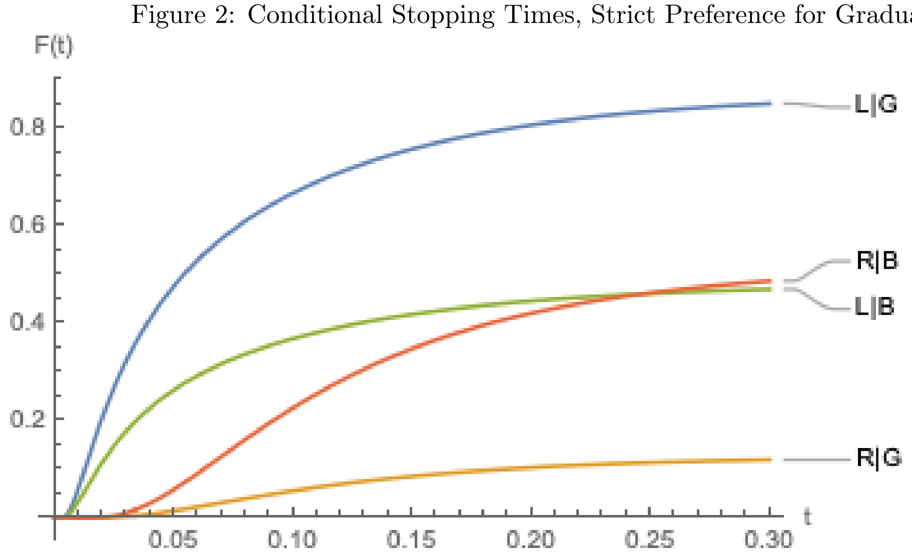


Figure 2: Conditional Stopping Times, Strict Preference for Gradual Learning



We now turn to the case of discrete learning. Consider first the symmetric case, case 1. In this case, because the value function is symmetric, it is minimized at $q = \frac{1}{2}$. It follows by our result that the value function always jumps up and drifts down (or the equivalent result in Zhong [2019]) that while $q > \frac{1}{2}$, the intensity of jumping to q_L is

$$\bar{\psi}_L(q) = \frac{\rho^{\frac{1}{\rho}} c}{D_{KL}(q_L || q)},$$

and the intensity of jumping to q_R is zero. After this point, this hazard rates will both be equal to half this value. Define $\tau_{\frac{1}{2}}$ as the time at which q_t reaches $\frac{1}{2}$. Because beliefs are martingales, the likelihood of this occurring (that is, of failing to jump before reaching $q = \frac{1}{2}$) is $\frac{q_L - q_0}{q_L - \frac{1}{2}}$. For all $t > \tau_{\frac{1}{2}}$,

$$F_L(t) = (1 - \frac{q_L - q_0}{q_L - \frac{1}{2}}) + \frac{1}{2}(\frac{q_L - q_0}{q_L - \frac{1}{2}})(1 - \exp(-\bar{\psi}_L(\frac{1}{2})(t - \tau_{\frac{1}{2}}),$$
$$F_R(t) = \frac{1}{2}(\frac{q_L - q_0}{q_L - \frac{1}{2}})(1 - \exp(-\bar{\psi}_L(\frac{1}{2})(t - \tau_{\frac{1}{2}}).$$

Prior to time $\tau_{\frac{1}{2}}$, conditional on not stopping, the beliefs $q(t)$ drift towards $q = \frac{1}{2}$,

$$q'(t) = \bar{\psi}_L(q(t))(q(t) - q_L),$$

and the probability of stopping accumulates at a hazard rate based on $\bar{\psi}_L(q(t))$,

$$F_L'(t) = (1 - F_L(t))\bar{\psi}_L(q(t)).$$

Solving this system of differential equations with initial conditions $F_L(0) = 0$ and $q(0) = q_0$, and finding the unique time value where $q(\tau_{\frac{1}{2}}) = \frac{1}{2}$ and $F_L(\tau_{\frac{1}{2}}) = \frac{q_0 - \frac{1}{2}}{q_L - \frac{1}{2}}$, which occur simultaneously, characterizes the system.

Our numerical solution is shown below in figure 3, and the conditional stopping times are shown in figure 4. Note that, aside from a few kinks and the like, the stopping time distribution is almost identical to the case of a preference for gradual learning.

Figure 3: Stopping Times, Preference for Discrete Learning, Case 1

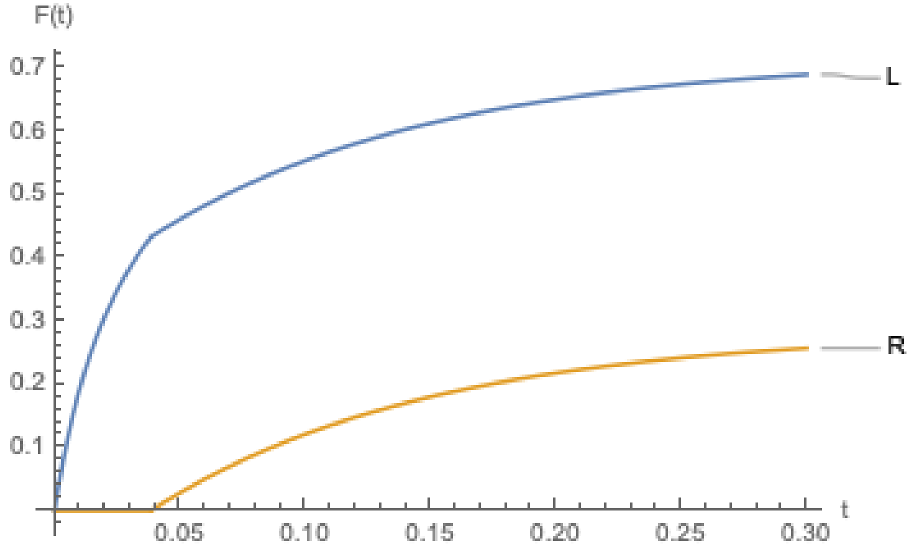
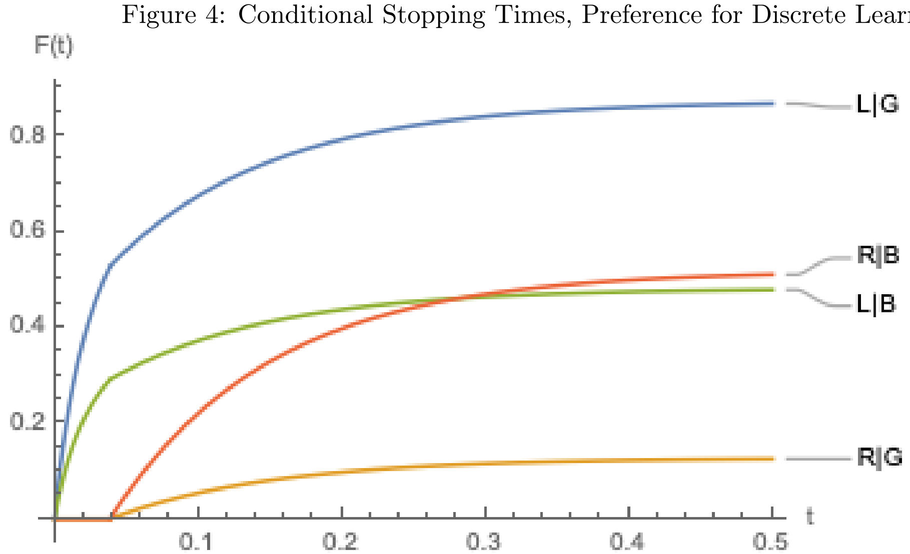


Figure 4: Conditional Stopping Times, Preference for Discrete Learning, Case 1



Now consider the asymmetric case, case 2. In this case, the value function is minimized at q_R, due to smooth pasting. That is, $\frac{d}{dq}\hat{u}(q)|_{q=q_R} = 0$ because $u(R,B) = u(R,G)$, and by the convexity of the value function[18], it follows that the value function is increasing on its domain $q \in [q_R, q_L]$. It immediately follows by Theorem 6 that, in the $\beta < 1$ case, the DM will always attempt to jump to q_L and drift towards q_R, a result Zhong [2019] calls one-sided learning.

The system that characterizes beliefs is identical to the $t < \tau_{\frac{1}{2}}$ domain of the previous case, stopping at $q = q_R$ and $F_L(\tau_R) = \frac{q_0 - q_R}{q_L - q_R}$ instead. After τ_R, the DM immediately decides to stop and choose action R. As a result, the cumulative distributions of stopping times look quite a bit different than the corresponding functions from case 1. We show these results below, in figure 5, and the conditional results in figure 6.

We are not aware of any direct experimental evidence that manipulates the level of payoffs while preserving the incentive for correct choices, though we find the predicted abrupt effect of a small change in payoffs implausible.[19] Note that the abrupt change in behavior under a preference for discrete learning depends on use of the $\beta \to 1$ limit as an optimal policy selection device. In a model with $\beta < 1$, the DM's behavior is not invariant to transformations of the utility function. The use of $\beta \to 1$ as a selection device yields predicted behavior even in the limit of no discounting that continues not to satisfy the invariance property.[20]

[18] This follows from appendix Lemma 13, which shows convexity for the discrete value functions, and Theorem 2.

[19] An experiment to test this prediction ought to be feasible.

[20] In our view, this issue should cast doubt on the usefulness of theories of behavior in the undiscounted case that are premised on the importance of continuity with behavior in the discounted case.

Figure 5: Stopping Times, Preference for Discrete Learning, Case 2

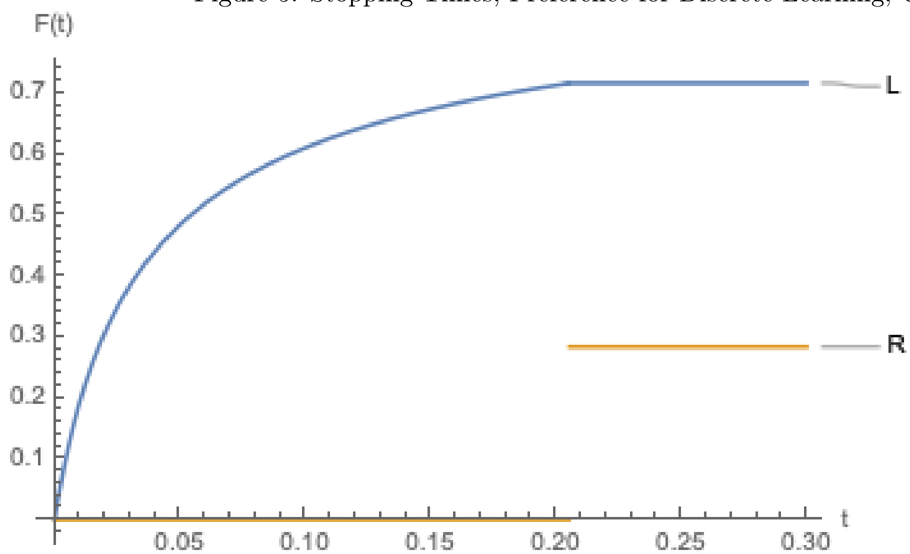
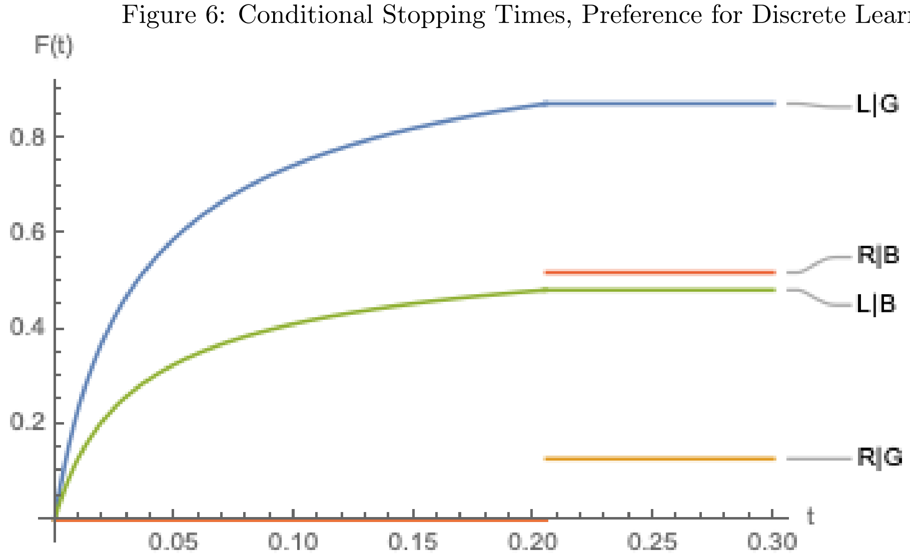


Figure 6: Conditional Stopping Times, Preference for Discrete Learning, Case 2



7 Discussion and Conclusion

We have derived a continuous-time rational-inattention model as the limit of a discrete-time sequential evidence accumulation problem. In the limit of a very large number of successive signals, each of which is only minimally informative, only the local properties of the flow cost function matter. Using these properties, we have demonstrated (without discounting) cases in which beliefs converge to either a pure diffusion process or a pure jump process. With discounting, we have described a more general limiting problem, and shown that it may result in "diffusion-like" pure jump processes. Furthermore, in the case without discounting, we have demonstrated that the resulting behavior (under an additional integrability assumption in the gradual learning case) is equivalent to the behavior predicted by a static rational inattention model. This equivalence provides a relatively simple way of generating predictions about the effects of variation in the opportunity cost of time for the probability of choosing different actions, without having to numerically solve a complex dynamic model.

We have left unresolved the question of whether it is appropriate to use the discounting or no-discounting models, and whether a preference for gradual learning is a reasonable way to characterize the evolution of beliefs. It is worth noting that in many decision-making contexts, the time it takes to make a decision (seconds or minutes) is likely to call for a cumulative discount factor (β^{τ}) very close to one. As a result, the limiting case of $\beta = 1$ may provide a useful and tractable approximation. We also note that when our key divergence D^* is a Bregman divergence, both our diffusion and jump results apply. However, this does not imply that everything is the same about the two problems– in particular, the joint distribution of decision times and actions taken may differ between the two models. This joint distribution is the subject of a great deal of existing research (e.g. Fudenberg et al. [2018]) and results in this literature might help determine whether a preference for gradual or discrete learning provides a better description of decision making.

The continuous time limit we have derived in this paper can serve as a foundation for analytically tractable models of rational inattention. One strength of our approach is its generality– we have imposed relatively minimal assumptions on the nature of information costs, and yet derived models that are specific and tractable enough to be tested in data. Exploring the properties of these models, and in particular whether jump processes or diffusions better characterize beliefs, and what cost functions should be employed, is the next step in this research agenda.

References

Syed Mumtaz Ali and Samuel D Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 131–142, 1966.

Carlos Alos-Ferrer, Ernst Fehr, and Nick Netzer. Time will tell: Recovering preferences when choices are noisy. *Unpublished manuscript*, October 2018.

Shun-ichi Amari and Hiroshi Nagaoka. *Methods of information geometry*, volume 191. American Mathematical Soc., 2007.

Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer. Information geometry and sufficient statistics. *Probability Theory and Related Fields*, pages 1–38, 2014.

Arindam Banerjee, Xin Guo, and Hui Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Transactions on Information Theory*, 51(7):2664–2669, 2005.

David Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, 24 (2):265–272, 1953.

Jerome R Busemeyer and James T Townsend. Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, 100:432–459, 1993.

Andrew Caplin and Mark Dean. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7):2183–2203, 2015.

Andrew Caplin, Mark Dean, and John Leahy. Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*, 86(3):1061–1094, 2018.

Andrew Caplin, Mark Dean, and John Leahy. Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. *Unpublished manuscript*, February 2019.

Yeon-Koo Che and Konrad Mierendorff. Optimal dynamic allocation of attention. *American Economic Review*, 109(8):2993–3029, 2019.

Nikolai Nikolaevich Chentsov. *Statistical decision rules and optimal inference*. Number 53. American Mathematical Soc., 1982.

John A Clithero. Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, 148:344–375, 2018.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

Henrique De Oliveira, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. Rationally inattentive preferences and hidden information costs. *Theoretical Economics*, 12:621–624, 2017.

Mark Dean and Nathaniel Neligh. Experimental tests of rational inattention. *Unpublished manuscript*, June 2019.

Ernst Fehr and Antonio Rangel. Neuroeconomic foundations of economic choice — recent advances. *Journal of Economic Perspectives*, 25(4):3–30, 2011.

Drew Fudenberg, Philipp Strack, and Tomasz Strzalecki. Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–84, 2018.

Benjamin Hébert and Michael Woodford. Information costs and sequential information sampling. *NBER Working Paper no. 25316*, November 2018.

Jean Jacod and Albert Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2013.

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6): 2590–2615, 2011.

Ian Krajbich, Bastiaan Oud, and Ernst Fehr. Benefits of neuroeconomics modeling: New policy interventions and predictors of preference. *American Economic Review*, 104(5):501–506, 2014.

Harold Kushner and Paul G Dupuis. *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Science & Business Media, 2013.

Filip Matêjka, Alisdair McKay, et al. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105(1):272–98, 2015.

Stephen Morris and Philipp Strack. The Wald problem and the relation of sequential sampling and ex-ante information costs. *Unpublished manuscript*, February 2019.

Giuseppe Moscarini and Lones Smith. The optimal level of experimentation. *Econometrica*, 69(6): 1629–1644, 2001.

Huyên Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.

Luciano Pomatto, Philipp Strack, and Omer Tamuz. The cost of information. *arXiv preprint arXiv:1812.04211*, 2018.

Roger Ratcliff. Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, 92:212–225, 1985.

Roger Ratcliff and Jeffrey N Rouder. Modeling response times for two-choice decisions. *Psychological Science*, 9:347–356, 1998.

R Tyrrell Rockafellar. Convex analysis. 1970.

JF Schouten and JAM Bekker. Reaction time and accuracy. *Acta Psychologica*, 27:143–153, 1967.

Michael Shadlen and Daphna Shohamy. Decision making and sequential sampling from memory. *Neuron*, 90(5):927–939, 2016.

Christopher A Sims. Rational inattention and monetary economics. *Handbook of Monetary Economics*, 3:155–181, 2010.

Jakub Steiner, Colin Stewart, and Filip Matějka. Rational inattention dynamics: Inertia and delay in decision-making. *Econometrica*, 85(2):521–553, 2017.

Satohiro Tajima, Jan Drugowitsch, and Alexandre Pouget. Optimal policy for value-based decision-making. *Nature communications*, 7, 2016.

Satohiro Tajima, Jan Drugowitsch, Nishit Patel, and Alexandre Pouget. Optimal policy for multi-alternative decisions. *Nature Neuroscience*, 22:1503–1511, 2019.

Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

Michael Woodford. Inattentive valuation and reference-dependent choice. *Unpublished manuscript*, May 2012.

Michael Woodford. Stochastic choice: An optimizing neuroeconomic model. Technical Report 5, 2014.

Michael Woodford. Modeling imprecision in perception, valuation and choice. *NBER Working Paper no. 26258*, September 2019.

Weijie Zhong. Indirect information measure and dynamic learning. *arXiv preprint arXiv:1809.00697*, June 2017.

Weijie Zhong. Optimal dynamic information acquisition. *Unpublished manuscript*, January 2019.

A Cost Functions Satisfying Conditions 1-5

In this appendix section, we discuss various families of cost functions that satisfy our conditions.

We first show that any posterior-separable cost function (in the terminology of Caplin et al. [2019]) that is sufficiently convex satisfies conditions 1-5. This includes all uniformly posterior-separable cost functions, such as mutual information and the neighborhood-based cost functions we describe in Hébert and Woodford [2018].

Posterior-separable cost functions are defined as

$$C(p,q;S)=\sum_{s\in S}\pi_s(p,q)D(q_s(p,q)||q),$$

where $D(\cdot||\cdot)$ is a divergence.

Lemma 8. *In the posterior-separable family if cost functions, if the divergence $D(\cdot||\cdot)$ is twice differentiable and strongly convex in its first argument, and differentiable in its second argument, the cost function $C(p,q;S)$ satisfies Conditions 1-5.*

Proof. See the appendix, section B.23. □

These posterior-separable cost functions a popular choice, but by no means the only possibility. As an alternative, consider "state-separable" cost functions,

$$C(p,q;S)=\sum_{x\in X}q_x D(p_x||\pi(p,q);S),$$

where $D(p_x||\pi(p,q);S)$ is a family of divergences defined on the signal alphabets S. Mutual information is both "posterior-separable" and "state-separable," but in there are many cost functions in one family but not the other.

We will assume that the family of divergences is constant with respect to the addition of signals that never occur. That is, if $S\subset S'$, and $e_s^T p'=e_s^T p$ for $s\in S$ and zero otherwise, where e_s is a vector with one corresponding to signal s and zero otherwise, then

$$D(p_x||\pi(p,q);S)=D(p'_x||\pi(p',q);S').$$

We also assume a Blackwell-type inequality for these divergences,

$$D(\Pi r||\Pi r';S')\leq D(r||r';S)$$

for all r,r' and all garbling matrices $\Pi:S\to S'$. Note that this implies that D is invariant in the sense of Chentsov [1982]. Under these assumptions, and some regularity assumptions, we prove that this family also satisfies our conditions.

Lemma 9. *In the state-separable family of cost functions, if the divergences $D(\cdot||\cdot;S)$ are convex in their first argument and twice differentiable, and satisfies, for some $m>0$ and all $r,r'\in\mathcal{P}(S)$ with*

$r' \ll r$,

$$D(r'||r;S) \geq m(r'-r)^T Diag(r)^+(r'-r),$$

then the cost function $C(p,q;S)$ satisfies Conditions 1-5.

Proof. See the appendix, section B.24. □

Lastly, we note that if some of cost function $C(p,q;S)$ satisfies our conditions, so does a convex transformation of that cost function, or a linear combination of two cost functions. Consequently, the two "separable" families described above can be used to generate a huge variety of non-separable cost functions.

Lemma 10. *If $C(p,q;S)$ is a family of cost functions satisfying Conditions 1-5, then so is $C_h(p,q;S) = h(C(p,q;S))$, where $h : \mathbb{R}_+ \to \mathbb{R}_+$ is a strongly convex, strictly-increasing function with $h(0) = 0$. Likewise, if $C_1(p,q;S)$ and $C_2(p,q;S)$ are families of cost functions satisfying Conditions 1-5, then so is $C_\alpha(p,q;S) = \alpha C_1(p,q;S) + (1-\alpha)C_2(p,q;S)$, for any $\alpha \in (0,1)$.*

Proof. The proofs are almost immediate, given the the assumptions. □

Technical Appendix

B Proofs

Notation

In our proofs, we often use a matrix notation in which a prior or posterior q is a vector in $\mathbb{R}^{|X|}$ and a signal structure p is an $|S| \times |X|$ matrix. $p_x \in \mathbb{R}^{|S|}$ and $p_s \in \mathbb{R}^{|X|}$ refer to specific columns and rows of this matrix. We use the notation $e_x \in \mathbb{R}^{|X|}$ and $e_s \in \mathbb{R}^{|X|}$ to refer to basis vectors with a one in the element corresponding to $x \in X$ and $s \in S$ respectively, and zero otherwise. We use the notation ι to refer to a vector of ones (in both $\mathbb{R}^{|S|}$ and $\mathbb{R}^{|X|}$ contexts).

B.1 Proof of Lemma 1

Let p and p' be information structures with signal alphabet S. First, we will show that mixture feasibility and Blackwell monotonicity imply convexity. By mixture feasibility, letting p_M denote the mixture information structure and S_M the signal alphabet,

$$C(p_M, q; S_M) \leq \lambda C(p, q; S) + (1-\lambda)C(p', q; S).$$

Consider the garbling $\Pi : S \times \{1,2\} \to S$, which maps each $(s,i) \in S_M$ to $s \in S$. By Blackwell monotonicity,

$$C(p_M, q; S_M) \geq C(\Pi p_M, q; S).$$

By construction,

$$e_s^T \Pi p_M = \lambda e_s^T p + (1-\lambda) e_s^T p',$$

and the result follows.

Now we show the other direction: that convexity and Blackwell monotonicity imply mixture feasibility. Let p_1 and p_2 be information structures with signal alphabets S_1 and S_2. Because the cost function satisfies Blackwell monotonicity, it is invariant to Markov congruent embeddings. Define $S_M = (S_1 \cup S_2) \times \{1,2\}$. There exists an embedding $\Pi_1 : S_1 \to S_M$ such that, for some $s_M = (s,i) \in S_M$,

$$e_{s_M}^T \Pi_1 p_1 = \begin{cases} 0 & i = 2 \\ 0 & s \notin S_1 \\ e_s^T p_1 & otherwise \end{cases}.$$

Define an embedding Π_2 along similar lines,

$$e_{s_M}^T \Pi_2 p_2 = \begin{cases} 0 & i = 1 \\ 0 & s \notin S_2 \\ e_s^T p_2 & otherwise, \end{cases}$$

and note that these embeddings are left-invertible. It follows by Blackwell monotonicity in both directions that

$$C(\Pi_1 p_1, q; S_M) = C(p_1, q; S_1),$$

and likewise that

$$C(\Pi_2 p_2, q; S_M) = C(p_2, q; S_2).$$

By convexity,

$$C(\lambda \Pi_1 p_1 + (1-\lambda)\Pi_2 p_2; q; S_M) \leq \lambda C(\Pi_1 p_1, q; S_M) + (1-\lambda) C(\Pi_2 p_2, q; S_M).$$

Observing that

$$\lambda \Pi_1 p_1 + (1-\lambda)\Pi_2 p_2 = p_M$$

proves the result.

B.2 Proof of Theorem 1

To prove the theorem, we use Taylor's theorem to approximate the cost function and its gradient up to order Δ_m (a second-order approximation for the cost function, first-order for the gradient).

We start by describing the local (second-order) properties of any information cost function satisfying our conditions. The condition requiring that Blackwell-dominant information structures cost weakly more (Condition 3) is of particular importance. Recall Blackwell's theorem:

Theorem. *(Blackwell [1953]) The information structure $\{p_x\}_{x\in X}$, with signal alphabet S, is more informative, in the Blackwell sense, than $\{p'_x\}_{x\in X}$, with signal alphabet S', if and only if there exists a Markov transition matrix $\Pi : S \to S'$ such that, for all $s' \in S'$ and $x \in X$,*

$$p'_x = \Pi p_x. \tag{25}$$

This Markov transition matrix is known as the "garbling" matrix. Another way of interpreting Condition 3 is that garbled signals are (weakly) less costly than the original signal.

There are certain kinds of garbling matrices that don't actually garble the signals. These garbling matrices have left inverses that are also Markov transition matrices. If we define an information structure $\{p_x\}_{x\in X}$, with signal alphabet S, and another information structure $\{p'_x\}_{x\in X}$, with signal alphabet S', using one of these left-invertible matrices, via equation (25), then $\{p_x\}_{x\in X}$ is more informative than $\{p'_x\}_{x\in X}$, but $\{p'_x\}_{x\in X}$ is also more informative than $\{p_x\}_{x\in X}$. These two information structures are called "Blackwell-equivalent," and it follows that the cost of these two information structures must be equal, by Condition 3. The left-invertible Markov transition matrices associated with Blackwell-equivalent information structures are called Markov congruent embeddings by Chentsov [1982]. Chentsov [1982] studied tensors and divergences that are invariant to Markov congruent embeddings (we will say "invariant" for brevity).

Let Π be a Markov congruent embedding from $\mathcal{P}(S)$ to $\mathcal{P}(S')$. By Condition 3, all information cost functions satisfying our conditions are invariant to Markov congruent embeddings. It necessarily follows that, for any Markov congruent embedding Π, that

$$C(\{p_x\}_{x\in X}, q; S) = C(\{\Pi p_x\}_{x\in X}, q; S').$$

Using this invariance, and results from Chentsov [1982], we will describe the local structure of all information cost functions satisfying our conditions.

The key results of Chentsov [1982] are expressed in terms of the Fisher information matrix. In our context, the Fisher information matrix on the simplex is

$$g(r) = Diag(r)^{+} - \iota\iota^{T},$$

where $Diag(r)^{+}$ is the pseudo-inverse of $Diag(r)$ and ι is a vector of ones. Chentsov establishes the following results:[21]

1. Any continuous function that is invariant over the probability simplex is equal to a constant.

2. Any continuous, invariant 1-form tensor field over the probability simplex is equal to zero.

3. Any continuous, invariant quadratic form tensor field over the probability simplex is proportional to the Fisher information matrix.[22]

These results allow us to characterize the local properties of rational inattention cost functions, via a Taylor expansion. Hold fixed the signal alphabet S, and consider an information structure

$$p_x(\epsilon, \nu) = r + \epsilon\nu_x + \nu\omega_x.$$

Here, $r \in \mathcal{P}(S)$ and $\nu_x \in \mathbb{R}^{|S|}$ satisfies $\iota^T\nu_x = 0$ for all x, where ι is a vector of ones. We also assume that, for all $s \in S$, $e_s^T\nu_x \neq 0$ only if $e_s^T r > 0$. That is, ν_x is an element of the tangent space of the probability simplex at r. The same properties hold true for ω_x. As a result, for values of the perturbation parameters ϵ and ν sufficiently close to zero, $p_x \in \mathcal{P}(S)$ for all $x \in X$. In other words, the parameters ϵ and ν index a two-parameter family of perturbations of an uninformative information structure (corresponding to $\epsilon = \nu = 0$), in which the perturbed information structures will generally be informative; the ν_x and ω_x specify two directions of perturbation. Each of the perturbed information structures has the property that p_x is absolutely continuous with respect to r.

By Condition 1, $C(\{p_x(0,0)\}_{x\in X}; q; S) = 0$. The first order term is

$$\frac{\partial}{\partial\epsilon} C(\{p_x(\epsilon,\nu)\}_{x\in X}, q; S)|_{\epsilon=\nu=0} = \sum_{x\in X} C_x(\{r\}_{x\in X}, q; S)\cdot\nu_x,$$

[21] See Lemma 11.1, Lemma 11.2, and Theorem 11.1 in Chentsov [1982]. See also Proposition 3.19 of Ay et al. [2014], who demonstrate how to extend the Chentsov results to infinite sets X and S.

[22] A 1-form tensor field on a probability simplex $\mathcal{P}$ is a function $T : V \times \mathcal{P} \to \mathbb{R}$, where V is the tangent space of the simplex. Let $\Pi : \mathcal{P} \to \mathcal{P}'$ be a mapping from the simplex $\mathcal{P}$ to the simplex $\mathcal{P}'$, let V' be the tangent space of the simplex $\mathcal{P}'$, and let $d\Pi : V \to V'$ be the pushforward of the mapping Π. The tensor field is invariant under Π if $T(d\Pi v, \Pi p) = T(v, p)$ for all $p \in \mathcal{P}$ and v in the tangent space at p, and a similar definition holds for quadratic form tensor fields.

where C_x denotes the derivative with respect to p_x. This derivative, $C_x(\{r\}; q; S)$, forms a continuous 1-form tensor field over the probability simplex $\mathcal{P}(S)$. By the invariance of $C(\cdot)$, it also follows that C_x is invariant, and therefore, by Chentsov's results, it is equal to zero.

We repeat the argument for the second derivative terms. Those terms can be written as

$$\frac{\partial}{\partial \nu}\frac{\partial}{\partial \epsilon}C(\{p_x(\epsilon,\nu)\}_{x\in X}, q; S)|_{\epsilon=\nu=0} = \sum_{x'\in X}\sum_{x\in X}\omega_{x'}^T \cdot C_{xx'}(\{r\}_{x\in X}, q; S)\cdot \nu_x.$$

By the invariance of $C(\cdot)$, the quadratic form $C_{xx'}(\cdot)$ is invariant for all $x, x' \in X$, and therefore is proportional to the Fisher information matrix for all $x, x' \in X$. We can define a matrix $k(q)$ consisting of the constants of proportionality associated with each $x, x' \in X$. That is,

$$\frac{\partial}{\partial \nu}\frac{\partial}{\partial \epsilon}C(\{p(\cdot|\cdot;\epsilon,\nu)\}, q)|_{\epsilon=\nu=0} = \sum_{x'\in X}\sum_{x\in X}k_{x,x'}(q)\omega_{x'}^T g(r)\nu_x,$$

where $g(r)$ is the Fisher information matrix evaluated at the unconditional distribution of signals $r \in \mathcal{P}(S)$. We note that the matrix-valued function $k(q)$ can depend on the prior q, but cannot depend on the unconditional distribution of signals, r; otherwise, invariance would not hold.

We begin by considering perturbations that preserve the support of the signal structure. As a result, this theorem should be interpreted as applying to "frequent but not very informative" signals, as opposed to "rare but informative" signals. We will discuss the latter type of signals shortly. Note that the pseudo-inverse of the Fisher information matrix is

$$g^+(q) \;=\; Diag(q) - qq^T.$$

Lemma 11. *Suppose that a sequence of information structures p_m, with signal alphabet S, is described by the equation*

$$p_{m,s,x} = \Delta_m^{\alpha(s)} r_s + \Delta_m^{\frac{1}{2}(1+\alpha(s))}\nu_{s,x} + o(\Delta_m^{\frac{1}{2}(1+\alpha(s))}),$$

where, for all $s \in S$, $x \in X$, and $\Delta_m \geq 0$, $p_{m,x,s} \neq 0 \Rightarrow r_s > 0$, $\alpha(s) \in [0, 1)$, and $\sum_{x\in X} v_{s,x} = 0$. Let $C(\cdot)$ be an information cost function that satisfies Conditions 1-4.

There exists a matrix valued function $k(q)$ such that

$$C(p_m; q; S) = \frac{1}{2}\Delta_m \sum_{x'\in X}\sum_{x\in X}k_{x,x'}(q)\nu_{x'}^T g(r)\nu_x + o(\Delta_m).$$

For all q, the matrix-valued function $k(q)$ is continuous, positive semi-definite and symmetric, and satisfies $v^T k(q)v = 0$ for any vector $z \in \mathbb{R}^{|X|}$ that is constant in the support of q.

If in addition the cost function satisfies Condition 5, then there exists a constant $m_g > 0$ such that the difference between $k(q)$ and the pseudo-inverse of the Fisher information matrix, $g^+(q)$, multiplied by that constant, is positive semi-definite: $k(q) \succeq m_g g^+(q)$.

Proof. See the appendix, section B.3. □

In the case of the mutual-information cost function, the matrix $k(q)$ is itself the pseudo-inverse of the Fisher information matrix.

Several authors (Caplin and Dean [2015], Kamenica and Gentzkow [2011]) have observed that it is easier to study rational inattention problems by considering the space of posteriors, conditional on receiving each signal, rather than space of signals. We can redefine the cost function using the posteriors and unconditional signal probabilities, rather than the prior and the conditional probabilities of signals. The corollary below expresses the results of Lemma 11 in terms of posterior beliefs.

Corollary 4. *Under the assumptions of Lemma 11, the posterior beliefs can be written, for any $s \in S$ such that $r_s > 0$, as*

$$q_{s,x}(p_m, q) = q_x + \Delta_m^{\frac{1}{2}(1-\alpha(s))} q_x \frac{\nu_{s,x} - \sum_{x' \in X} q_{x'} \nu_{s,x'}}{r_s} + o(\Delta_m^{\frac{1}{2}(1-\alpha(s))}).$$

The cost function can be written as

$$C(p_m, q; S) = \frac{1}{2} \sum_{s \in S: r_s > 0} \pi_s(p_m, q)(q_{s,x}(p_m, q) - q)^T \bar{k}(q)(q_{s,x}(p_m, q) - q) + o(\Delta_m),$$

where $\bar{k}(q) = Diag(q)^+ k(q) Diag(q)^+$.

Proof. See the appendix, section B.4. □

There are, in effect, two ways for a signal to be contain a small amount of information, and different costs associated with these different types of signals. The results of Lemma 11 characterize, for any rational inattention cost function satisfying our conditions, the cost of receiving frequently, but relatively uninformative, signals. We next consider the cost of receiving a rare but informative signal.

Lemma 12. *Under the assumptions of Lemma 11, define the signal structure*

$$\hat{p}_m = p_m + \Delta_m \omega,$$

where p_m is a signal structure of the type described in Lemma 11, with $\sum_{s \in S} \omega_x = 0$ for all $x \in X$ and with $\omega_{s,x} \geq 0$ for all $s \in S$ such that $p_{m,s,x} = 0$.

The cost of this information structure can be written in the form

$$\begin{aligned} C(\hat{p}_m; q; S) = \frac{1}{2} &\sum_{s \in S: r_s > 0} \pi_s(\hat{p}_m, q)(q_{s,x}(\hat{p}_m, q) - q)^T \bar{k}(q)(q_{s,x}(\hat{p}_m, q) - q) \\ &+ \sum_{s \in S: r_s = 0} \pi_s(\hat{p}_m, q) D^*(q_{s,x}(\hat{p}_m, q) || q) + o(\Delta_m), \end{aligned} \tag{26}$$

where the divergence D^* is finite, convex in its first argument, and twice-differentiable in its first *argument for q' sufficiently close to q, with*

$$\frac{\partial^2 D^*(r||q)}{\partial r^i \partial r^j}|_{r=q} \;=\; \bar{k}(q). \tag{27}$$

Proof. See the appendix, section B.5. □

The divergence D^* represents the cost of acquiring an infrequent, but potentially informative, signal. Naturally, if the signal is in fact not very informative, this cost must be closely related to the costs of other uninformative signals, which gives rise to the condition on the Hessian of the divergence. Note that the lemma demonstrates that the cost is additive with respect to the other signals being received (at least up to order Δ). The result follows from the directional differentiability of the cost function with respect to signals that occur with zero probability and the continuity of that directional derivative (Condition 4) and invariance.

To conclude the proof, observe that by assumption,

$$\pi_s(p_m,q)||q_s(p_m,q)-q||_X^2 \leq B\Delta_m$$

for all $m \in \mathbb{N}$ and $s \in S$. Consequently, for all convergent subsequences of m (denote them by n), either

$$\lim_{n\to\infty} \frac{\pi_s(p_n,q)}{\Delta_n^{\alpha(s)}} \leq B$$

for some $\alpha(s) \in [0,1)$, or $\pi_s(p_n,q) = O(\Delta_n)$.

In the first case, we must have

$$||q_s(p_n,q)-q||_X^2 = O(\Delta_n^{1-\alpha(s)}).$$

In this case, it follows by Taylor's theorem that

$$\frac{1}{2}(q_{s,x}(p_m,q)-q)^T\bar{k}(q)(q_{s,x}(p_m,q)-q) = D^*(q_s(p_m,q)||q) + o(\Delta_m^{1-\alpha(s)}).$$

Defining

$$r_s = \lim_{n\to\infty} \frac{\pi_s(p_n,q)}{\Delta_n^{\alpha(s)}}$$

and

$$v_{s,x} = \lim_{n\to\infty} \frac{q_{s,x}(p_m,q)-q_x}{q_x\pi_s(p_n,q)}\Delta_n^{-\frac{1}{2}(1+\alpha(s))},$$

we can apply Lemma 11.

In the second case, defining

$$\omega_{s,x} = \lim_{n\to\infty} \frac{q_{s,x}(p_n,q)}{q_x\Delta_n}\pi_s(p_n,q)$$

allows us to apply Lemma 12. It follows that, for all convergent subsequences (therefore by boundedness for all m),

$$C(p_m;q;S) = \sum_{s\in S} \pi_s(p_m,q) D^*(q_s(p_m,q)||q) + o(\Delta_m).$$

The claimed properties of the divergence and its Hessian follow from the two lemmas.

B.3 Proof of Lemma 11

Consider a Taylor expansion of $C(p_m,q;S)$ around the value of

$$r_{m,s} = \Delta_m^{\alpha(s)} r_s.$$

We have

$$p_{m,s,x} - r_{m,s} = \Delta_m^{\frac{1}{2}(1+\alpha(s))} \nu_{s,x} + o(\Delta_m^{\frac{1}{2}(1+\alpha(s))}),$$

and therefore by Taylor's theorem we have

$$C(p_m,q;S) = \frac{1}{2} \sum_{x'\in X} \sum_{x\in X} k_{x,x'}(q)(p_{m,x}-r_m)^T g(r_m)(p_{m,x}-r_m) + o(\sum_{s\in S} |p_{m,x,s} - r_{m,s}|^2).$$

By construction,

$$o(\sum_{s\in S} |p_{m,x,s} - r_{m,s}|^2) = o(\Delta_m).$$

Observing that

$$\sum_{x\in X} v_{s,x} = 0,$$

that $p_{m,s,x} - r_{m,s} \neq 0$ if and only if $r_{m,s} > 0$, and using the definition

$$g(r_m) = Diag(r_m)^+ - \iota\iota^T,$$

we have

$$(p_{m,x} - r_m)^T g(r_m)(p_{m,x} - r_m) = \Delta_m^{(1+\alpha(s))} \nu_x^T Diag(r_m)^+ \nu_x + o(\Delta_m),$$

which is

$$(p_{m,x} - r_m)^T g(r_m)(p_{m,x} - r_m) = \Delta_m \nu_x^T Diag(r)^+ \nu_x + o(\Delta_m).$$

It therefore follows that

$$C(p_m;q;S) = \frac{1}{2}\Delta_m \sum_{x'\in X} \sum_{x\in X} k_{x,x'}(q) \nu_{x'}^T g(r) \nu_x + o(\Delta_m).$$

We next demonstrate the claimed properties of $k(q)$. First, $k(q)$ is symmetric and continuous in q, by the symmetry of partial derivatives and the assumption of continuous second derivatives (Condition 4).

Now consider a particular sequence of information structures for which $\nu_{s,x} = \phi_s v_x$, where $v \in \mathbb{R}^{|X|}$ and $\phi \in \mathbb{R}^{|S|}$, with $\sum_{s\in S} e_s^T \phi = 0$, and $\alpha(s) = 0$ for all $s \in S$. Suppose that both v and ϕ

are not zero. For this sequence of information structures,

$$C(p_m, q; S) = \frac{1}{2}\Delta_m \bar{g} v^T k(q) v + o(\Delta_m),$$

where $\phi^T g(r)\phi = \bar{g} > 0$. Suppose the information structure is uninformative for all Δ_m. This would be the case if v is proportional to ι, or any vector constant in the support of q, because such a signal structure has the same distribution of signals in each state in the support of q. Therefore, for such a v,

$$v^T k(q) v = 0$$

by Condition 1. Regardless of whether the information structure is informative, by Condition 1, we must have

$$v^T k(q) v \geq 0,$$

implying that $k(q)$ is positive semi-definite. If z and $-z$ are in the tangent space of the simplex at q, there exists an x, x' $e_x^T z \neq e_{x'}^T z$ with x, x' in the support of q. Using z in the place of v above, by Condition 1, we must have

$$z^T k(q) z > 0.$$

Suppose now that the cost function satisfies Condition 5. Let v be as above, non-zero, and not proportional to ι. We have

$$C(p_m, q; S) = \frac{1}{2}\Delta_m \bar{g} v^T k(q) v + o(\Delta_m),$$

and therefore for the B defined in Condition 5 there exists a Δ_B such that, for all $\Delta < \Delta_B$, $C(p, q; S) < B$. Therefore, we must have

$$C(p_m, q; S) \geq \frac{m}{2} \sum_{s \in S} \pi_s(p, q) ||q_s(p_m, q) - q||_X^2.$$

By Bayes' rule, for any signal that is received with positive probability,

$$q_s(p_m, q) - q = \frac{(Diag(q) - qq^T) p_{m,s}^T}{q^T p_{m,s}^T}.$$

By convention, $q_s = q$ for any s such that $\pi_s(p, q) = 0$.

The support of q_s is always a subset of the support of q, and therefore (by the equivalence of norms),

$$C(p_m, q; S) \geq \frac{m_g}{2} \sum_{s \in S} \pi_s(p, q)(q_s(p_m, q) - q)^T Diag^+(q)(q_s(p_m, q) - q)$$

for some constant $m_g > 0$.

For sufficiently small Δ_m, $\pi_s(p, q) > 0$ if $r_s > 0$, and therefore

$$C(p_m, q; S) \geq \frac{m_g}{2} \sum_{s \in S : r_s > 0} \frac{(p_{m,s}(Diag(q) - qq^T) Diag^+(q)(Diag(q) - qq^T) p_{m,s}^T)}{\pi_s(p, q)},$$

or, for the particular sequence defined by the vectors ϕ and v,

$$C(p_m,q;S) \geq \frac{m_g}{2}\Delta_m \sum_{s\in S:r_s>0} (\phi_s)^2 \frac{v^T(Diag(q)-qq^T)Diag^+(q)(Diag(q)-qq^T)v}{(r_s)} + o(\Delta).$$

Noting that

$$\sum_{s\in S:r_s>0} \frac{(\phi_s)^2}{r_s} = \phi^T g(r)\phi = \bar{g},$$

and that

$$(Diag(q)-qq^T)Diag^+(q)(Diag(q)-qq^T) = g^+(q),$$

we have

$$C(p_m,q;S) \geq \frac{m_g}{2}\Delta_m \bar{g} v^T g^+(q)v + o(\Delta_m).$$

It follows that we must have

$$\frac{1}{2}v^T k(q)v \geq \frac{m_g}{2}v^T g^+(q)v$$

for all v.

B.4 Proof of Corollary 4

Under the stated assumptions,

$$p_{m,s,x} = \Delta_m^{\alpha(s)} r_s + \Delta_m^{\frac{1}{2}(1+\alpha(s))}\nu_{s,x} + o(\Delta_m^{\frac{1}{2}(1+\alpha(s))}).$$

By Bayes' rule, for any $s \in S$ such that $\pi_s(p_m,q) > 0$,

$$q_s(p_m,q) = \frac{Diag(q)p_{m,s}^T}{q^T p_{m,s}^T}.$$

It follows immediately that

$$\lim_{\Delta\to 0^+} q_s(p_m,q) = Diag(q)\frac{\iota r_s}{r_s} = q.$$

Next, using the notation $v_s \in \mathbb{R}^{|X|}$ to denote the vector of $\{v_{s,x}\}_{x\in X}$,

$$\begin{aligned}
\Delta_m^{-\frac{1}{2}(1-\alpha(s))}(q_s(p_m,q)-q) &= \frac{\Delta_m^{-\frac{1}{2}(1+\alpha(s))}}{\Delta_m^{-\alpha(s)}} \frac{(Diag(q)-qq^T)p_s^T}{\pi_s(p_m,q)} \\
&= Diag(q)\frac{\nu_s - \iota q^T\nu_s + o(1)}{\Delta_m^{-\alpha(s)}\pi_s(p_m,q)}.
\end{aligned}$$

For any s such that $r_s > 0$,

$$\lim_{\Delta\to 0^+} \Delta_m^{-\frac{1}{2}(1-\alpha(s))}(q_s(p_m,q)-q) = Diag(q)\frac{\nu_s - \iota q^T\nu_s}{r_s}.$$

By Lemma 11,

$$C(p_m, q; S) = \frac{1}{2}\Delta_m \sum_{x' \in X} \sum_{x \in X} k_{x,x'}(q)\nu_{x'}^T g(r)\nu_x + o(\Delta_m).$$

By the definition of the Fisher matrix, and the observation that $\iota^T \nu_x = 0$ for all $x \in X$,

$$\nu_{x'}^T g(r)\nu_x = \sum_{s \in S: r_s > 0} r_s \frac{\nu_{x',s}}{r_s}\frac{\nu_{x,s}}{r_s}.$$

Substituting in the result regarding the posterior,

$$C(p_m, q; S) = \frac{1}{2} \sum_{s \in S: r_s > 0} \pi_s(p, q)(q_s(p_m, q) - q)^T Diag^+(q)k(q)Diag^+(q)(q_s(p_m, q) - q) + o(\Delta_m),$$

which is the result, observing that $q^T Diag^+(q)$ is constant in the support of q and applying Lemma 11.

B.5 Proof of Corollary 12

By directional differentiability and the continuity of the directional derivatives, there exists a function

$$f(\omega, r, q; S) = \lim_{m \to \infty} \frac{C(p_m + \Delta_m \omega, q; S) - C(p_m, q; S)}{\Delta_m}.$$

Observe that, if ωe_x is in the support of r for all x in the support of q, we must have $f(\omega, \bar{p}, q; S) = 0$, by the results of Lemma 11. Relatedly, if ω and ω' differ only with respect to the frequency of signals in the support of r for all x in the support of q, we must have

$$f(\omega, r, q; S) = f(\omega', r, q; S).$$

Assuming there are signals with $\pi_s(p_m, q) = 0$, we can write $\omega = \omega_1 + \omega_2 + \ldots$, where each ω_i is a perturbation that contains only one signal with $\pi_s(p_m, q) = 0$. Let $N \leq |S|$ denote the number of these perturbations. We can define

$$f_i(\omega_i, r, q; S) = \lim_{m \to \infty} \frac{C(p_{i-1,m} + \Delta\omega_i, q; S) - C(p_{i-1,m}, q; S)}{\Delta_m},$$

where $p_{i-1,m} = p_m + \Delta \sum_{j=1}^{i-1} \omega_i$. By the assumption of the continuity of the directional derivatives,

$$f_i(\omega_i, r, q; S) = f(\omega_i, r, q; S).$$

It follows that

$$f(\omega, r, q; S) = \sum_{i=1}^{N} f(\omega_i, r, q; S).$$

By invariance, the function $f(\omega_i, r, q; S)$ does not depend on r or S. By the argument above, it is only a function of $\omega_{i,s_i} \in \mathbb{R}^{|X|}$, where $s_i \in S$ is the unique signal in ω_i with $r_{s_i} = 0$. By Bayes' rule, if the prior q has full support,

$$\omega_{i,s_i} = (q^T \omega_{i,s_i}) Diag(q)^+ q_{s_i},$$

where q_{s_i} is the posterior associated with signal s_i. If not, by Condition 1, it is without loss of generality to assume $\omega_{i,s_i} e_x = r$ for all x not the support of q, and hence that this equation holds for all q. By the homogeneity of the directional derivative, we can rewrite this as

$$f(\omega_i, r, q; S) = (q^T \omega_{i,s_i}) F(q_{s_i}, q).$$

By the requirement that the cost of an uninformative signal structure is zero, and everything else is costly, we must have

$$F(q, q) = 0,$$

$$F(q', q) > 0$$

for all $q' \neq q$. Therefore, F is a divergence, which we write $D^*(q'||q)$. The finiteness and continuity of $D^*(q'||q)$ is implied by the existence and continuity of the directional derivative. The approximation of the cost function follows from this result and Corollary 4, observing that $\pi_{s_i}(\hat{p}_m, q) = q^T \omega_{i,s_i}$.

By invariance, there exists a Markov congruent embedding that splits each signal in S into $M > 1$ distinct signals in S'. As M becomes arbitrarily large, the probability of each signal becomes small — and in particular, can be of order Δ. It follows for all $s \in S'$ such that $||q_s - q|| = O(\Delta_m^{\frac{1}{2}(1-\alpha(s))})$ (e.g. the signals described in Corollary 4), we must have

$$D^*(q_s||q) = \frac{1}{2} \Delta_m^{(1-\alpha(s))} (q_s^T - q) \bar{k}(q) (q_s - q) + o(\Delta_m^{(1-\alpha(s))}),$$

and therefore $D^*(q'||q)$ must be twice differentiable in q' evaluated at q.

Lastly, we prove that D^* is convex in its first argument. By the convexity of $C(p, q; S)$,

$$C(p_m, q; S) \geq \sum_{s \in S} \pi_s(p_m, q) D^*(q_s(p_m, q)||q).$$

Therefore, for all signal structures p_m^1 and p_m^2 satisfying the conditions of the lemma and all $\lambda \in (0, 1)$, letting $p_m = \lambda p_m^1 + (1 - \lambda) p_m^2$, by convexity

$$\lambda C(p_m^1, q; S) + (1 - \lambda) C(p_m^2, q; S) \geq \sum_{s \in S} \pi_s(p_m, q) D^*(q_s(p_m, q)||q).$$

Taking the limit as $m \to \infty$, we must have

$$\lim_{m \to \infty} \lambda \sum_{s \in S} \pi_s(p_m^1, q) D^*(q_s(p_m^1, q)||q) +$$
$$\lim_{m \to \infty} (1 - \lambda) \sum_{s \in S} \pi_s(p_m^2, q) D^*(q_s(p_m^2, q)||q) \geq \lim_{m \to \infty} \sum_{s \in S} \pi_s(p_m, q) D^*(q_s(p_m, q)||q).$$

For any $q^1, q^2 \in \mathcal{P}(X)$ absolutely continuous with respect to q, define

$$p_m^1 = r + \Delta_m \omega^1,$$
$$p_m^2 = r + \Delta_m \omega^2,$$

where ω^1 and ω^2 are both non-zero only for some $s \in S$ with $r_s = 0$, and satisfy $q^1 = q_s(p_m^1, q)$, $q^2 = q_s(p_m^2, q)$, and $\pi_s(p_m^1, q) = \pi_s(p_m^2, q)$ (which can be achieved by Bayes' rule). It follows that

$$q_s(p_m, q) = \lambda q^1 + (1-\lambda) q^2,$$

and therefore that D^* is convex in its first argument.

B.6 Proof of Theorem 2

We begin by describing three lemmas that we will employ to prove the convergence result. Our first lemma shows that the dual discrete time value function $W(q_t, \lambda; \Delta)$ is well-behaved:

Lemma 13. *If $\lambda \in (0, \kappa c^{-\rho})$ and $\beta = 1$, or if $\beta < 1$, for all $\Delta \leq 1$ the value function $W(q_t, \lambda; \Delta)$ is bounded above on $q_t \in \mathcal{P}(X)$ by a constant $\bar{W}$, bounded below by zero, and is convex in q. Moreover, for all $\Delta \leq 1$,*

$$\kappa - \lambda c^\rho - \ln(\beta)\bar{W} > 0.$$

Proof. See the appendix, section B.7. □

Our next lemma shows that, because of the curvature (ρ) that we impose, the DM will choose, under any optimal policy, to gather only a small amount of information in each time period, as the length of each time period shrinks.

Lemma 14. *Let $n \in \mathbb{N}$ denote a sequence such that $\Delta_n \leq 1$ and $\lim_{n\to\infty} \Delta_n = 0$. Under the assumptions of Lemma 13, any associated sequence of optimal policies $p_{t,n}^*$ satisfies, for all elements of the sequence,*

$$C(p_{t,n}^*, q_{t,n}; S) \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} \Delta_n,$$

where $\theta = \lambda(\rho \frac{\kappa - \lambda c^\rho - \ln(\beta)\bar{W}}{\lambda(\rho-1)})^{\frac{\rho-1}{\rho}}$ and $\bar{W}$ is the upper bound of Lemma 13.

Proof. See appendix, section B.8. □

Our next lemma discuss the convergence of an arbitrary sequence of stochastic processes for beliefs (denoted $q_{t,m}$) and of stopping times (denoted τ_m) to their continuous-time limits, under the assumption that the policies generating them satisfy the bound in Lemma 14 and a bound on expected stopping times. This lemma applies to a sequence of optimal policies, but also to sequences of sub-optimal policies. The lemma describes the convergence of the beliefs process to a martingale, which is not necessarily a diffusion (it may have jumps, or even be a semi-martingale that is not a jump-diffusion).

Lemma 15. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that $\lim_{m\to\infty} \Delta_m = 0$. Let $p_m(q)$ denote a sequence of Markov policies satisfying the bound in Lemma 14. Let $q_{t,m}$ denote the stochastic process for the DM's beliefs at time t, under such a policy, and let τ_m be a sequence of stopping policies such that $E_0[\tau_m] \leq \bar{\tau}$.*

There exists a sub-sequence $n \in \mathbb{N}$ and a probability space such that:

1. *The beliefs $q_{t,n}$ and the stopping time τ_n converge almost surely to a martingale q_t and a stopping time τ.*

2. *The martingale q_t can be represented in terms of its semi-martingale characteristics,*

$$B_t = -\int_0^t (\int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) z dz) dA_s$$

$$C_t = \int_0^t Diag(q_{s^-})\sigma_s \sigma_s^T Diag(q_{s^-}) dA_s$$

$$\nu_t(z) = dA_t \psi_t(x),$$

where σ_s is an $|X| \times |X|$ matrix-valued predictable stochastic process, satisfying $q_{s^-}^T \sigma_s = \vec{0}$, ψ_s is a measure on $\mathbb{R}^{|X|} \setminus \{0\}$ such that $q_{s^-} + z \in \mathcal{P}(X)$ and $q_{s^-} + z \ll q_{s^-}$ for all z in the support of ψ_s, and dA_s is the increment of a weakly increasing process.

3. *For all stopping times T,*

$$E_t[\int_t^T \beta^{s-t} \{\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-} + z || q_{s^-}) dz\} dA_s] \leq$$
$$(\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} \frac{1-\beta^{T-t}}{-\ln(\beta)}.$$

4. *The limit of the cumulative information cost is bounded below,*

$$\lim_{n\to\infty} E_0[\Delta_n^{1-\rho} \sum_{j=0}^{\tau_n \Delta_n^{-1} - 1} \beta^{\Delta_n j} C(p_n(q_{\Delta_n j,n}), q_{\Delta_n j,n}; S)^\rho] \geq$$
$$E_t[\int_0^\tau \beta^s \{\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-} + z || q_{s^-}) dz\}^\rho (\frac{dA_s}{ds})^\rho ds].$$

Proof. See the appendix, section B.9. □

Having described these three lemmas (all proven below), we now proceed to the main proof. Assume that $\lambda \in (0, \kappa c^{-\rho})$ if $\beta = 1$, $\lambda > 0$ if $\beta < 1$. Under this assumption, lemmas 13, 14, and 15 apply.

Let m index a sequence of Markov optimal policies, $p_m^*(q)$, and of stopping times τ_m^*. Let $q_{t,n}^*$ denote the associated process for beliefs. By the uniform boundedness and convexity of the family of value functions $W(q, \lambda; \Delta_m)$, a uniformly convergent sub-sequence exists. Rockafellar [1970] Theorem 10.9 demonstrates that a uniformly convergent sub-sequence exists on the relative interior of the simplex, and Rockafellar [1970] Theorem 10.3 demonstrates that there is a unique extension to a convex and continuous function on the boundary of the simplex. Therefore, by Lemma 13, that proving the discrete time value function converges to the stated continuous time limit also proves that the continuous time limit is bounded and convex.

Pass to this sub-sequence, which (for simplicity) we also index by m, and let $W(q,\lambda)$ denote its limit. Let $W^{+}(q,\lambda)$ denote the continuous time problem defined in Definition 1. We will prove that $W(q,\lambda) = W^{+}(q,\lambda)$.

By Lemmas 13 and 14, the sequence of optimal policies and stopping time satisfies the conditions of Lemma 15. It follows by that lemma that

$$\lim_{n\to\infty} E_0[\int_0^{\tau_n^*} \beta^{\Delta_n\lfloor\Delta_n^{-1}t\rfloor}\Delta_n^{1-\rho}C(p_n^*(q_{t,n}^*),q_{t,n}^*;S)^{\rho}dt] \geq$$
$$E_t[\int_0^{\tau} \beta^{s}\{\frac{1}{2}tr[\sigma_s^*\sigma_s^{*T}k(q_{s-}^*)]+$$
$$\int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^*(z)D^*(q_{s-}+z||q_{s-})dz\}^{\rho}\}^{\rho}(\frac{dA_s^*}{ds})^{\rho}ds],$$

where q_s^* is the limiting stochastic process and $\sigma_s^*, \psi_s^*, dA_s^*$ are associated with the characteristics of the martingale q_s^*.

We also have, by weak convergence,

$$\lim_{n\to\infty} E_0[\beta^{\tau_n}\hat{u}(q_{\tau_n^*,n}) - \Delta_n\frac{1-\beta^{\tau_n^*}}{1-\beta^{\Delta_n}}(\kappa-\lambda c^{\rho}))] = E_0[\beta^{\tau^*}\hat{u}(q_{\tau^*}) - \frac{1}{-\ln(\beta)}(1-\beta^{\tau^*})(\kappa-\lambda c^{\rho}))].$$

Recall also the bound, for any stopping time T measurable with respect filtration generated by q_s^* ,

$$E_t[\int_t^{T} \beta^{s}\{\frac{1}{2}tr[\sigma_s^*\sigma_s^{*T}k(q_{s-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^*(z)D^*(q_{s-}^*+z||q_{s-}^*)dz\}dA_s^*] \leq$$
$$(\frac{\theta}{\lambda})^{\frac{1}{\rho-1}}E_t[\frac{1-\beta^{(T-t)}}{-\ln(\beta)}].$$

It follows that

$$W(q,\lambda) \leq W^{+}(q,\lambda)$$

for all $q \in \mathcal{P}(X)$, where

$$W^{+}(q_t,\lambda) = \sup_{\{\sigma_s,\psi_s,dA_s,\tau\}} E_t[\beta^{(\tau^*-t)}\hat{u}(q_{\tau^*}) - \frac{1}{-\ln(\beta)}(1-\beta^{(\tau^*-t)})(\kappa-\lambda c^{\rho})] -$$
$$-\frac{\lambda}{\rho}E_t[\int_t^{\tau}\beta^{(s-t)}\{\frac{1}{2}tr[\sigma_s\sigma_s^{T}k(q_s)] + \int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s(z)D^*(q_{s-}+z||q_{s-})dz\}^{\rho}(\frac{dA_s}{ds})^{\rho}ds],$$

subject to the constraints, for all stopping times T measurable with respect filtration generated by

q_s^* ,

$$E_t[\int_t^T \beta^{(s-t)}\{\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_t)] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z)D^*(q_{s^-}+z||q_{s^-})dz\}dA_s] \leq$$
$$(\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} E_t[\frac{1-\beta^{(T-t)}}{-\ln(\beta)}].$$

and the evolution of beliefs as implied by the characteristics derived from σ_s, ψ_s, dA_s. Observe, by the arguments in the proof of Lemma 13, that $W^+(q,\lambda)$ is bounded, and convex in q, and satisfies $\kappa - \lambda c^\rho - \ln(\beta)W^+(q_t,\lambda) > 0$.

Also note that, for W^+, it is without loss of generality to set $dA_s = ds$. Scaling dA_s up and scaling $\sigma_s\sigma_s^T$ and ψ_s down, or vice versa, does not change the constraint, and setting $dA_s = 0$ is clearly sub-optimal by the result that $\kappa - \lambda c^\rho - \ln(\beta)W^+(q_t,\lambda) > 0$. Note also that there is a version of the optimal policies which satisfy the constraint everywhere:

$$\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z)D^*(q_{s^-}+z||q_{s^-})dz \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}}.$$

Next, observe that increasing $\sigma_s\sigma_s^T$ by a quantity ϵzz^T results in a first order condition, anywhere W^+ is twice-differentiable, of

$$\lambda\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z'')D^*(q_{s^-}+z''||q_{s^-})dz''\}^{\rho-1}\frac{1}{2}tr[zz^T k(q_{s^-})] \geq$$
$$\frac{1}{2}z^T W_{qq}^+(q_{s^-})z,$$

with equality if the diffusion terms are non-zero in that direction. Note that the bound that the optimal policies satisfy implies that $W_{qq}^+(q_{s^-})$, interpreted in a distributional sense, is finite and hence that W^+ is differentiable.

Similarly, for any z such that $\psi_s^+(z) > 0$, the first-order condition requires that

$$\lambda\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z'')D^*(q_{s^-}+z''||q_{s^-})dz''\}^{\rho-1}D^*(q_{s^-}+z||q_{s^-}) =$$
$$W^+(q_{s^-}+x,\lambda) - W^+(q_{s^-},\lambda) - x^T \cdot W_q^+(q_{s^-},\lambda), \tag{28}$$

where the differentiability of W^+ in the continuation region follows from the envelope theorem.

Combining these two first order conditions, consider a perturbation that decreases $\sigma_s\sigma_s^T$ by ϵzz^T and increases $\psi_s(\nu z)$ and $\psi_s(-\nu z)$ by $\frac{1}{2}\frac{\epsilon}{\nu^2}$. The first-order conditions for this perturbation is

$$\lambda\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z'')D^*(q_{s^-}+z''||q_{s^-})dz''\}^{\rho-1}\times$$
$$\{\frac{1}{2\nu^2}D^*(q_{s^-}+\nu z||q_{s^-}) + \frac{1}{2\nu^2}D^*(q_{s^-}-\nu z||q_{s^-}) - \frac{1}{2}tr[zz^T k(q_{s^-})]\} =$$
$$\frac{1}{2\nu^2}(W^+(q_{s^-}+\nu z,\lambda) + W^+(q_{s^-}-\nu z,\lambda) - 2W^+(q_{s^-},\lambda)) - \frac{1}{2}z^T W_{qq}^+(q_{s^-})z.$$

In the limit as $\nu \to 0^+$, this equation is always satisfied, and therefore it is without loss of generality to suppose that the diffusion term is zero.

Lastly, if there exists an z, z' with $\psi_s^+(z) > 0$ and $\psi_s^+(z') > 0$, an alternative policy that sets $\tilde{\psi}_s^+(z) = \psi_s^+(z) + \frac{D^*(q_{s-}+z'||q_{s-})}{D^*(q_{s-}+z||q_{s-})}\psi_s^+(z')$ and $\tilde{\psi}_s^+(z') = 0$ generates the same cost, and changes utility by

$$\frac{D^*(q_{s^-}+z'||q_{s^-})}{D^*(q_{s^-}+z||q_{s^-})}(W^+(q_{s^-}+z,\lambda) - W^+(q_{s^-},\lambda) - z^T \cdot W_q^+(q_{s^-},\lambda))\psi_s^+(z') -$$
$$(W^+(q_{s^-}+z',\lambda) - W^+(q_{s^-},\lambda) - z'^T \cdot W_q^+(q_{s^-},\lambda))\psi_s^+(z') = 0.$$

It follows that it is without loss of generality to assume that $\psi_s^+(z) > 0$ for at most one value of z. Recalling that the optimal policies are Markov, let $\sigma^+(q_s)$ denote the optimal policy for the diffusion, let $\bar{\psi}^+(q)$ denote the optimal jump intensity, and let $z^+(q)$ denote the Markov optimal jump direction. Any semi-martingale with these characteristics generates a law that is identical to the jump-diffusion process described in Lemma 15.

Noting that $W^+(q,\lambda) \geq W(q,\lambda)$, it follows that if there exists a sequence of policies that converge to the stochastic process q_t^+, characterized by σ^+, $\bar{\psi}^+$, z^+, and whose cumulative information costs $\Delta_n^{-1}C(\cdot)$ converge to the total information costs in definition 1, then such a sequence of policies achieves, in the limit, at least as much utility as any other sequence of policies. It would then be the case that there must be sequence of optimal policies that converges a.s. (as in Lemma 15) to some optimal policy of W^+ (not necessarily the policies that generate q_t^+). Note also by the result above that it is without loss of generality to suppose $\sigma^+ = 0$.

We can rewrite our controls in terms of the jump destination, $q^+(q_s) = q_s + z^+(q_s)$. To construct such a sequence of convergent policies, consider the "constant control" described in chapter 13.2 of Kushner and Dupuis [2013] ("constant controls", in this context, being a constant $q^+, \bar{\psi}_t^+$ pair over the interval $[t, t+\Delta_n)$, switching to $\psi_t = 0$ after the first jump). By theorem 2.3 of that chapter, there exists a sequence of constant controls that converge (weakly) to the optimal policies of W^+. Moreover, these controls result, of the intervals $[t, t+\Delta_n)$, in a two-point distribution, with support on $q^+(q_t)$ and $q_t - \bar{\psi}_t^+(q^+(q_t) - q_t)\Delta_n$ for the left limit of the process at time $t + \Delta_n$.

Define the constant

$$\theta^+ = \frac{E_t[\int_t^\tau \beta^{(s-t)}\bar{\psi}_s^+(q_{s^-})D^*(q_{s^-}+z^+(q_{s^-})||q_{s^-})ds]}{E_t[\frac{1-\beta^{(\tau-t)}}{-\ln(\beta)}]}.$$

Now consider a modification of these constant control policies, which scale the intensity $\bar{\psi}_t^+$ by the quantity $\alpha_n(q_t)$, so that, for the modified policy,

$$\Delta_n C(\cdot) = \theta^+.$$

By the first-order condition with respect to $\bar{\psi}_t$, and the Bellman equation,

$$\kappa - \lambda c^\rho - \ln(\beta)W^+(q_t,\lambda) = \lambda(1-\frac{1}{\rho})(\bar{\psi}_t^+(q_{t^-})D^*(q_{t^-}+z^+(q_{t^-})||q_{t^-}))^\rho.$$

By the convexity of $C(\cdot)$,

$$C(\cdot) \geq \alpha_n(q_t)\bar{\psi}_t^+(q_{t^-})D^*(q_{t^-}+z^+(q_{t^-})||q_{t^-}).$$

Observe that the lower bound on the value function that

$$\kappa - \lambda c^\rho - \ln(\beta)W^+(q_t,\lambda) > 0$$

for all q_t. It follows that

$$\alpha_n(q_t) \in [0, \frac{\theta^+}{\kappa - \lambda c^\rho - \ln(\beta)\underline{\mathrm{W}}}],$$

where $\underline{\mathrm{W}} = \min_{q\in\mathcal{P}(X)} W^+(q,\lambda)$, and that

$$\lim_{n\to\infty} \alpha_n(q_t) = 1.$$

Therefore, by this uniform bound, the modified policies converge weakly to the same limit as the constant control policies, and hence to an optimal policy of W^+. Moreover, by construction, the costs converge, and hence the dual value function W^+ is achievable and therefore $W(q,\lambda) = W^+(q,\lambda)$.

We next demonstrate equality of the primal and dual. The associated Bellman equation for the dual value function W^+, in the continuation region, is

$$\begin{aligned} -\ln(\beta)W^+(q_s,\lambda) = \max_{\sigma_s,\psi_s} & E[dW^+(q_s,\lambda)] - (\kappa-\lambda c^\rho)ds \\ & - \frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_s)] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z)D^*(q_{s^-}+z||q_{s^-})dz\}^\rho. \end{aligned}$$

Consider a perturbation which scales $\sigma_s^+\sigma_s^{+T}$ and ψ_s^+ be some constant $(1+\epsilon)$. Note that such a perturbation would also scale $E[dW^+]$ by $(1+\epsilon)$, and that at least one of σ_s^+ and ψ_s^+ must be non-zero by the assumption that $-\ln(\beta)W^+(q_s,\lambda)+\kappa-\lambda c^\rho > 0$. The first order condition for this perturbation is

$$\begin{aligned} -\ln(\beta)W^+(q_s,\lambda) + (\kappa-\lambda c^\rho) + \frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T}k(q_{s^-})]+ \\ \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z)D^*(q_{s^-}+z||q_{s^-})dz\}^\rho = \\ \lambda\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T}k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z)D^*(q_{s^-}+z||q_{s^-})dz\}^\rho, \end{aligned}$$

which must hold at the optimal policies for this problem. We can rewrite the Bellman equation as

$$-\ln(\beta)W^{+}(q_s,\lambda)ds + (\kappa - \lambda c^{\rho})ds =$$
$$E[dW^{+}(q_s,\lambda)] - \frac{\lambda}{\rho}(\rho\frac{rW^{+}(q_s,\lambda) + (\kappa - \lambda c^{\rho})}{\lambda(\rho-1)})ds,$$

or

$$(-\ln(\beta)W^{+}(q_s,\lambda) + (\kappa - \lambda c^{\rho}))\frac{\rho}{\rho-1} = E[dW^{+}(q_s,\lambda)].$$

Solving this equation,

$$W^{+}(q_s,\lambda) = E_s[\beta^{\frac{\rho}{\rho-1}(\tau^{*}-s)}\hat{u}(q_{\tau^{*}}) - \frac{\rho}{\rho-1}(\kappa - \lambda c^{\rho})\int_s^{\tau^{*}} \beta^{\frac{\rho}{\rho-1}(l-s)}dl].$$

Define λ^{*} by

$$E_s[\beta^{\frac{\rho}{\rho-1}(\tau^{*}-s)}\hat{u}(q_{\tau^{*}}) - \frac{\rho}{\rho-1}(\kappa - \lambda^{*} c^{\rho})\int_s^{\tau^{*}} \beta^{\frac{\rho}{\rho-1}(l-s)}dl] =$$
$$E_0[\beta^{(\tau^{*}-s)}\hat{u}(q_{\tau^{*}}) - \kappa\int_s^{\tau^{*}} \beta^{(l-s)}dl].$$

We can rewrite this as

$$(\frac{1}{\rho-1}\kappa - \frac{\rho}{\rho-1}\lambda^{*}c^{\rho})E_0[\int_s^{\tau^{*}} \beta^{\frac{\rho}{\rho-1}(l-s)}dl] =$$
$$E_0[\beta^{\frac{\rho}{\rho-1}(\tau^{*}-s)}\hat{u}(q_{\tau^{*}}) - \kappa\int_s^{\tau^{*}} \beta^{\frac{\rho}{\rho-1}(l-s)}dl] -$$
$$E_0[\beta^{(\tau^{*}-s)}\hat{u}(q_{\tau^{*}}) - \kappa\int_0^{\tau^{*}} \beta^{(l-s)}dl].$$

The right-hand side is weakly negative, and zero if $\beta = 1$. Consequently, $\lambda^{*} > 0$, and $\lambda^{*} = \frac{\kappa}{\rho c^{\rho}} < \kappa c^{-\rho}$ if $\beta = 1$.

Consider a convergent sub-sequence of $V(q_0;\Delta_n)$ (which exists by the uniform boundedness and convexity of the problem), and denote its limit $V(q_0)$ (again, we will index this sequence by n). By the standard duality inequalities, for all λ,

$$V(q_0;\Delta_n) \leq W(q_0,\lambda;\Delta_n),$$

for all n, and therefore

$$V(q_0) \leq W^{+}(q_0,\lambda^{*}).$$

Consider the value function $\tilde{V}(q_0)$, which is the value function under the feasible optimal policies for $W^{+}(q_0,\lambda^{*})$. It follows that $\tilde{V}(q_0) = W(q_0,\lambda^{*})$, and $\tilde{V}(q_0) \leq V(q_0)$, and therefore $V(q_0) = W(q_0,\lambda^{*})$.

Note that every convergent sub-sequence of $V(q_0;\Delta_n)$ converges to the same function. It follows that

$$V(q_0) = \lim_{\Delta \to 0^+} V(q_0; \Delta).$$
$$= E_0[\beta^{\tau^*} \hat{u}(q_{\tau^*}) - \kappa \int_0^{\tau^*} \beta^l dl].$$

By the definition of λ^* and the Bellman equation,

$$E_0[\int_0^{\tau^*} \beta^s \frac{1}{\rho} \{\frac{1}{2} tr[\sigma_s^* \sigma_s^{*T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|} \setminus \{0\}} \psi_s^*(z) D^*(q_{s^-} + z || q_{s^-}) dz\}^\rho ds] \leq$$
$$c^\rho E_0[\int_0^{\tau^*} \beta^l dl],$$

as required. It follows that the value function is the maximized over all policies satisfying the above constraint (which is the limiting constraint, by the dominated convergence theorem), concluding the proof.

B.7 Proof of Lemma 13

Write the value function in sequence-problem form, for the $\beta < 1$ case:

$$W(q_0, \lambda; \Delta) = \max_{\{p_{j\Delta}\}, \tau} E_0[\beta^\tau \hat{u}(q_\tau) - \kappa \Delta \frac{1 - \beta^\tau}{1 - \beta^\Delta})] -$$
$$\lambda E_0[\Delta^{1-\rho} \sum_{j=0}^{\tau \Delta^{-1}} \beta^{j\Delta} \{\frac{1}{\rho} C(\{p_{j\Delta,x}\}_{x \in X}, q_{j\Delta}(\cdot))^\rho - \Delta^\rho c^\rho\}],$$

Define

$$\bar{u} = \max_{a \in A, x \in X} u(a, x).$$

By the weak positivity of the cost function $C(\cdot)$, it follows that

$$W(q_0, \lambda; \Delta) \leq \bar{u} + \Delta E_0[\frac{1 - \beta^\tau}{1 - \beta^\Delta}](\lambda c^\rho - \kappa).$$

If $\lambda \in [0, \kappa c^{-\rho}]$, the value function is bounded above by $\bar{u}$. If $\lambda > \kappa c^{-\rho}$,

$$W(q_0, \lambda; \Delta) \leq \bar{u} + \frac{\Delta}{1 - \beta^\Delta}(\lambda c^\rho - \kappa),$$

and

$$1 - \beta^\Delta > \frac{-\Delta \ln(\beta)}{1 - \Delta \ln(\beta)},$$

implying

$$\frac{\Delta}{1-\beta^{\Delta}} < \frac{1-\ln(\beta)}{-\ln(\beta)}$$

for all $\Delta \leq 1$. Therefore,

$$\bar{W} = \bar{u} + \frac{1-\ln(\beta)}{-\ln(\beta)} \max\{\lambda c^{\rho} - \kappa, 0\}.$$

It follows immediately that

$$\kappa - \lambda c^{\rho} - \ln(\beta)\bar{W} = \begin{cases} -\ln(\beta)\bar{u} + \kappa - \lambda c^{\rho} & \kappa \geq \lambda c^{\rho} \\ -\ln(\beta)\bar{u} & \kappa < \lambda c^{\rho}, \end{cases}$$

and therefore

$$\kappa - \lambda c^{\rho} - \ln(\beta)\bar{W} > 0.$$

For the $\beta = 1$ case, by the assumption that $\lambda c^{\rho} \leq \kappa$, $W(q_0, \lambda; \Delta) \leq \bar{u} = \bar{W}$, and the result holds immediately.

There is a smallest possible decision utility which is strictly positive, and because stopping now and deciding is always feasible,

$$W(q_0, \lambda; \Delta) \geq 0.$$

We can define the "state-specific" value function, $W(q_t, \lambda; \Delta, x)$, which is the value function conditional on the true state being x. The state-specific value function has a recursive representation, in the region in which the DM continues to gather information:

$$\begin{aligned} W(q_t, \lambda; \Delta, x) \quad = \quad & -\kappa\Delta + \lambda\Delta^{1-\rho}(\Delta^{\rho} c^{\rho} - \frac{1}{\rho} C(\cdot)^{\rho}) + \\ & \beta^{\Delta} \sum_{s \in S:\, e_s^T p_t e_x > 0} (e_s^T p_t^* e_x) W(q_{t+\Delta,s}^*, \lambda; \Delta, x). \end{aligned}$$

In this equation, we take the optimal information structure as given. Note that, by construction, wherever the DM does not choose to stop, the expected value of the state-specific value functions is equal to the value function.

$$\sum_{x \in X} q_{t,x} W(q_t, \lambda; \Delta, x) = W(q_t, \lambda; \Delta).$$

By the optimality of the policies, we have

$$W(q_t, \lambda; \Delta) \geq \sum_{x \in X} q_{t,x} W(q', \lambda; \Delta, x),$$

for any q' in $\mathcal{P}(X)$. Suppose not; then the DM could simply adopt the information structure associated with beliefs q' and achieve higher utility, contradicting the optimality of the policy.

The convexity of the value function follows from the observation that

$$\begin{aligned} W(\alpha q + (1-\alpha)q', \lambda; \Delta) &= \alpha \sum_{x \in X} q_x W(\alpha q + (1-\alpha)q', \lambda; \Delta, x) + \\ & (1-\alpha) \sum_{x \in X} q'_x W(\alpha q + (1-\alpha)q', \lambda; \Delta, x), \end{aligned}$$

and using the inequality above,

$$W(\alpha q + (1-\alpha)q', \lambda; \Delta) \leq \alpha W(q, \lambda; \Delta) + (1-\alpha) W(q', \lambda; \Delta).$$

B.8 Proof of Lemma 14

Consider an alternative policy that mixes (in the sense of Condition 2) the optimal signal structure and an uninformative signal, with probabilities $1-a$ and a, respectively. We must have

$$-\beta^{\Delta_n} \sum_{s \in S} (e_s^T r_{t,n}^*)(W(q_{t,n,s}^*, \lambda; \Delta_n) - W(q_{t,n}, \lambda; \Delta_n)) - \\ \lambda \Delta_n^{1-\rho} C(p_{t,n}^*, q_{t,n})^{\rho-1} \frac{\partial C(p_{t,n}(a), q_{t,n})}{\partial a}|_{a=0^+} \leq 0,$$

which is the first-order condition at the optimal policy in the direction of adding a little bit of the uninformative signal (decreasing a). By the convexity of $C(\cdot)$ and Condition 1,

$$C(p_{t,n}^*, q_{t,n}) + \frac{\partial C(p_{t,n}(a), q_{t,n})}{\partial a}|_{a=0^+} \leq 0,$$

and therefore we must have

$$\beta^{\Delta_n} \sum_{s \in S} (e_s^T r_{t,n}^*)(W(q_{t,n,s}^*, \lambda; \Delta_n) - W(q_{t,n}, \lambda; \Delta_n)) \geq \lambda \Delta_n^{1-\rho} C(p_{t,n}^*, q_{t,n})^{\rho}.$$

Applying the Bellman equation in the continuation region,

$$(1-\beta^{\Delta_n}) W(q_{t,n}, \lambda; \Delta_n) + (\kappa - \lambda c^{\rho})\Delta_n + \frac{\lambda}{\rho} \Delta_n^{1-\rho} C(p_{t,n}^*, q_{t,n})^{\rho} \geq \lambda \Delta_n^{1-\rho} C(p_{t,n}^*, q_{t,n})^{\rho}.$$

Therefore,

$$\lambda(1 - \frac{1}{\rho}) \Delta_n^{-\rho} C(p_{t,n}^*, q_{t,n})^{\rho} \leq (\kappa - \lambda c^{\rho}) + \frac{(1-\beta^{\Delta_n})}{\Delta_n} W(q_{t,n}, \lambda; \Delta_n).$$

If $\beta = 1$, then

$$C(p_{t,n}^*, q_{t,n}) \leq \Delta_n (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}},$$

for the constant $\theta = \lambda(\rho \frac{\kappa - \lambda c^{\rho}}{\lambda(\rho-1)})^{\frac{\rho-1}{\rho}} > 0$.

If $\beta < 1$, note that

$$\frac{(1-\beta^{\Delta_n})}{\Delta_n} < -\ln(\beta).$$

Let $\bar{W}$ denote the upper bound on $W(q_{t,n}, \lambda; \Delta_n)$, which exists by Lemma 13. We have

$$C(p^*_{t,n}, q_{t,n}) \leq \Delta_n (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}},$$

where

$$\theta = \lambda(\rho \frac{\kappa - \lambda c^\rho - \ln(\beta)\bar{W}}{\lambda(\rho-1)})^{\frac{\rho-1}{\rho}}.$$

The constant θ is positive by (13). Note that this generalizes the formula of the $\beta = 1$ case.

B.9 Proof of Lemma 15

We begin by discussing the convergence of stopping times. Let $\bar{W}$ denote the upper bound on $W(q_{t,n}, \lambda; \Delta_n)$, which exists by Lemma 13. Suppose that under an optimal policy,

$$\lim_{T\to\infty} Pr\{\tau_n < T\} = 1 - \alpha < 1.$$

The value function at time T must be bounded above by

$$W(q_T, \lambda; \Delta) \leq (1-\alpha)\bar{W},$$

as the payoff conditional on never stopping is negative. Now consider an alternative policy that follows the optimal policy until time T, and then stops. The difference in the initial value functions is bounded above by the possibility of making the best possible decision under the optimal policy vs. the worst possible decision under the alternative policy, with utility $\underline{\mathrm{u}} > 0$:

$$(1 - \alpha - Pr\{\tau_n < T\})\beta^T \bar{W} \geq (1 - Pr\{\tau_n < T\})\beta^T \underline{\mathrm{u}}.$$

This inequality cannot hold in the limit as $T \to \infty$. Therefore, by the positivity of τ_n, the laws of τ_n are tight, and therefore there exists a sub-sequence that converges in measure. Pass to this sub-sequence (which we will also index by n), and let τ denote the limit of this sub-sequence.

The beliefs $q_{t,n}$ are a family of $\mathbb{R}^{|X|}$-valued stochastic processes, with $q_{t,n} \in \mathcal{P}(X)$ for all $t \in [0,\infty)$ and $n \in \mathbb{N}$. Construct them as RCLL processes by assuming that $q_{\Delta_n j+\epsilon,n} = q_{\Delta_n j,n}$ for all m, $\epsilon \in [0, \Delta_n)$, and $j \in \mathbb{N}$. We next establish that the laws of $q_{t,n}$ are tight. By Condition 5 and Lemma 14,

$$\frac{m}{2}\sum_{s\in S}(e_s^T p_n(q_{t,n})q_{t,n})||q_{s,n}(q_{t,n}) - q_{t,n}||_2^2 \leq C(p_n(q_{t,n}), q_{t,n}; S) \leq \Delta_n(\frac{\theta}{\lambda})^{\frac{1}{\rho-1}},$$

where $q_{s,n}(q)$ is defined by $p_n(q)$ and Bayes' rule. It follows that, for any $\epsilon > 0$, there exists an N_ϵ such that, for all $n > N_\epsilon$,

$$P(||q_{t+\Delta_n,n} - q_{t,n}|| > \epsilon) \leq K_\epsilon \Delta_n,$$

for the constant $K_\epsilon = 2m^{-1}\epsilon^{-2}\theta^{\frac{1}{\rho-1}}$. By Theorem 3.21 in chapter 6 of Jacod and Shiryaev [2013], and the boundedness of $q_{t,n}$, it follows that the laws of $q_{t,n}$ are tight. By Prokhorov's theorem (Theorem 3.9 in chapter 6 of Jacod and Shiryaev [2013]), it follows that there exists a convergent sub-sequence. Pass to this sub-sequence, and let q_t denote the limiting stochastic process. By Proposition 1.1 in chapter 9 of Jacod and Shiryaev [2013], q_t is a martingale with respect to the filtration it generates. By Skorohod's representation theorem, there exists a probability space and random variables (which we will also denote with $q_{t,n}$ and q_t) such the convergence is almost sure. We refer to this probability space and these random variables in what follows.

Note that, by Bayes' rule, if $e_x^T q_{t,n} = 0$ for some $x \in X$ and time t, then $e_x^T q_{s,n} = 0$ for all $s > t$. By Proposition 2.9 and Corollary 2.38 in chapter 2 of Jacod and Shiryaev [2013], we can write the "good" version of the martingale with characteristics

$$B = -\int_0^t (\int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) z dz) dA_s$$

$$C = \int_0^t \Sigma_s dA_s$$

$$\nu = dA_s \psi_s(x).$$

Because beliefs remain in the simplex, $\psi_s(z)$ has support only on z such that $q_s + z \in \mathcal{P}(X)$ and $q_s + z \ll q_s$. Relatedly, $\iota^T \Sigma_s = 0$, and Σ_s can be decomposed as $\Sigma_s = D(q_{s^-})\sigma_s\sigma_s^T D(q_{s^-})$.

By the convexity of the cost function and Theorem 1,

$$C(p_n(q_{t,n}), q_{t,n}; S) \geq \sum_{s\in S} (e_s^T p_n(q_{t,n}) q_{t,n}) D^*(q_{s,n}(q_{t,n}) || q_{t,n}).$$

Defining the process, for arbitrary stopping time T,

$$D_{s,n} = \lim_{\epsilon\to 0^+} D^*(q_{s^-+\epsilon,n} || q_{s^-,n})$$

and

$$D_{t,T,n} = E_t[\int_t^T \beta^{\Delta_n \lfloor \Delta_n^{-1}(s-t)\rfloor} D_{s,n} ds] \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} \Delta_n E_t[\sum_{j=0}^{\lfloor \Delta_n^{-1}(s-t)\rfloor} \beta^{j\Delta_n}],$$

we have by Ito's lemma, almost sure convergence, and the dominated convergence theorem,

$$D_{t,T} = \lim_{n\to\infty} D_{t,T,n} = E_t[\int_t^T \beta^{s-t}\{\frac{1}{2} tr[\sigma_s\sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-} + z || q_{s^-}) dz\} dA_s].$$

Hence, for all such stopping times T,

$$E_t[\int_t^T \beta^{s-t}\{\frac{1}{2} tr[\sigma_s\sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-} + z || q_{s^-}) dz\} dA_s] \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}} \frac{1-\beta^{T-t}}{-\ln(\beta)}.$$

Note also by this argument that

$$
\begin{aligned}
\lim_{n\to\infty} E_0[\Delta_n^{1-\rho} \sum_{j=0}^{\tau_n \Delta_n^{-1}-1} & \beta^{\Delta_n j} C(p_n(q_{\Delta_n j,n}), q_{\Delta_n j,n}; S)^\rho] \\
&= \lim_{n\to\infty} E_0[\int_0^{\tau_n} \beta^{\Delta_n \lfloor \Delta_n^{-1} t \rfloor} \Delta_n^{-\rho} C(p_n(q_{t,n}), q_{t,n}; S)^\rho dt] \\
\geq E_t[\int_0^{\tau} \beta^s \{\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_{s^-})] & + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-}+z||q_{s^-}) dz\}^\rho (\frac{dA_s}{ds})^\rho ds]
\end{aligned}
$$

B.10 Proof of Corollary 1

With $\beta = 1$, the associated Bellman equation for the candidate value function W^+, in the continuation region, is

$$
\begin{aligned}
0 = \max_{\sigma_s,\psi_s} & E[dW^+(q_s,\lambda)] - (\kappa - \lambda c^\rho) ds \\
& - \frac{\lambda}{\rho}\{\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_s)] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z) D^*(q_{s^-}+z||q_{s^-}) dz\}^\rho.
\end{aligned}
$$

Let σ_s^+ and ψ_s^+ denote optimal policies for this problem. Consider a perturbation which scales $\sigma_s^+ \sigma_s^{+T}$ and ψ_s^+ be some constant $(1+\epsilon)$. Note that such a perturbation would also scale $E[dW^+]$ by $(1+\epsilon)$, and that at least one of σ_s^+ and ψ_s^+ must be non-zero by the assumption that $\kappa - \lambda c^\rho > 0$. The first order condition for this perturbation is

$$
\begin{aligned}
(\kappa - \lambda c^\rho) + \frac{\lambda}{\rho}\{\frac{1}{2} tr[\sigma_s^+ \sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z) D^*(q_{s^-}+z||q_{s^-}) dz\}^\rho = \\
\lambda\{\frac{1}{2} tr[\sigma_s^+ \sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s^+(z) D^*(q_{s^-}+z||q_{s^-}) dz\}^\rho,
\end{aligned}
$$

which must hold at the optimal policies for this problem. It follows by the definition of θ in the $\beta = 1$ case (see the proof of Lemma 14),

$$
\theta = \lambda(\rho \frac{\kappa - \lambda c^\rho}{\lambda(\rho - 1)})^{\frac{\rho-1}{\rho}},
$$

that the constraint

$$
\frac{1}{2} tr[\sigma_s \sigma_s^T k(q_{s^-})] + \bar{\psi}_s D^*(q_{s^-}+z_s||q_{s^-}) \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}}
$$

binds with equality everywhere, where we have used the result in the proof of Theorem 2 that it is without loss of generality to suppose $\psi_s(z)$ has support on at most one value of z, which we denote z_s.

Consequently, the Bellman equation can be rewritten as

$$
\max_{\sigma_s,\bar{\psi}_s,z_s} E[dW^+(q_s,\lambda)] - (\kappa - \lambda c^\rho) ds - \frac{\lambda}{\rho}(\frac{\theta}{\lambda})^{\frac{\rho}{\rho-1}} ds
$$

subject to

$$\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_{s^-})] + \bar{\psi}_s D^*(q_{s^-} + z_s||q_{s^-}) \leq (\frac{\theta}{\lambda})^{\frac{1}{\rho-1}}.$$

Defining

$$\chi(\lambda) = (\rho\frac{\kappa - \lambda c^\rho}{\lambda(\rho-1)})^{\frac{1}{\rho}}$$

and observing that

$$\begin{aligned}\kappa - \lambda c^\rho + \frac{\lambda}{\rho}(\frac{\theta}{\lambda})^{\frac{\rho}{\rho-1}} &= \kappa - \lambda c^\rho + \frac{\kappa - \lambda c^\rho}{\rho - 1} \\ &= (\kappa - \lambda c^\rho)\frac{\rho}{\rho-1},\end{aligned}$$

the result follows, noting from the proof of Theorem 2 that $\lambda^* = \frac{1}{\rho}\kappa c^{-\rho}$ when $\beta = 1$, and therefore

$$\chi(\lambda^*) = c\rho^{\frac{1}{\rho}}.$$

B.11 Proof of Lemma 2

Consider a two-signal alphabet, $s \in \{s_1, s_2\}$, with $\pi_{s_1} = \pi_{s_2}$, and $q_{s_1} = (1+\epsilon)q' - \epsilon q$ and $q_{s_2} = (1-\epsilon)q' + \epsilon q$. Applying the "chain rule" inequality,

$$\begin{aligned}D^*(q'||q) + \frac{1}{2}D^*(q' + \epsilon(q'-q)||q') + \frac{1}{2}D^*(q' - \epsilon(q'-q)||q') \\ \leq \frac{1}{2}D^*(q' + \epsilon(q'-q)||q) + \frac{1}{2}D^*(q' - \epsilon(q'-q)||q).\end{aligned}$$

Dividing by ϵ^2 and taking the limit as $\epsilon \to 0^+$,

$$(q'-q)^T \cdot \bar{k}(q') \cdot (q'-q) \leq \frac{d^2}{d\epsilon^2}D^*(q' + \epsilon(q'-q)||q)|_{\epsilon=0}.$$

Since this must hold for all $q' \ll q$, it holds for $q' = q + t(q''-q)$, with some arbitrary $q'' \ll q$ and $t \in [0,1]$. Therefore,

$$\frac{d^2}{dt^2}D^*(q + t(q''-q)||q)|_{t=0} \geq (q''-q)^T \cdot \bar{k}(q + t(q''-q)) \cdot (q''-q).$$

Integrating,

$$D^*(q''||q) \geq \int_0^1 \int_0^s (q''-q)^T \cdot \bar{k}(q + t(q''-q)) \cdot (q''-q)dtds,$$

which is

$$D^*(q''||q) \geq \int_0^1 (1-t)(q''-q)^T \cdot \bar{k}(q + t(q''-q)) \cdot (q''-q)dt.$$

B.12 Proof of Theorem 3

Conjecture that $\lambda \in (0, \kappa c^{-\rho})$. Under this conjecture, lemmas 13, 14, 15, and 2 apply.

Consider a possibly sub-optimal policy which sets $\psi_s(z) = 0$ for all z and satisfies the constraint. The above FOC applies, and therefore we must have

$$tr[\tilde{\sigma}_s \tilde{\sigma}_s^T (D(q_{s^-}) W_{qq}^+(q_{s^-}, \lambda) D(q_{s^-}) - \theta k(q_{s^-}))] \leq 0,$$

where W_{qq}^+ is understood in a distributional sense. It follows that, for all feasible z,

$$W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) \leq \int_0^1 \int_0^s z^T \bar{k}(q_{s^-} + lz) z dl ds.$$

By our assumption of gradual learning (definition 3), this implies that

$$W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) \leq \theta D^*(q_{s^-} + z || q_{s^-}).$$

Hence, it is without loss of generality to assume that $\psi_s^+(z) = 0$ for all z. Note that, if there is a strict preference for gradual learning, the above inequality is strict for all non-zero z. As a result, in this case we must have $\psi_s^+(z) = 0$ for all z. Note also that our control problem involves direct control of the diffusion coefficients, and hence satisfies the standard requirements for the existence and uniqueness of a strong solution to the resulting SDE (Pham [2009] sections 1.3 and 3.2).

B.13 Proof of Theorem 4

The associated Bellman equation, in the continuation region, is (letting $W^+(q, \lambda)$ denote the continuous time value function of Definition 1, generalized to allow multiple jumps)

$$\begin{aligned} 0 = \max_{\sigma_s, \psi_s} & E[dW^+(q_s, \lambda)] + \ln(\beta) W^+(q_{s^-}, \lambda) ds - (\kappa - \lambda c^\rho) ds \\ & - \frac{\lambda}{\rho} \{ \frac{1}{2} tr[\sigma_s \sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|} \setminus \{0\}} \psi_s(z) D^*(q_{s^-} + z || q_{s^-}) dz \}^\rho ds. \end{aligned} \tag{29}$$

Let σ_s^+ and ψ_s^+ denote optimal policies for this problem. Suppose that the constraint does not bind, and consider a perturbation which scales $\sigma_s^+ \sigma_s^{+T}$ and ψ_s^+ be some constant $(1 + \epsilon)$. Note that such a perturbation would also scale $E[dW^+]$ by $(1 + \epsilon)$, and that at least one of σ_s^+ and ψ_s^+ must be non-zero by the assumption that $-\ln(\beta) W^+(q_s, \lambda) + \kappa - \lambda c^\rho > 0$. The first order condition for this perturbation is

$$\begin{aligned} -\ln(\beta) W^+(q_{s^-}, \lambda) + (\kappa - \lambda c^\rho) + \frac{\lambda}{\rho} \{ \frac{1}{2} tr[\sigma_s^+ \sigma_s^{+T} k(q_{s^-})] + \\ \int_{\mathbb{R}^{|X|} \setminus \{0\}} \psi_s^+(z) D^*(q_{s^-} + z || q_{s^-}) dz \}^\rho = \\ \lambda \{ \frac{1}{2} tr[\sigma_s^+ \sigma_s^{+T} k(q_{s^-})] + \int_{\mathbb{R}^{|X|} \setminus \{0\}} \psi_s^+(z) D^*(q_{s^-} + z || q_{s^-}) dz \}^\rho, \end{aligned}$$

which must hold at the optimal policies for this problem.

Define

$$\theta(q_{s^-}) = \lambda(\frac{-\ln(\beta)W^+(q_{s^-},\lambda) + (\kappa - \lambda c^\rho)}{\lambda(1-\frac{1}{\rho})})^{\frac{\rho-1}{\rho}}$$

Observe by Lemma 13 that $\theta(q_{s^-}) > 0$.

For any feasible z, define

$$\tilde{\theta}(q_{s^-}, z) = \min_{\alpha \in [0,1]} \theta(q_{s^-} + \alpha z)$$

and let $\alpha^*(q_{s^-}, z)$ denote the minimizer.

Consider a sub-optimal policy $\tilde{\sigma}_s$ which sets $\psi_s(z) = 0$ and satisfies

$$-\ln(\beta)W^+(q_{s^-} + \alpha^*(q_{s^-}, z)z, \lambda) + (\kappa - \lambda c^\rho) = \lambda(1 - \frac{1}{\rho})\{\frac{1}{2}tr[\sigma_s^+ \sigma_s^{+T} k(q_{s^-})]\}^\rho,$$

which is

$$\frac{1}{2}tr[\tilde{\sigma}_s \tilde{\sigma}_s^T k(q_{s^-})] = (\frac{\tilde{\theta}(q_{s^-}, z)}{\lambda})^{\frac{1}{\rho-1}}.$$

For such a policy, the Bellman equation must be an inequality,

$$\begin{aligned}\frac{1}{2}tr[\tilde{\sigma}_s \tilde{\sigma}_s^T (Diag(q_{s^-})W_{qq}^+(q_{s^-}, \lambda)Diag(q_{s^-})]ds \leq \\ -\ln(\beta)W^+(q_{s^-}, \lambda)ds + (\kappa - \lambda c^\rho)ds + \frac{\lambda}{\rho}\{\frac{1}{2}tr[\tilde{\sigma}_s \tilde{\sigma}_s^T k(q_{s^-})]\}^\rho ds,\end{aligned}$$

where W_{qq}^+ is understood in a distributional sense. We simplify this expression to

$$\begin{aligned}\frac{1}{2}tr[\tilde{\sigma}_s \tilde{\sigma}_s^T (Diag(q_{s^-})W_{qq}^+(q_{s^-}, \lambda)Diag(q_{s^-})] \leq -\ln(\beta)[W^+(q_{s^-}, \lambda) - W^+(q_{s^-} + \alpha^*(q_{s^-}, z)z, \lambda)] \\ + \frac{\tilde{\theta}(q_{s^-}, z)}{2}tr[\tilde{\sigma}_s \tilde{\sigma}_s^T k(q_{s^-})].\end{aligned}$$

This inequality must hold for all $\tilde{\sigma}_s$ with optimal scale. It follows that, integrating along a line (which must lie in the continuation region) and using the positivity of the value function, that

$$\begin{aligned}W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) \leq \tilde{\theta}(q_{s^-}, z) \int_0^1 \int_0^s z^T \bar{k}(q_{s^-} + lz)z dl ds \\ - 2\ln(\beta) \int_0^1 \int_0^t W^+(q_{s^-} + lz) dl dt,\end{aligned}$$

where $W_q^+(q_{s^-}, \lambda)$, the derivative, exists by Theorem 2.

By the strong preference for gradual learning and the upper bound on utility,

$$\begin{aligned}W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) - \tilde{\theta}(q_{s^-}, z) D^*(q_{s^-} + z || q_{s^-}) \leq \\ -\ln(\beta)\bar{u}||z||_2^2 - m||z||_2^{2+\delta}.\end{aligned} \tag{30}$$

If a jump is optimal, we must have (by the first-order condition)

$$W^{+}(q_{s^-}+z,\lambda)-W^{+}(q_{s^-},\lambda)-z^{T}W_{q}^{+}(q_{s^-},\lambda)=$$
$$\lambda\{\psi_s^{+}(z)D^{*}(q_{s^-}+z||q_{s^-})\}^{\rho-1}D^{*}(q_{s^-}+z||q_{s^-}),$$

and

$$\lambda(1-\rho^{-1})\{\psi_s^{+}(z)D^{*}(q_{s^-}+z||q_{s^-})\}^{\rho}=-\ln(\beta)W^{+}(q_{s^-},\lambda)+(\kappa-\lambda c^{\rho}).$$

Therefore, by the monotone relationship between $\theta(q_{s^-})$ and $W^{+}(q_{s^-},\lambda)$,

$$\lambda\{\psi_s^{+}(x)D^{*}(q_{s^-}+z||q_{s^-})\}^{\rho-1}\geq\lambda(\frac{-\ln(\beta)W^{+}(q_{s^-}+\alpha^{*}(q_{s^-},z)z,\lambda)+(\kappa-\lambda c^{\rho})}{\lambda(1-\rho^{-1})})^{\frac{\rho-1}{\rho}},$$

which implies

$$W^{+}(q_{s^-}+z,\lambda)-W^{+}(q_{s^-},\lambda)-z^{T}W_{q}^{+}(q_{s^-},\lambda)\geq\tilde{\theta}(q_{s^-},z)D^{*}(q_{s^-}+z||q_{s^-}).$$

Using equation (30) above,

$$m||z||_2^{\delta}\leq-\ln(\beta)\bar{u},$$

which is

$$||z||_2\leq(-\frac{\bar{u}\ln(\beta)}{m})^{\delta^{-1}}.$$

Now suppose that the jump reduces the value function,

$$W^{+}(q_{s^-}+z,\lambda)\leq W^{+}(q_{s^-},\lambda).$$

Consider again a sub-optimal diffusion policy, but with (for all q)

$$-\ln(\beta)W^{+}(q,\lambda)+(\kappa-\lambda c^{\rho})=\lambda(1-\frac{1}{\rho})\{\frac{1}{2}tr[\sigma_s^{+}\sigma_s^{+T}k(q)]\}^{\rho},$$

which is

$$\frac{1}{2}tr[\tilde{\sigma}_s\tilde{\sigma}_s^{T}k(q)]=(\frac{\theta(q)}{\lambda})^{\frac{1}{\rho-1}}.$$

The Bellman inequality in this case simplifies to

$$\frac{1}{2}tr[\tilde{\sigma}_s\tilde{\sigma}_s^{T}(Diag(q)W_{qq}^{+}(q,\lambda)Diag(q)]\leq\frac{\theta(q)}{2}tr[\tilde{\sigma}_s\tilde{\sigma}_s^{T}k(q)].$$

Observe by Lemma 13 and Theorem 5 that W^{+} is the limit of a sequence of bounded and convex functions, and hence convex. By the convexity of W^{+}, for all $\alpha\in[0,1)$,

$$W^{+}(q_{s^-}+\alpha z,\lambda)\leq W^{+}(q_{s^-},\lambda),$$

and therefore (by the definition of $\theta(q)$) $\theta(q_{s^-} + \alpha z) < \theta(q_{s^-})$. Consequently, integrating, we have

$$W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) \leq \theta(q_{s^-}) \int_0^1 \int_0^s z^T \bar{k}(q_{s^-} + lz) z dl ds,$$

and by the definition of a strong (and hence strict) preference for gradual learning,

$$W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) < \theta(q_{s^-}) D^*(q_{s^-} + x || q_{s^-}).$$

However, if a jump downwards is optimal, we must have (as argued above)

$$W^+(q_{s^-} + z, \lambda) - W^+(q_{s^-}, \lambda) - z^T W_q^+(q_{s^-}, \lambda) = \theta(q_{s^-}) D^*(q_{s^-} + z || q_{s^-}),$$

and therefore downwards jumps are never optimal.

B.14 Proof of Lemma 3

Suppose the cost function satisfies a preference for discrete learning. Consider a two-signal alphabet, $s \in \{s_1, s_2\}$, with $\pi_{s_1} = \pi_{s_2}$, and $q_{s_1} = (1+\epsilon)q' - \epsilon q$ and $q_{s_2} = (1-\epsilon)q' + \epsilon q$. Applying the "chain rule" inequality,

$$\begin{aligned} D^*(q' || q) + \frac{1}{2} D^*(q' + \epsilon(q' - q) || q') + \frac{1}{2} D^*(q' - \epsilon(q' - q) || q') \\ \geq \frac{1}{2} D^*(q' + \epsilon(q' - q) || q) + \frac{1}{2} D^*(q' - \epsilon(q' - q) || q), \end{aligned} \tag{31}$$

strictly if the preference is strict and $q' \neq q$. Dividing by ϵ^2 and taking the limit as $\epsilon \to 0^+$,

$$(q' - q)^T \cdot \bar{k}(q') \cdot (q' - q) \geq \frac{d^2}{d\epsilon^2} D^*(q' + \epsilon(q' - q) || q)|_{\epsilon=0}.$$

Since this must hold for all $q' \ll q$, it holds for $q' = q + t(q'' - q)$, with some arbitrary $q'' \ll q$ and $t \in [0, 1]$. Therefore,

$$\frac{d^2}{dt^2} D^*(q + t(q'' - q) || q)|_{\epsilon=0} \leq (q'' - q)^T \cdot \bar{k}(q + t(q'' - q)) \cdot (q'' - q).$$

Integrating,

$$D^*(q'' || q) \leq \int_0^1 \int_0^s (q'' - q)^T \cdot \bar{k}(q + t(q'' - q)) \cdot (q'' - q) dt ds,$$

which is

$$D^*(q'' || q) \leq \int_0^1 (1 - t)(q'' - q)^T \cdot \bar{k}(q + t(q'' - q)) \cdot (q'' - q) dt.$$

It follows that equality in this equation must hold if the cost function satisfies both a preference for discrete learning and a preference for gradual learning. Consequently, a strict preference for gradual learning is incompatible with a preference for discrete learning. Moreover, in this case equation (31) cannot hold strictly, and therefore a strict preference for discrete learning implies no preference for gradual learning.

B.15 Proof of Theorem 5

The problem described in Corollary 1, using the fact that it is without loss of generality to assume a pure jump process, is

$$W^{+}(q_t,\lambda)=\sup_{\{\bar{\psi}_s,z_s\},\tau} E_t[\hat{u}(q_{\tau^*})-\tau\frac{\rho}{\rho-1}(\kappa-\lambda c^{\rho})]$$

subject to

$$\bar{\psi}_s D^*(q_{s^-}+z_s||q_{s^-})\leq\chi(\lambda).$$

Suppose that the theorem is false– that for some q_{t^-} and z_t^*, $q_{t^-}+z_t^*=q'$ is in the continuation region. The first-order condition (see equation (28) in the proof of Theorem 2, which proves differentiability) can be written as

$$W^{+}(q_{t^-}+z_t^*,\lambda)-W^{+}(q_{t^-},\lambda)-z_t^{*T}\cdot W_q^{+}(q_{t^-},\lambda)=\theta D^*(q'||q_{t^-}).$$

If q' is in the continuation region, there must be some $q''\ll q'$ such that

$$W^{+}(q'',\lambda)-W^{+}(q',\lambda)-(q''-q')^T\cdot W_q^{+}(q',\lambda)=\theta D^*(q''||q').$$

Adding these two equations together and re-arranging,

$$\begin{aligned}W^{+}(q'',\lambda)-W^{+}(q_{t^-},\lambda)-(q''-q_{t^-})^T\cdot W_q^{+}(q_{t^-},\lambda)=\\ \theta D^*(q'||q_{t^-})+\theta D^*(q''||q')+(q''-q')^T\cdot(W_q^{+}(q',\lambda)-W_q^{+}(q_{t^-},\lambda)).\end{aligned}$$

Observe that, because $q''\ll q'$, there exists an $\bar{\epsilon}>0$ such that, for all $\epsilon\in[0,\bar{\epsilon}]$, $q'-\frac{\epsilon}{1-\epsilon}(q''-q')$ remains in the the simplex. By the fact that z_t^* is optimal, for all such ϵ,

$$\begin{aligned}W^{+}(q',\lambda)-W^{+}(q_{t^-},\lambda)-(q'-q_{t^-})^T\cdot W_q^{+}(q_{t^-},\lambda)-\theta D^*(q'||q_{t^-})\geq\\ W^{+}(q'-\frac{\epsilon}{1-\epsilon}(q''-q'),\lambda)-W^{+}(q_{t^-},\lambda)-(q'-\frac{\epsilon}{1-\epsilon}(q''-q')-q_{t^-})^T\cdot W_q^{+}(q_{t^-},\lambda)-\\ \theta D^*(q'-\frac{\epsilon}{1-\epsilon}(q''-q')||q_{t^-}),\end{aligned}$$

or

$$\begin{aligned}\theta D^*(q'||q_{t^-})-\theta D^*(q'-\frac{\epsilon}{1-\epsilon}(q''-q')||q_{t^-})\leq\\ W^{+}(q',\lambda)-W^{+}(q'-\frac{\epsilon}{1-\epsilon}(q''-q'),\lambda)-\frac{\epsilon}{1-\epsilon}(q''-q')^T W_q^{+}(q_{t^-},\lambda).\end{aligned}$$

Now consider the chain rule inequality, supposing that $s\in\{s_1,s_2\}$ with $\pi_{s_1}=\epsilon$, $\pi_{s_2}=1-\epsilon$,

$q_{s_1} = q''$, and $q_{s_2} = q' - \frac{\epsilon}{1-\epsilon}(q'' - q')$,

$$D^*(q'||q_{t^-}) + \epsilon D^*(q''||q') + (1-\epsilon)D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q') \geq$$
$$\epsilon D^*(q''||q_{t^-}) + (1-\epsilon)D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t^-}).$$

We have, for all $\epsilon \in [0, \bar{\epsilon}]$,

$$W^+(q',\lambda) - W^+(q' - \frac{\epsilon}{1-\epsilon}(q''-q'),\lambda) - \frac{\epsilon}{1-\epsilon}(q''-q')^T W_q^+(q_{t^-},\lambda) +$$
$$\epsilon\theta D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t^-}) + \epsilon\theta D^*(q''||q') \geq$$
$$\epsilon\theta D^*(q''||q_{t^-}) - (1-\epsilon)\theta D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q').$$

Dividing by ϵ and taking limits,

$$(q''-q')^T \cdot (W_q^+(q',\lambda) - W_q^+(q_{t^-},\lambda)) +$$
$$\theta D^*(q'||q_{t^-}) + \theta D^*(q''||q') \geq \theta D^*(q''||q_{t^-}).$$

Consequently,

$$W^+(q'',\lambda) - W^+(q_{t^-},\lambda) - (q''-q_{t^-})^T \cdot W_q^+(q_{t^-},\lambda) \geq \theta D^*(q''||q_{t^-}),$$

meaning that it is without loss of generality to suppose that beliefs jump directly to q'' instead of to q'. Therefore, it is without loss of generality to suppose beliefs jump directly to the stopping region.

B.16 Proof of Theorem 6

The associated Bellman equation, in the continuation region, is (letting $W^+(q,\lambda)$ denote the continuous time value function of Definition 1, generalized to allow multiple jumps)

$$0 = \max_{\sigma_s,\psi_s} E[dW^+(q_s,\lambda)] + \ln(\beta)W^+(q_{s^-},\lambda)ds - (\kappa - \lambda c^\rho)ds$$
$$- \frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_{s^-})] + \int_{\mathbb{R}^{|X|}\setminus\{0\}} \psi_s(z)D^*(q_{s^-}+z||q_{s^-})dz\}^\rho ds.$$

Let σ_s^+ and ψ_s^+ denote optimal policies for this problem. Suppose that the constraint does not bind, and consider a perturbation which scales $\sigma_s^+\sigma_s^{+T}$ and ψ_s^+ be some constant $(1+\epsilon)$. Note that such a perturbation would also scale $E[dW^+]$ by $(1+\epsilon)$, and that at least one of σ_s^+ and ψ_s^+ must be non-zero by the assumption that $-\ln(\beta)W^+(q_s,\lambda) + \kappa - \lambda c^\rho > 0$. The first order condition for this perturbation is

$$-\ln(\beta)W^{+}(q_{s^-},\lambda)+(\kappa-\lambda c^{\rho})+\frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s^{+}\sigma_s^{+T}k(q_{s^-})]+$$
$$\int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^{+}(z)D^{*}(q_{s^-}+z||q_{s^-})dz\}^{\rho}=$$
$$\lambda\{\frac{1}{2}tr[\sigma_s^{+}\sigma_s^{+T}k(q_{s^-})]+\int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^{+}(z)D^{*}(q_{s^-}+z||q_{s^-})dz\}^{\rho},$$

which must hold at the optimal policies for this problem.

Define

$$\theta(q_{s^-})=\lambda(\frac{-\ln(\beta)W^{+}(q_{s^-},\lambda)+(\kappa-\lambda c^{\rho})}{\lambda(1-\frac{1}{\rho})})^{\frac{\rho-1}{\rho}}$$

and observe that it is strictly positive by Theorem 2.

If a jump is optimal, we must have (by the above first-order condition)

$$W^{+}(q_{s^-}+z_s^{*},\lambda)-W^{+}(q_{s^-},\lambda)-z_s^{*T}W_q^{+}(q_{s^-},\lambda)=\theta(q_{s^-})D^{*}(q_{s^-}+z_s^{*}||q_{s^-}),$$

where $W_q^{+}(q_{s^-},\lambda)$ is the derivative that exists by Theorem 2.

Suppose that for some q_{t^-} and z_t^{*}, $q_{t^-}+z_t^{*}=q'$ is in the continuation region and that $W^{+}(q',\lambda)\geq W^{+}(q_{t^-},\lambda)$. Then we have

$$\theta(q')\geq\theta(q_{t^-})$$

and, for some $q''\ll q'$,

$$W^{+}(q'',\lambda)-W^{+}(q',\lambda)-(q''-q')^{T}\cdot W_q^{+}(q',\lambda)=\theta(q')D^{*}(q''||q'),$$

and therefore

$$W^{+}(q'',\lambda)-W^{+}(q',\lambda)-(q''-q')^{T}\cdot W_q^{+}(q',\lambda)\geq\theta(q_{t^-})D^{*}(q''||q').$$

We also have the first order condition

$$W^{+}(q_{t^-}+z_t^{*},\lambda)-W^{+}(q_{t^-},\lambda)-z_t^{*T}\cdot W_q^{+}(q_{t^-},\lambda)=\theta(q_{t^-})D^{*}(q'||q_{t^-})$$

and, putting these two equations together,

$$W^{+}(q'',\lambda)-W^{+}(q_{t^-},\lambda)-(q''-q_{t^-})^{T}\cdot W_q^{+}(q_{t^-},\lambda)\geq$$
$$\theta(q_{t^-})D^{*}(q'||q_{t^-})+\theta(q_{t^-})D^{*}(q''||q')+(q''-q')^{T}\cdot(W_q^{+}(q',\lambda)-W_q^{+}(q_{t^-},\lambda)).$$

Observe that, because $q''\ll q'$, there exists an $\bar{\epsilon}>0$ such that, for all $\epsilon\in[0,\bar{\epsilon}]$, $q'-\frac{\epsilon}{1-\epsilon}(q''-q')$

remains in the the simplex. By the fact that z_t^* is optimal, for all such ϵ,

$$\begin{aligned}
W^+(q',\lambda) - W^+(q_{t-},\lambda) - (q'-q_{t-})^T \cdot W_q^+(q_{t-},\lambda) - \theta(q_{t-})D^*(q'||q_{t-}) \geq \\
W^+(q' - \frac{\epsilon}{1-\epsilon}(q''-q'),\lambda) - W^+(q_{t-},\lambda) - (q' - \frac{\epsilon}{1-\epsilon}(q''-q') - q_{t-})^T \cdot W_q^+(q_{t-},\lambda) - \\
\theta(q_{t-})D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t-}),
\end{aligned}$$

or

$$\begin{aligned}
\theta(q_{t-})D^*(q'||q_{t-}) - \theta(q_{t-})D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t-}) \leq \\
W^+(q',\lambda) - W^+(q' - \frac{\epsilon}{1-\epsilon}(q''-q'),\lambda) - \frac{\epsilon}{1-\epsilon}(q''-q')^T W_q^+(q_{t-},\lambda).
\end{aligned}$$

Now consider the chain rule inequality, supposing that $s \in \{s_1, s_2\}$ with $\pi_{s_1} = \epsilon$, $\pi_{s_2} = 1-\epsilon$, $q_{s_1} = q''$, and $q_{s_2} = q' - \frac{\epsilon}{1-\epsilon}(q''-q')$,

$$\begin{aligned}
D^*(q'||q_{t-}) + \epsilon D^*(q''||q') + (1-\epsilon)D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q') \geq \\
\epsilon D^*(q''||q_{t-}) + (1-\epsilon)D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t-}).
\end{aligned}$$

We have, for all $\epsilon \in [0,\bar{\epsilon}]$,

$$\begin{aligned}
W^+(q',\lambda) - W^+(q' - \frac{\epsilon}{1-\epsilon}(q''-q'),\lambda) - \frac{\epsilon}{1-\epsilon}(q''-q')^T W_q^+(q_{t-},\lambda) + \\
\epsilon\theta(q_{t-})D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q_{t-}) + \epsilon\theta(q_{t-})D^*(q''||q') \geq \\
\epsilon\theta(q_{t-})D^*(q''||q_{t-}) - (1-\epsilon)\theta(q_{t-})D^*(q' - \frac{\epsilon}{1-\epsilon}(q''-q')||q').
\end{aligned}$$

Dividing by ϵ and taking limits,

$$\begin{aligned}
(q''-q')^T \cdot (W_q^+(q',\lambda) - W_q^+(q_{t-},\lambda)) + \\
\theta(q_{t-})D^*(q'||q_{t-}) + \theta(q_{t-})D^*(q''||q') \geq \theta(q_{t-})D^*(q''||q_{t-}).
\end{aligned}$$

Consequently,

$$W^+(q'',\lambda) - W^+(q_{t-},\lambda) - (q''-q_{t-})^T \cdot W_q^+(q_{t-},\lambda) \geq \theta(q_{t-})D^*(q''||q_{t-}),$$

and therefore it is without loss of generality to suppose that beliefs jump directly to q'' instead of to q'. Note that this inequality is strict if $W^+(q',\lambda) > W^+(q_{t-},\lambda)$. Hence it follows that if beliefs jump in such a way that increases the value function, they must jump to the stopping region.

Observe by Lemma 13 and Theorem 2 that W^+ is the limit of a sequence of bounded and convex functions, and hence convex. It follows that if $W^+(q',\lambda) = W^+(q_{t-},\lambda)$, we must have (by the mean value theorem) $(q'-q_{t-})^T \cdot W_q^+(\alpha q_{t-} + (1-\alpha)q',\lambda) = 0$ for some $\alpha \in (0,1)$, and therefore by convexity $(q'-q_{t-})^T \cdot W_q^+(q'_{t-},\lambda) \leq 0$. If such a jump were optimal, we would require $\theta(q_{s-})D^*(q'||q_{t-}) \leq 0$, which cannot hold. Therefore, jumps to the same level of the value function do not occur.

Now suppose that

$$W^{+}(q', \lambda) < W^{+}(q_{t^-}, \lambda)$$

and therefore $\theta(q') < \theta(q_{t^-})$. Define

$$q'' = \alpha q' + (1 - \alpha) q_{t^-}$$

for some $\alpha \in (0, 1)$.

By the convexity of W^{+}, for all $\alpha \in [0, 1)$, $W^{+}(q'', \lambda) < W^{+}(q_{t^-}, \lambda)$, and therefore

$$W^{+}(q', \lambda) - W^{+}(q'', \lambda) - (q' - q'')^{T} \cdot W_{q}^{+}(q'', \lambda) \geq \theta(q_{t^-}) D^{*}(q' || q'')$$

and

$$W^{+}(q', \lambda) - W^{+}(q'', \lambda) - (q' - q'')^{T} \cdot W_{q}^{+}(q'', \lambda) > \theta(q'') D^{*}(q' || q'').$$

Define $\bar{\psi}'$ by

$$\lambda(\bar{\psi}' D^{*}(q' || q''))^{\rho - 1} D^{*}(q' || q'') = W^{+}(q', \lambda) - W^{+}(q'', \lambda) - (q' - q'')^{T} \cdot W_{q}^{+}(q', \lambda).$$

We have

$$\lambda(\bar{\psi}' D^{*}(q' || q''))^{\rho - 1} > \lambda\left(\frac{-\ln(\beta) W^{+}(q'', \lambda) + (\kappa - \lambda c^{\rho})}{\lambda(1 - \frac{1}{\rho})}\right)^{\frac{\rho - 1}{\rho}}$$

and therefore

$$-\ln(\beta) W^{+}(q'', \lambda) + (\kappa - \lambda c^{\rho}) < \lambda(1 - \frac{1}{\rho})(\bar{\psi}' D^{*}(q' || q''))^{\rho},$$

which is

$$\begin{aligned} -\ln(\beta) W^{+}(q'', \lambda) + (\kappa - \lambda c^{\rho}) < & -\frac{\lambda}{\rho}(\bar{\psi}' D^{*}(q' || q''))^{\rho} \\ & + \bar{\psi}'[W^{+}(q', \lambda) - W^{+}(q'', \lambda) - (q' - q'')^{T} \cdot W_{q}^{+}(q', \lambda)]. \end{aligned}$$

It follows that the policy $z' = (q' - q'')$ and $\bar{\psi}'$ violates the HJB equation at q'', and therefore jumps downward never occur.

Hence we conclude that only upward jumps in the value function occur, and only to the stopping region.

B.17 Proof of Theorem 7

The associated Bellman equation, in the continuation region, is (letting $W^{+}(q,\lambda)$ denote the continuous time value function of Definition 1, generalized to allow multiple jumps)

$$0=\max_{\sigma_s,\psi_s} E[dW^{+}(q_s,\lambda)]+\ln(\beta)W^{+}(q_{s^-},\lambda)ds-(\kappa-\lambda c^{\rho})ds \\ -\frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s\sigma_s^T k(q_{s^-})]+\int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s(z)D^*(q_{s^-}+z||q_{s^-})dz\}^{\rho}ds.$$

Let σ_s^+ and ψ_s^+ denote optimal policies for this problem. Suppose that the constraint does not bind, and consider a perturbation which scales $\sigma_s^+\sigma_s^{+T}$ and ψ_s^+ be some constant $(1+\epsilon)$. Note that such a perturbation would also scale $E[dW^+]$ by $(1+\epsilon)$, and that at least one of σ_s^+ and ψ_s^+ must be non-zero by the assumption that $-\ln(\beta)W^{+}(q_s,\lambda)+\kappa-\lambda c^{\rho}>0$. The first order condition for this perturbation is

$$-\ln(\beta)W^{+}(q_{s^-},\lambda)+(\kappa-\lambda c^{\rho})+\frac{\lambda}{\rho}\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T}k(q_{s^-})]+ \\ \int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^+(z)D^*(q_{s^-}+z||q_{s^-})dz\}^{\rho}= \\ \lambda\{\frac{1}{2}tr[\sigma_s^+\sigma_s^{+T}k(q_{s^-})]+\int_{\mathbb{R}^{|X|}\setminus\{0\}}\psi_s^+(z)D^*(q_{s^-}+z||q_{s^-})dz\}^{\rho},$$

which must hold at the optimal policies for this problem.

Define

$$\theta(q_{s^-})=\lambda(\frac{-\ln(\beta)W^{+}(q_{s^-},\lambda)+(\kappa-\lambda c^{\rho})}{\lambda(1-\frac{1}{\rho})})^{\frac{\rho-1}{\rho}}$$

and observe that it is strictly positive by Theorem 2.

Because a jump is optimal, we must have (by the above first-order condition)

$$W^{+}(q_{s^-}+z_s^*,\lambda)-W^{+}(q_{s^-},\lambda)-z_s^{*T}W_q^{+}(q_{s^-},\lambda)=\theta(q_{s^-})D^*(q_{s^-}+z_s^*||q_{s^-}),$$

where $W_q^{+}(q_{s^-},\lambda)$ is the derivative that exists by Theorem 2, and for all feasible jumps,

$$W^{+}(q_{s^-}+z,\lambda)-W^{+}(q_{s^-},\lambda)-z^{T}W_q^{+}(q_{s^-},\lambda)\leq\theta(q_{s^-})D^*(q_{s^-}+z||q_{s^-}).$$

We begin by proving that a preference for gradual learning exists for two-signal alphabets, and assuming that all of the relevant elements of the simplex are interior. We then extend the result to prove the full preference for discrete learning.

Proof by contradiction: suppose there exists an interior $q,q',q_1,q_2\in\mathcal{P}(X)$ and $\pi\in(0,1)$ such that

$$\pi q_1+(1-\pi)q_2=q'$$

and

$$D^*(q'|q)+\pi D^*(q_1|q')+(1-\pi)D^*(q_2||q')<\pi D^*(q_1|q)+(1-\pi)D^*(q_2||q).$$

Now suppose that there exists a utility function such that $z = q_1 - q$ and $z = q_2 - q$ are both optimal policies from q, and for which

$$W^+(q', \lambda) \le W^+(q, \lambda).$$

Then we must have, for $i \in \{1, 2\}$,

$$W^+(q_i, \lambda) - W^+(q, \lambda) - (q_i - q)^T \cdot W_q^+(q, \lambda) = \theta(q) D^*(q_i || q),$$

$$W^+(q', \lambda) - W^+(q, \lambda) - (q' - q)^T \cdot W_q^+(q, \lambda) \le \theta(q) D^*(q' || q),$$

$$W^+(q_i, \lambda) - W^+(q', \lambda) - (q_i - q')^T \cdot W_q^+(q', \lambda) \le \theta(q') D^*(q_i || q') \le \theta(q) D^*(q_i || q'),$$

where $\theta(q') \le \theta(q)$ by the definition of $\theta(\cdot)$ and $W^+(q', \lambda) \le W^+(q, \lambda)$. Putting these together,

$$\theta(q) D^*(q' || q) + \theta(q) D^*(q_i || q') - \theta(q) D^*(q_i || q) \ge -(q_i - q')^T \cdot [W_q^+(q', \lambda) - W_q^+(q, \lambda)].$$

It would follow in this case that

$$D^*(q' | q) + \pi D^*(q_1 | q') + (1 - \pi) D^*(q_2 || q') \ge \pi D^*(q_1 | q) + (1 - \pi) D^*(q_2 || q),$$

a contradiction. To prove the result, we construct such a utility function. Note that our construction below will assume there are three actions; when applying this proof to the case of $|X| = 2$, one of the actions will be redundant.

Define, for some $\mu = (0, 1)$, a q_3 such that

$$\mu q_3 + (1 - \mu) q' = q.$$

Note that such a q_3 exists by the assumption that q is in the interior of the simplex.

Let $v \in \mathbb{R}^{|X|}$ be a vector and let k_1, k_2, k_3, K be constants. Suppose there are three actions, and let their utilities satisfy

$$u_i \in \theta(q) \partial D^*(q_i || q) + v + \iota k_i,$$

where $\partial D^*(q_i || q)$ denotes the sub-gradient with respect to the first argument. This sub-gradient exists by the convexity of D^* in its first argument and the assumption that q_i is interior. Define

$$k_i = \theta(q) D^*(q_i || q) - q_i^T \cdot \theta(q) \partial D^*(q_i || q)] + K - q^T v$$

so that

$$\theta(q) D^*(q_i || q) = q_i^T \cdot u_i - K + (q - q_i)^T v.$$

Observe by convexity that

$$\begin{aligned} q_i^T(u_i - u_j) &= \theta(q)D^*(q_i||q) + K - (q - q_i)^T v \\ &- \theta(q)D^*(q_j||q) - K + (q - q_j)^T v \\ &- (q_i - q_j)^T \cdot u_j, \end{aligned}$$

and by the definition of the sub-gradient this yields

$$q_i^T(u_i - u_j) \geq 0.$$

By sub-optimality, for any $q'' \ll q$ and any $i \in \{1, 2, 3\}$,

$$\theta(q)D^*(q''||q) \geq (q'')^T u_i - W^+(q, \lambda) - (q'' - q)^T \cdot W_q^+(q, \lambda).$$

Therefore, for all $i \in \{1, 2, 3\}$,

$$(q_i - q)^T \cdot (W_q^+(q, \lambda) - v) \geq K - V(q)$$

Since this must hold for all q_i, we must have

$$W^+(q, \lambda) = K$$

and

$$(q_i - q)^T \cdot (W_q^+(q, \lambda) - v) = 0.$$

Hence it follows that

$$(q' - q)^T \cdot W_q^+(q, \lambda) = (q' - q)v.$$

By sub-optimality,

$$\theta(q)D^*(q'||q) + (q' - q)^T \cdot v \geq W^+(q', \lambda) - W^+(q, \lambda).$$

Setting

$$v \in -\theta(q)\partial D^*(q'||q)$$

ensures by convexity that $W^+(q, \lambda) \geq W^+(q', \lambda)$.

Observe that

$$\begin{aligned} \theta(q)D^*(q_i||q) &= q_i^T \cdot u_i - K + (q - q_i)^T v \\ &= q_i^T \cdot u_i - W^+(q', \lambda) + (q - q_i)^T W_q^+(q, \lambda) \end{aligned}$$

and therefore that jumps to the points q_i are optimal. Observe also that for any other q'', by the definition of the sub-gradient,

$$\theta(q)D^*(q''||q) - \theta(q)D^*(q_i||q) \geq (q'' - q_i)^T(u_i - v)$$

and therefore

$$\theta(q)D^*(q''||q) \geq (q'')^T u_i - W^+(q', \lambda) + (q - q'')^T W_q^+(q, \lambda)$$

as required. Therefore, the stationary policy of jumping to $\{q_1, q_2, q_3\}$ in proportions $\pi(1-\mu), (1-\pi)(1-\mu), \mu$ is optimal.

We conclude that for all interior pairs,

$$D^*(q'|q) + \pi D^*(q_1|q') + (1-\pi)D^*(q_2||q') \geq \pi D^*(q_1|q) + (1-\pi)D^*(q_2||q).$$

The result extends immediately to more than two $\{q_s\}$ by adding this expression for different pairs. The result extends to the boundary of the simplex by continuity.

B.18 Proof of Lemma 4

Recall the definition of a preference for discrete learning: for all $q, q', \{q_s\}_{s\in S}$ with $q' \ll q$ and $\sum_{s\in S} \pi_s q_s = q'$,

$$D^*(q'||q) + \sum_{s\in S} \pi_s D^*(q_s||q') \geq \sum_{s\in S} \pi_s D^*(q_s||q)$$

Therefore, for all $z \in \mathbb{R}^{|X|}$ with support on the support of q' and ϵ sufficiently small,

$$D^*(q'||q' + \epsilon z) + \sum_{s\in S} \pi_s D^*(q_s||q') \geq \sum_{s\in S} \pi_s D^*(q_s||q' + \epsilon z).$$

It follows immediately by the differentiability assumption that

$$\sum_{s\in S} \pi_s \frac{\partial}{\partial \epsilon} D^*(q_s||q' + \epsilon z)|_{\epsilon=0} = 0.$$

By step 1 in the proof of theorem 4 of Banerjee et al. [2005], it follows immediately that

$$D^*(q'||q) = H(q') - H(q) - (q' - q)^T H_q(q)$$

for some convex function H, where H_q denotes the gradient. Note that theorem 4 of Banerjee et al. [2005] is stated as requiring that

$$\sum_{s\in S} \pi_s D^*(q_s||q' + \epsilon z)$$

be minimized at $\epsilon = 0$ for all z, but step 1 of the proof in fact only requires that $\epsilon = 0$ correspond to a critical value for all z. Step 2 of the proof relaxes slightly the regularity conditions, but we have simply assumed these. Minimization is only required to establish the last step of the proof, step 3, which proves strict convexity of H. Strict convexity of $H(q)$ on the support of q follows in our setting immediately from the properties of the $k(q)$ matrix (Theorem 1).

B.19 Proof of Lemma 5

Define $M(q_t)$ as the set of $|X| \times |X|$ matrices such that, for all $\sigma \in M(q_t)$, $q_t^T \sigma = \vec{0}$.

In the continuation region, everywhere the value function is twice differentiable,

$$\sup_{\sigma_t \in M(q_t)} \frac{1}{2} tr[\sigma_t^T D(q_t) V_{qq}(q_t) D(q_t) \sigma_t] = \kappa,$$

subject to

$$\frac{1}{2} tr[\sigma_t^T k(q_t) \sigma_t] \leq \chi.$$

First, suppose that the constraint does not bind and a maximizing optimal policy exists:

$$\frac{1}{2} tr[\sigma_t^{*T} k(q_t) \sigma_t^*] = a\chi,$$

where σ_t^* is a maximizer, for some $a \in [0,1)$ ($a \geq 0$ by the positive semi-definiteness of $k(q_t)$). For any $c \in (1, a^{-1})$, with $a^{-1} = \infty$ for $a = 0$, if we used $\sigma_t = c\sigma_t^*$ instead, the policy would be feasible and we would have

$$\frac{1}{2} tr[\sigma_t^T D(q_t) V_{qq}(q_t) D(q_t) \sigma_t] = c^2 \kappa > \frac{1}{2} tr[\sigma_t^{*T} D(q_t) V_{qq}(q_t) D(q_t) \sigma_t^*] = \kappa,$$

a contradiction by the fact that $\kappa > 0$. Therefore, either the constraint binds under the optimal policy or an optimal policy does not exist. The latter would require that, for some non-zero vector $z \in \mathbb{R}^{|X|}$ with $zz^T \in M(q_t)$,

$$z^T D(q_t) V_{qq}(q_t) D(q_t) z > 0$$

and $z^T k(q_t) z = 0$, but the null space of $k(q_t)$ consists only of vectors whose elements are constant over the support of q_t by Theorem 1, and therefore satisfy $q^T z \neq 0$, implying that $zz^T \notin M(q_t)$. Therefore, the constraint binds, and an optimal policy exists.

Using θ as defined in the lemma, it must be the case (anywhere the DM chooses not to stop and the value function is twice differentiable) that

$$\max_{\sigma_t \in M(q_t)} \frac{1}{2} tr[\sigma_t \sigma_t^T (Diag(q_t) V_{qq}(q_t) Diag(q_t) - \theta k(q_t))] = 0.$$

B.20 Proof of Theorem 8

Define $\phi(q_t)$ as the static value function in the statement of the theorem (we will prove that it is equal to $V(q_t)$, the value function of the dynamic problem). We first show that $\phi(q_t)$ satisfies the HJB equation, can be implemented by a particular strategy for the DM, and that any other strategy for the DM achieves weakly less utility. We begin by observing that

$$\iota^T k(q_t) Diag(q_t)^{-1} = 0 = \iota^T Diag(q_t) H_{qq}(q_t) = q_t^T H_{qq}(q_t),$$

and therefore converse of Euler's homogenous function theorem applies. That is, $H_q(q_t)$ is homogenous of degree zero, and $H(q_t)$ is homogeneous of degree one.

We start by showing that the function $\phi(q_t)$ is twice-differentiable in certain directions. Substituting the definition of the divergence into the statement of theorem,

$$\phi(q_0) = \max_{\pi \in \mathcal{P}(A), \{q_a \in \mathcal{P}(X)\}_{a \in A}} \sum_{a \in A} \pi(a) u_a^T \cdot q_a + \theta H(q_0) - \theta \sum_{a \in A} \pi(a) H(q_a),$$

subject to the same constraint. Define a new choice variable, $\hat{q}_a = \pi(a) q_a$. By definition, $\hat{q}_a \in \mathbb{R}_+^{|X|}$, and the constraint is $\sum_{a \in A} \hat{q}_a = q_0$. By the homogeneity of H, the objective is

$$\phi(q_0) = \max_{\pi \in \mathcal{P}(A), \{q_a \in \mathcal{P}(X)\}_{a \in A}, \{\hat{q}_a \in \mathcal{P}(X)\}_{a \in A}} \sum_{a \in A} u_a^T \cdot \hat{q}_a + \theta H(q_0) - \theta \sum_{a \in A} H(\hat{q}_a).$$

Any choice of $\hat{q}_a$ satisfying the constraint can be implemented by some choice of π and q_a in the following way: set $\pi(a) = \iota^T \hat{q}_a$, and (if $\pi(a) > 0$) set

$$q_a = \frac{\hat{q}_a}{\pi(a)}.$$

If $\pi(a) = 0$, set $q_a = q_0$. By construction, the constraint will require that $\pi(a) \leq 1$, $\sum_{a \in A} \pi(a) = 1$, and the fact that the elements of q_a are weakly positive will ensure $\pi(a) \geq 0$. Similarly, $\iota^T q_a = 1$ for all $a \in A$, and the elements of q_a are weakly greater than zero. Therefore, we can implement any set of $\hat{q}_a$ satisfying the constraints.

Rewriting the problem in Lagrangian form,

$$\begin{aligned}\phi(q_0) = \max_{\{\hat{q}_a \in \mathbb{R}^{|X|}\}_{a \in A}} \min_{\kappa \in \mathbb{R}^{|X|}, \{\nu_a \in \mathbb{R}_+^{|X|}\}_{a \in A}} & \sum_{a \in A} u_a^T \cdot \hat{q}_a + \theta H(q_0) \\ & - \theta \sum_{a \in A} H(\hat{q}_a) + \kappa^T (q_0 - \sum_{a \in A} \hat{q}_a) + \sum_{a \in A} \nu_a^T \hat{q}_a.\end{aligned}$$

Observe that $\phi(q_0)$ is convex in q_0. Suppose not: for some $q = \lambda q_0 + (1-\lambda) q_1$, with $\lambda \in (0,1)$, $\phi(q) < \lambda \phi(q_0) + (1-\lambda)\phi(q_1)$. Consider a relaxed version of the problem in which the DM is allowed to choose two different $\hat{q}_a$ for each a. Because of the convexity of H, even with this option, the DM will set both of the $\hat{q}_a$ to the same value, and therefore the relaxed problem reaches the same value as the original problem. However, in the relaxed problem, choosing the optimal policies for q_0 and q_1 in the original problem, scaled by λ and $(1-\lambda)$ respectively, is feasible. It follows that $\phi(q) \geq \lambda \phi(q_0) + (1-\lambda)\phi(q_1)$. Note also that $\phi(q_0)$ is bounded on the interior of the simplex. It follows by Alexandrov's theorem that is is twice-differentiable almost everywhere on the interior of the simplex.

By the convexity of H, the objective function is concave, and the constraints are affine and a feasible point exists. Therefore, the KKT conditions are necessary. Anywhere the objective function is continuously differentiable in the choice variables and in q_0, and therefore the envelope theorem applies. We have, by the envelope theorem,

$$\phi_q(q_0) = \theta H_q(q_0) + \kappa,$$

and the first-order conditions (for all $a \in A$ with $\hat{q}_a \neq \vec{0}$),

$$u_a - \theta H_q(\hat{q}_a) - \kappa + \nu_a = 0. \tag{32}$$

If $\hat{q}_a = \vec{0}$, we must have $q^T(u_a - \kappa) \leq \theta H(q)$ for all q, meaning that $u_a - \kappa$ is a sub-gradient of $H(q)$ at $q = 0$. In this case, we can define $\nu_a = \vec{0}$ and observe that the first-order condition holds for an appropriately-chosen sub-gradient. Define $\hat{q}_a(q_0)$, $\kappa(q_0)$, and $\nu_a(q_0)$ as functions that are solutions to the first-order conditions and constraints.

We next prove the "locally invariant posteriors" property described by Caplin et al. [2019]. Consider an alternative prior, $\tilde{q}_0 \in \mathcal{P}(X)$, such that

$$\tilde{q}_0 = \sum_{a \in A} \alpha(a)\hat{q}_a(q_0)$$

for some $\alpha(a) \geq 0$. Conjecture that $\hat{q}_a(\tilde{q}_0) = \alpha(a)\hat{q}_a(q_0)$, $\kappa(\tilde{q}_0) = \kappa(q_0)$, and $\nu_a(\tilde{q}_0) = \nu_a(q_0)$. By the homogeneity property,

$$H_q(\alpha(a)\hat{q}_a(q_0)) = H_q(\hat{q}_a(q_0)),$$

and therefore the first-order conditions are satisfied. By construction, the constraint is satisfied, the complementary slackness conditions are satisfied, and $\hat{q}_a$ and ν_a are weakly positive. Therefore, all necessary conditions are satisfied, and by the concavity of the problem, this is sufficient. It follows that the conjecture is verified.

Consider a perturbation

$$q_0(\epsilon; z) = q_0 + \epsilon z,$$

with $z \in \mathbb{R}^{|X|}$, such that $q_0(\epsilon; z)$ remains in $\mathcal{P}(X)$ for some $\epsilon > 0$. If z is in the span of $\hat{q}_a(q_0)$, then there exists a sufficiently small $\epsilon > 0$ such that the above conjecture applies. In this case that κ is constant, and therefore $\phi_q(q_0(\epsilon; z))$ is directionally differentiable with respect to ϵ. If $q_0(-\epsilon; z) \in \mathcal{P}(X)$ for some $\epsilon > 0$, then ϕ_q is differentiable, with

$$\phi_{qq}(q_0) \cdot z = \theta H_{qq}(q_0) \cdot z,$$

proving twice-differentiability in this direction. This perturbation exists anywhere the span of $\hat{q}_a(q_0)$ is strictly larger than the line segment connecting zero and q_0 (in other words, all $\hat{q}_a(q_0)$ are not proportional to q_0). Define this region as the continuation region, Ω. Outside of this region, all $\hat{q}_a(q_0)$ are proportional to q_0, implying that

$$\phi(q_0) = \max_{a \in A} u_a^T \cdot q_0,$$

as required for the stopping region. Within the continuation region, the strict convexity of $H(q_0)$ in all directions orthogonal to q_0 implies that, as required,

$$\phi(q_0) > \max_{a \in A} u_a^T \cdot q_0.$$

Now consider an arbitrary perturbation z such that $q_0(\epsilon; z) \in \mathbb{R}_+^{|X|}$ and $q_0(-\epsilon; z) \in \mathbb{R}_+^{|X|}$ for some $\epsilon > 0$. Observe that, by the constraint,

$$\epsilon z = \sum_{a \in A} (\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)).$$

It follows that

$$(\kappa^T(q_0(\epsilon; z)) - \kappa^T(q_0))\epsilon z = \sum_{a \in A} (\kappa^T(q_0(\epsilon; z)) - \kappa^T(q_0))(\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)).$$

By the first-order condition,

$$\begin{aligned} &(\kappa^T(q_0(\epsilon; z)) - \kappa^T(q_0))(\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)) = \\ [\theta H_q(\hat{q}_a(q_0)) - \theta H_q(\hat{q}_a(\epsilon; z)) &+ \nu_a^T(q_0(\epsilon; z)) - \nu_a^T(q_0)](\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)). \end{aligned}$$

Consider the term

$$(\nu_a^T(q_0(\epsilon; z)) - \nu_a^T(q_0))(\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)) = \sum_{x \in X} (\nu_a^T(q_0(\epsilon; z)) - \nu_a^T(q_0)) e_x e_x^T (\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)).$$

By the complementary slackness condition,

$$(\nu_a^T(q_0(\epsilon; z)) - \nu_a^T(q_0))(\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)) = -\nu_a^T(q_0(\epsilon; z))\hat{q}_a(q_0) - \nu_a^T(q_0)\hat{q}_a(\epsilon; z) \leq 0.$$

By the convexity of H,

$$\theta(H_q(\hat{q}_a(q_0)) - \theta H_q(\hat{q}_a(\epsilon; z)))(\hat{q}_a(\epsilon; z) - \hat{q}_a(q_0)) \leq 0.$$

Therefore,

$$(\kappa^T(q_0(\epsilon; z)) - \kappa^T(q_0))\epsilon z \leq 0.$$

Thus, anywhere ϕ is twice differentiable (almost everywhere on the interior of the simplex),

$$\phi_{qq}(q) \preceq \theta H_{qq}(q),$$

with equality in certain directions. Therefore, it satisfies the HJB equation almost everywhere in the continuation region. Moreover, by the convexity of ϕ,

$$(H_q(q_0(\epsilon; z)) - H_q(q_0))^T \epsilon z \geq (\phi_q(q_0(\epsilon; z)) - \phi_q(q_0))^T \epsilon z \geq 0,$$

implying that the "Hessian measure" (see Villani [2003]) associated with ϕ_{qq} has no pure point component. This implies that ϕ is continuously differentiable.

Next, we show that there is a strategy for the DM in the dynamic problem which can implement this value function. Suppose the DM starts with beliefs q_0, and generates some $\hat{q}_a(q_0)$ as described above. As shown previously, this can be mapped into a policy $\pi(a, q_0)$ and $q_a(q_0)$, with the property that

$$\sum_{a \in A} \pi(a, q_0) q_a(q_0) = q_0.$$

We will construct a policy such that, for all times t,

$$q_t = \sum_{a \in A} \pi_t(a) q_a(q_0)$$

for some $\pi_t(a) \in \mathcal{P}(A)$. Let Ω (the continuation region) be the set of q_t such that a $\pi_t \in \mathcal{P}(A)$ satisfying the above property exists and $\pi_t(a) < 1$ for all $a \in A$. The associated stopping rule will be the stop whenever $\pi_t(a) = 1$ for some $a \in A$.

For all $q_t \in \Omega$, there is a linear map from $\mathcal{P}(A)$ to Ω, which we will denote $Q(q_0)$:

$$Q(q_0)\pi_t = q_t.$$

Therefore, we must have

$$Q(q_0) d\pi_t = Diag(q_t) \sigma_t dB_t.$$

By the assumption that $|X| \geq |A|$, there exists a $|A| \times |X|$ matrix $\sigma_{\pi,t}$ such that

$$Q(q_0)\sigma_{\pi,t} = Diag(q_t)\sigma_t$$

and $d\pi_t = \sigma_{\pi,t} dB_t$. Define $\tilde{\phi}(\pi_t) = \phi(q_t)$. As shown above,

$$Q^T(q_0)\phi_{qq}(q_t)Q(q_0)$$

exists everywhere in Ω, and therefore

$$\tilde{\phi}(\pi_t) - \theta H(Q(q_0)\pi_t)$$

is a martingale. We also have to scale $\sigma_{\pi,t}$ to respect the constraint,

$$\frac{1}{2} tr[\sigma_t \sigma_t^T k(q_t)] = \chi > 0.$$

This can be rewritten as

$$\frac{1}{2} tr[\sigma_{\pi,t}\sigma_{\pi,t}^T Q^T(q_0) Diag^+(Q(q_0)\pi_t) k(Q(q_0)\pi_t)) Diag^+(Q(q_0)\pi_t) Q(q_0)] = \chi,$$

where $Diag^+$ denotes the pseudo-inverse of the diagonal matrix.

By the positive-definiteness of k in all directions except those constant in the support of $Q(q_0)\pi_t$, we will always have $\frac{1}{2}tr[\sigma_{\pi,t}\sigma_{\pi,t}^T] > 0$. Under the stopping rule described previously, the boundary will be hit a.s. as the horizon goes to infinity. As a result, by the martingale property described above, initializing $\pi_0(a) = \pi(a, q_0)$,

$$\tilde{\phi}(\pi_0) = E_0[\tilde{\phi}(\pi_\tau) - \theta H(Q(q_0)\pi_\tau) + \theta H(Q(q_0)\pi_0)].$$

By Ito's lemma,

$$\theta H(Q(q_0)\pi_\tau) - \theta H(Q(q_0)\pi_0) = \int_0^\tau \chi\theta dt = \mu\tau.$$

By the value-matching property of ϕ, $\tilde{\phi}(\pi_\tau) = \hat{u}(Q(q_0)\pi_\tau)$. It follows that, as required,

$$\phi(q_0) = \tilde{\phi}(\pi_0) = E_0[\hat{u}(q_\tau) - \mu\tau].$$

Finally, we verify that alternative policies are sub-optimal. Consider an arbitrary control process σ_t and stopping rule described by the stopping time τ. We have, by the convexity of ϕ and the generalized Ito formula for convex functions (noting that we have shown that the Hessian measure associated with ϕ_{qq} has no pure point component), interpreting ϕ_{qq} in a distributional sense,

$$E_0[\phi(q_\tau)] - \phi(q_0) = \frac{1}{2}E_0[\int_0^\tau tr[\sigma_t^T D(q_t)\phi_{qq}(q_t)D(q_t)\sigma_t]dt].$$

By the feasibility of the policies, anywhere in the continuation region of the optimal policy,

$$\frac{1}{2}tr[\sigma_t^T D(q_t)\phi_{qq}(q_t)D(q_t)\sigma_t] \leq \frac{1}{2}\theta tr[\sigma_t^T k(q_t)\sigma_t] \leq \theta\chi.$$

In the stopping region of the optimal policy,

$$\frac{1}{2}tr[\sigma_t^T D(q_t)\phi_{qq}(q_t)D(q_t)\sigma_t] = 0 < \theta\chi.$$

Therefore,

$$\phi(q_0) \geq E_0[\phi(q_\tau)] - \int_0^\tau \theta\chi dt.$$

By inequality $\phi(q_\tau) \geq \hat{u}(q_\tau)$, $\phi(q_0) \geq E_0[\hat{u}(q_\tau) - \mu\tau]$ for all policies, verifying optimality.

B.21 Proof of Corollary 2

We begin by observing that Theorem 8 characterizes the solution to the value function in this case–the proof of Theorem 8 requires only that the problem of Definition 2 be further restricted to have no jumps, not that there be a preference for gradual learning per se.

Now consider in particular utility functions with only two actions, L and R (all other action in A are dominated by those two and hence will never occur with positive probability). Using the first-order conditions for the static problem (equation (22)), we have, assuming interior solutions,

$$u_L - \theta H_q(q_L^*(q_0)) = u_R - \theta H_q(q_R^*(q_0))$$

and

$$\pi_L^*(q_0)q_L^*(q_0) + (1-\pi_L^*(q_0))q_R^*(q_0) = q_0.$$

Now pick any q_0, q_L, q_R such that $q_0 = \pi q_L + (1-\pi)q_R$ for some $\pi \in (0,1)$. Set

$$u_L = \theta H_q(q_L)) - \theta H_q(q_0) + K\iota$$

and

$$u_R = \theta H_q(q_R)) - \theta H_q(q_0) + K\iota$$

for some K such that both u_L and u_R are strictly positive. Observe that if the solution is interior, q_L, q_R, and π are optimal policies.

If the solution is not interior, stopping must be optimal. By the convexity of H,

$$q_L^T \cdot u_L - \theta H(q_L) + \theta H(q_0) + \theta(q_L - q_0)^T H_q(q_0)) - q_0^T \cdot u_L = \\ \theta(q_L - q_0)^T H_q(q_L) - \theta H(q_L) + \theta H(q_0) \geq 0,$$

and likewise for q_R. It follows that the q_0 is in the continuation region, and therefore that (q_L, q_R, π) are indeed optimal policies in the static problem.

By the "locally invariant posteriors" property described by Caplin et al. [2019], it follows that for any $q = \alpha q_L + (1-\alpha)q_R$ with $\alpha \in [0,1]$, (q_L, q_R, α) are optimal policies given initial prior q_0.

As in the proof of Theorem 8, this implies that the value function is twice-differentiable on the line segment between q_L and q_R, with

$$(q_L - q_0)^T \cdot W_{qq}(q,\lambda) \cdot (q_L - q_0) = \theta(q_L - q_0)^T \bar{k}(q)(q_L - q_0)$$

for all q on that line segment. Integrating,

$$W(q_L,\lambda) - W(q_0,\lambda) - (q_L - q_0)^T \cdot W_q(q_0,\lambda) = \\ \theta(q_L - q_0)^T \cdot (\int_0^1 (1-s)\bar{k}(sq_L + (1-s)q_0)ds) \cdot (q_L - q_0) = \\ \theta H(q_L) - \theta H(q_0) - \theta(q_L - q_0)^T H_q(q_0)).$$

By the sub-optimality of jumping directly from q_0 to q_L, it must be the case that

$$W(q_L,\lambda) - W(q_0,\lambda) - (q_L - q_0)^T \cdot W_q(q_0,\lambda) \leq \theta D^*(q_L||q_0)$$

and therefore a preference for gradual learning holds between the points q_0 and q_L.

This argument can be repeated for all (q_0, q_L) in the relative interior of the simplex. By the convexity of D^* and H, we can extend the result to the entirety of the simplex by continuity, proving that a preference for gradual learning must hold.

B.22 Proof of Lemma 7

Let $\phi(q,t,q_0)$ denote the likelihood that $q_t = q \in [q_L, q_H]$ given the initial beliefs q_0. The forward equation is

$$\phi_t(q,t;q_0) = \frac{\partial^2}{\partial q^2}[\sigma^*(q)^2\phi(q,t;q_0)].$$

By Lemma 5, the constraint binds,

$$\frac{1}{2}tr[\sigma_t^T Diag(q_t)\bar{k}(q_t)Diag(q_t)\sigma_t] = \rho^{\frac{1}{\rho}}c.$$

In the two-state model, with an alpha-divergence, $\bar{k}$ is the Fisher matrix,

$$\bar{k}(q_t) = \begin{bmatrix} \frac{1}{q_t} - 1 & -1 \\ -1 & \frac{1}{1-q_t} - 1 \end{bmatrix}$$

and assuming one-dimension of Brownian motion without loss of generality,

$$Diag(q_t)\sigma_t = \begin{bmatrix} \sigma^*(q_t) & 0 \\ -\sigma^*(q_t) & 0 \end{bmatrix}.$$

Therefore,

$$\frac{1}{2}\sigma^*(q_t)^2(\frac{1}{q_t} + \frac{1}{1-q_t}) = \rho^{\frac{1}{\rho}}c,$$

which is

$$\sigma^*(q_t)^2 = 2\rho^{\frac{1}{\rho}}cq_t(1-q_t),$$

as required.

For the conditional dynamics,

$$\begin{aligned} e_1 Diag(q_t)\sigma_t\sigma_t^T e_1 &= e_1 Diag(q_t)\sigma_t\sigma_t^T Diag(q_t)Diag(q_t)^{-1}e_1 \\ &= \frac{\sigma^*(q)^2}{q}, \end{aligned}$$

and the result follows, and likewise

$$e_1 Diag(q_t)\sigma_t\sigma_t^T e_2 = -\frac{\sigma^*(q)^2}{1-q}.$$

B.23 Proof of Lemma 8

We will show that Conditions 1-5 are satisfied. Recall the definition:

$$C(p,q;S) = \sum_{s\in S}\pi_s(p,q)D(q_s(p,q)||q).$$

B.23.1 Condition 1

Condition 1 requires that if the information structure is uninformative, the cost is zero, and if it is not, the cost is weakly positive. If the signal is uninformative, for any signal received with positive probability,

$$q_s = q,$$

and by our convention that $q_s = q$ if $\pi_s(p,q) = 0$, this also holds for zero-probability signals. By the definition of a divergence, $D(q||q) = 0$ for all q, and therefore the cost of an uninformative information structure is zero.

The cost is strictly positive by the definition of a divergence (being strictly positive if $q_s \neq q$) and the fact that probabilities must sum to one.

B.23.2 Condition 2

Mixture feasibility requires that

$$C(p_M, q; S_M) \leq \lambda C(p_1, q; S_1) + (1-\lambda)C(p_2, q; S_2).$$

By definition,

$$\begin{aligned} C(p_M, q; S) &= \sum_{s \in S} \pi_s(p_M, q) D(q_s(p_M, q)||q) \\ &= \lambda \sum_{s \in S_1} \pi_s(p_1, q) D(q_s(p_1, q)||q) + (1-\lambda) \sum_{s \in S_2} \pi_s(p_2, q) D(q_s(p_2, q)||q) \\ &= \lambda C(p_1, q; S_1) + (1-\lambda) C(p_2, q; S_2). \end{aligned}$$

verifying that the condition holds.

B.23.3 Condition 3

By Blackwell's theorem, for any Markov mapping $\Pi : S \to S'$, we require that

$$C(\Pi p, q; S') \leq C(p, q; S).$$

By definition,

$$\pi_{s'}(\Pi p, q) = \sum_{s \in S} \Pi_{s',s} \pi_s(p, q)$$

and by Bayes' rule, treating p as an $|S| \times |X|$ matrix and letting e_s denote a vector with one corresponding to s and zero otherwise,

$$D(q) p^T \Pi^T e_{s'} = \pi_{s'}(\Pi p, q) q_{s'}(\Pi p, q),$$

where $q_{s'}$ is the posterior associated with $s' \in S'$. This is

$$q_{s'}(\Pi p, q) = \frac{\sum_{s \in S} \pi_s(p,q) q_s(p,q) \Pi_{s',s}}{\pi_{s'}(\Pi p, q)}$$

It follows by the convexity of D in its first argument and Jensen's inequality that

$$\pi_{s'}(\Pi p, q) D(q_{s'}(\Pi p, q)||q) \leq \sum_{s \in S} \Pi_{s',s} \pi_s(p,q) D(q_s(p,q)||q).$$

It immediately follows that

$$\sum_{s' \in S'} \pi_{s'}(\Pi p, q) D(q_{s'}(\Pi p, q)||q) \leq \sum_{s \in S} \pi_s(p,q) D(q_s(p,q)||q).$$

B.23.4 Condition 4

We begin by showing twice-differentiability with respect to perturbations that do not change the support of the signal structure. By the definition of the cost function and the twice-differentiability of D in its first argument, it is sufficient to show that $\pi_s(p,q)$ and $q_s(p,q)$ are both twice-differentiable with respect to these perturbations, in the neighborhood of an uninformative information structure.

Suppose that

$$p(\epsilon, \nu) = r\iota^T + \epsilon\tau + \nu\omega,$$

where $r \in \mathcal{P}(S)$ and the support of τe_x is in the support of r, and likewise for ωe_x, for all $x \in X$.

By Bayes' rule, for all $s \in S$ such that $e_s^T r > 0$,

$$q_s(\epsilon, \nu) = \frac{D(q) p(\epsilon, \nu)^T e_s}{q^T p(\epsilon, \nu)^T e_s}.$$

Simplifying,

$$\begin{aligned} q_s(\epsilon, \nu) = {} & q \frac{r^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} + \frac{\epsilon D(q) \tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} \\ & + \frac{\nu D(q) \omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}. \end{aligned}$$

In the neighborhood around $\epsilon = \nu = 0$, the denominator is strictly positive, and therefore

$$\frac{\partial}{\partial \nu} q_s(\epsilon, \nu) = -q_s(\epsilon, v) \frac{q^T \omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} + \frac{D(q) \omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}$$

and

$$
\begin{aligned}
\frac{\partial}{\partial \epsilon}\frac{\partial}{\partial \nu}q_s(\epsilon,\nu) = {} & q_s(\epsilon,v)\frac{q^T\omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}\frac{q^T\tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} \\
& - \frac{q^T\omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}\frac{D(q)\tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} \\
& - q_s(\epsilon,v)\frac{q^T\omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}\frac{q^T\tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s} \\
& - \frac{D(q)\omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}\frac{q^T\tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \nu q^T \omega^T e_s}.
\end{aligned}
$$

For $s \in S$ such that $e_s^T r = 0$, $q_s(\epsilon,\nu) = q$, and therefore $\frac{\partial}{\partial \epsilon}\frac{\partial}{\partial \nu}q_s(\epsilon,\nu) = 0$. Therefore, $\frac{\partial}{\partial \nu}q_s(\epsilon,\nu)$ can be written as a quadratic form in $vec(\tau)$ and $vec(\omega)$. It follows that $q_s(\epsilon,\nu)$, in the neighborhood of an uninformative information structure, is twice-differentiable in the directions that do not change the support of the distribution of signals. By construction, $\pi_s(p,q) = (e_s^T pq)$ is twice-differentiable.

Now consider a perturbation that changes the support of the signals,

$$
p(\epsilon) = r\iota^T + \epsilon\tau + \epsilon\omega,
$$

where $e_s^T\omega = 0$ for all s such that $e_s^T r > 0$, and greater than or equal to zero otherwise, and the support of τe_x is in the support of r for all $x \in X$. We have

$$
\begin{aligned}
q_s(\epsilon) = {} & q\frac{r^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \epsilon q^T \omega^T e_s} + \frac{\epsilon D(q)\tau^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \epsilon q^T \omega^T e_s} \\
& + \frac{\epsilon D(q)\omega^T e_s}{r^T e_s + \epsilon q^T \tau^T e_s + \epsilon q^T \omega^T e_s}.
\end{aligned}
$$

For s such that $e_s^T\omega > 0$,

$$
q_s(\epsilon) = \frac{D(q)\omega^T e_s}{q^T\omega^T e_s},
$$

and hence does not depend on ϵ. We also have $(e_s^T pq) = \epsilon q^T\omega^T e_s$ for such s. Directional differentiability, continuous in (ω,τ), follows immediately.

B.23.5 Condition 5

This condition requires that, for some $m > 0$ and $B > 0$, for all $C(p,q;S) < B$,

$$
C(p,q;S) \geq \frac{m}{2}\sum_{s\in S}\pi_s(p,q)||q_s(p,q) - q||_X^2,
$$

where $||\cdot||_X$ is an arbitrary norm on the tangent space of $\mathcal{P}(X)$. It follows immediately by the strong convexity of the divergence.

B.24 Proof of Lemma 9

We will show that Conditions 1-5 are satisfied. Recall the definition:

$$C(p,q;S) = \sum_{x \in X} q_x D(p_x || \pi(p,q); S).$$

B.24.1 Condition 1

Condition 1 requires that if the information structure is uninformative, the cost is zero, and if it is not, the cost is strictly positive. If the signal is uninformative, $pe_x = pq$ for all $x \in X$, and the result holds by the definition of a divergence. The cost for informative signals is strictly positive by the definition of a divergence.

B.24.2 Condition 2

Mixture feasibility requires that

$$C(p_M, q; S_M) \leq \lambda C(p_1, q; S_1) + (1 - \lambda) C(p_2, q; S_2).$$

This follows by the convexity of the divergence, the Blackwell condition, and Lemma 1.

B.24.3 Condition 3

The result follows immediately by the Blackwell assumption on the divergence.

B.24.4 Condition 4

Twice differentiability follows by assumption. Directional differentiability, with continuous directional derivatives, follows from convexity (for the existence of directional derivatives) and twice-differentiability in the interior (which ensures continuity), and the assumption of continuity in the limit (as the signal probability reaches zero, and the signal alphabet changes).

B.24.5 Condition 5

This condition requires that, for some $m > 0$ and $B > 0$, for all $C(p,q;S) < B$,

$$C(p,q;S) \geq \frac{m}{2} \sum_{s \in S} \pi_s(p,q) || q_s(p,q) - q ||_X^2,$$

where $|| \cdot ||_X$ is an arbitrary norm on the tangent space of $\mathcal{P}(X)$.

By assumption,

$$D(r' || r; S) \geq m (r' - r)^T g(r) (r' - r),$$

where $g(r)$ is the Fisher information matrix.

Consequently,

$$\sum_{x \in X} q_x D(p_x || \pi(p,q); S) \geq m \sum_{x \in X} \sum_{s \in S} \frac{q_x}{\pi_s(p,q)} (e_x^T - q^T) p^T e_s e_s^T p (q - e_x),$$

which by Bayes' rule is

$$\sum_{x \in X} (e_x^T q) D(pe_x || pq; S) \geq m \sum_{s \in S} \pi_s(p,q) (q_s^T(p,q) - q^T) g(q) (q_s(p,q) - q).$$

Therefore, by $g(q) \succeq I$,

$$\sum_{x \in X} q_x D(p_x || \pi(p,q); S) \geq m \sum_{s \in S} \pi_s(p,q) || q_s(p,q) - q ||_2^2$$

where $|| \cdot ||_2$ denotes the Euclidean norm. The result follows by the equivalence of norms.