Rational Inattention with Sequential Information Sampling*

Benjamin Hébert †
Stanford University

Michael Woodford ‡
Columbia University

May 14, 2016

Preliminary and Incomplete. Comments welcome.

Abstract

We generalize the rationalize inattention framework proposed by Sims [2010] to allow for cost functions other than Shannon's mutual information. We study a problem in which a large number of successive samples of information about the decision situation, each only minimally informative by itself, can be taken before a decision must be made. We assume that the cost required for each individual increment of information satisfies standard assumptions about continuity, differentiability, convexity, and monotonicity with respect to the Blackwell ordering of experiments, but need not correspond to mutual information. We show that any information cost function satisfying these axioms exhibits an invariance property which allows us to approximate it by a particular form of Taylor expansion, which is accurate whenever the signals acquired by an agent are not very informative. We give particular attention to the case in which the cost function for an individual increment to information satisfies the additional condition of "prior invariance," so that it depends only on the way in which the probabilities of different observations depend on the state of the world, and not on the prior probabilities of different states. This case leads to a cost function for the overall quantity of information gathered that is not prior-invariant, but "sequentially prior-invariant." We offer a complete characterization of the family of sequentially prior-invariant cost functions in the continuous-time limit of our dynamic model of information sampling, and show how the resulting theory of rationally inattentive choice differs from both static and dynamic versions of a theory based on mutual information.

*The authors would like to thank Mira Frick, Matthew Gentzkow, Emir Kamenica, Divya Kirti, and Ran Shorrer for helpful discussions on this topic. Portions of this paper 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

The theory of rational inattention, proposed by Christopher Sims and surveyed in Sims [2010], seeks to endogenize the imperfect awareness that decision makers have about the circumstances under which they must choose their actions. According to the theory, a decision maker (DM) can choose an arbitrary signal structure with which to gather information about her situation before making a decision; the cost of each possible signal structure is described by a cost function, which in Sims' theory is an increasing function of the Shannon mutual information between the state of the world (that determines the DM's reward from choosing different actions) and the DM's signal.

Shannon's mutual information, a measure of the degree to which the signal is informative about the state of the world, plays a central role in information theory (Cover and Thomas [2012]), as a consequence of powerful mathematical results that are considerable practical relevance in communications engineering. It not obvious, though, that the theorems that justify the use of mutual information in communications engineering provide any warrant for using it as a cost function in a theory of attention allocation, in the case of either economic decisions or perceptual judgments.[1] Moreover, the mutual information cost function has implications that are unappealing on their face, and that seem inconsistent with evidence on the nature of sensory processing, as discussed in Woodford [2012].

For example, the mutual information cost function imposes a type of symmetry across different states of nature, so that it is equally easy or difficult to learn about any two states that occur with the same probability. In the case of an experimental task of the kind discussed in Caplin and Dean [2015] — in which subjects are presented with an array of 100

[1] As explained in (Cover and Thomas [2012]), these theorems rely upon the possibility of "block coding" of a large number of independent instances of a given type of message, that can be jointly transmitted before any of the messages have to be decodable by the recipient. In the kind of situations with which we are concerned here, instead, there is a constraint on the informativeness of each individual transmission, that can be measured independently of the signals that are sent earlier and later.

red and blue balls, and must determine whether there are more red balls or more blue — Sims' theory of rational inattention implies that given that the reward from any action (e.g., declaring that there are more red balls) is the same for all states with the property that there are more red balls than blue, the probability of a subject's choosing that response will be the same in each of those states. In fact, it is much easier to quickly and reliably determine that there are more red balls in the case of some arrays in this class (for example, one with 98 red balls and only two blue balls) than others (for example, one with 51 red balls and 49 blue balls, relatively evenly dispersed), and subjects make more payoff-maximizing responses in the former case.

As we review below, the mutual information cost function also implies that the cost of performing an experiment– receiving a signal whose distribution of outcomes is conditional on the state of the world– depends on the prior probabilities that different states of the world will be encountered. As a consequence, the mutual information cost function implies that it is not costly to arrange to receive a signal that can take a very large number of different values, allowing fine differentiation between many different possible states, as long as most of the states that can be distinguished from one another are unlikely to occur. We suppose instead that the cost of performing a given experiment should be the same, regardless of whether it is performed under circumstances in which it is expected to be highly informative or under circumstances where the outcome is a forgone conclusion; it is exactly for this reason that the observations that will be worth making should vary depending on circumstances. In assuming that the cost function for individual observations is prior-invariant, we follow the standard literature on the statistical theory of optimal experimentation, originated by authors such as Wald [1947].

One response to these unappealing features of the mutual information cost function is to develop a theory of rational inattention that makes only much weaker assumptions about the cost function (consistent with mutual information, but not requiring it), as authors such as

De Oliveira et al. [2013] and Caplin and Dean [2015] have done. This approach results in a theory with correspondingly weaker predictions. We instead propose a specific parametric family of information-cost functions, that uniquely possess certain desirable properties, while avoiding the disadvantages of mutual information just mentioned.

Our approach exploits the special structure implied by an assumption that information sampling occurs through a sequential process, in which each additional signal that is received determines whether additional information will be sampled, and if so, the kind of experiment to be performed next. We particularly emphasize the limiting case in which each individual experiment is only minimally informative, but a very large number of independent experiments can be performed within a given time interval. In the continuous-time limit of our sequential sampling process, we obtain strong and relatively simple characterizations of the implications of rational inattention, owing to the fact that only local properties of the assumed cost function for individual experiments matter in this case, and that (under relatively weak general assumptions about the cost of more informative experiments) these local properties can be summarized by a finite number of parameters.

In assuming a sequential sampling process in which only a very small amount of information arrives in any short time interval, we follow studies such as that of Fudenberg et al. [2015], which considers the optimal stopping problem for an information sampling process described by the sample path of a Brownian motion with a drift that depends on the unknown state of the world. This can be thought of as a problem in which a given experiment (which produces an outcome given by a single real number, with the probability of different outcomes dependent on the state of the world) can be repeated an indefinite number of times, with a fixed cost per repetition of the experiment (independent of what one may believe at that time about the probability of different states). The sequence of outcomes of the successive experiments becomes a Brownian motion in the limiting case in which individual experiments require only an infinitesimal amount of time (and hence involve only an infinitesimal cost, as a fixed cost of sampling per unit time is assumed), and are correspondingly minimal in the information that they reveal about the state (because the difference in the mean outcome of an individual experiment across different states of the world is tiny relative to the standard deviation of the outcome). We consider a sequential sampling process similar to that assumed by Fudenberg et al. [2015], except that instead of assuming that there is only one specific experiment that can be repeated an arbitrary number of times, we assume that in each tiny interval of time any of a wide range of experiments can be performed, with a different cost for each of the possible experiments, in the spirit of the rational inattention literature. We can then consider not only when it is optimal to terminate information sampling and make a decision, but also the optimal kind of information to sample at each stage of the sequential process, given the DM's objective and prior beliefs.

We believe that it is often quite realistic to assume that information is acquired through a sequential sampling process. As discussed in Fehr and Rangel [2011] and Woodford [2014], an extensive literature in psychology and neuroscience has argued that data on both the frequency of perceptual errors and the frequency distribution of response times can be explained by models of perceptual classification based on sequential sampling, and more recently, some authors have proposed that data on stochastic choice and response time in economic contexts can be similarly modeled. In the case of preferential choice between goods, the sampling is presumably a process of drawing from memory associations (or recollections of past experiences) that bear upon the assignment of value to the presented options, rather than a sequence of repeated observations of the item presented.

A simple and widely used model of this kind is the "drift-diffusion model" (DDM),[2] which postulates that an (unobserved) "relative decision value" (RDV) evolves according

[2] In addition to the references in Fehr and Rangel [2011], more recent examples of the application of this model to preferential choice experiments include Clithero and Rangel [2013] and Krajbich et al. [2014].

to a Brownian motion with a drift proportional to the true difference in value of the two options, and that the DM makes a decision in favor of one option or the other when this variable first crosses either an upper or lower threshold. Models of this kind were first considered by mathematical psychologists because the model was understood to correspond to an optimal Bayesian decision rule (taking the successive increments of the RDV to represent additional sample evidence) when there is a cost per unit time of additional sampling, in the case of a problem in which it is known *a priori* that only two possible states (i.e., two different relative values for the two options) are possible; in this case, the RDV tracks the posterior log odds of the two possible states, and provides a sufficient statistic for the decision whether to continue sampling additional information. The model is frequently fit to experimental data, however, in cases in which there are many more than two possible states (even among the cases encountered on separate experimental trials), so that the degree to which such a decision algorithm can be viewed as desirable remains an open question.

Fudenberg et al. [2015] consider an optimal decision problem with sequential sampling of evidence about the relative value of the two goods, in which the random evidence is of the kind assumed in the DDM: the new information over any time interval is given by the further evolution of a variable that follows a Brownian motion with a constant drift proportional to the true relative value. There is no choice about the kind of information that can be sampled, only about how long to continue sampling additional evidence; but Fudenberg *et al.* consider the optimal stopping rule rather than assuming arbitrary time-invariant thresholds, as in the DDM, and show that a time-invariant threshold is generally not optimal when there is a large number of possible states. Woodford [2014] instead takes as given that the DM's belief state can be summarized by a single real variable (that can be thought of as an RDV) that must evolve with a continuous sample path, and that a decision is made only when this variable reaches either an upper or lower threshold, as in the DDM. However, rather than assuming a particular stochastic evolution of this variable in the case of each possible true relative value, Woodford [2014] considers which evolution process would be optimal, in the sense of maximizing the expected value of the DM's eventual decision, net of a measure of the cost of information used in controlling the evolution of the belief state; and finds that while it is optimal for the drift of the belief state to be an increasing function of the true relative value, as assumed by the DDM, it is generally optimal for the drift also to depend on the current belief state. This corresponds to choosing to sample information of a different type depending on the belief state that has been reached as a result of the information collected to that point.

In the sequential information sampling problem considered here, we allow the information sampled at each stage to be chosen very flexibly, as in Woodford [2014], subject only to an information cost function; but we also allow the decision when to stop sampling and make a decision to be made optimally, on the basis of the entire history of information sampled to that point, as in Fudenberg et al. [2015]. In this way, the complete information sampling and decision procedure is endogenized, given our assumptions about the costs of alternative information sampling strategies. Interestingly, we find under relatively general assumptions that in the case of a binary decision, the evolution of the DM's belief state corresponds to a diffusion along a one-dimensional line segment, with a decision being made when either of the two endpoints (the locations of which are independent of time) is reached, as postulated by the DDM. Fudenberg *et al.* find that this is *not* the case, when there are more than two possible states, even when the information sampling process is exactly the one assumed in the DDM; and this might suggest that the conditions required for optimal belief dynamics to take such a form are very special. Instead we find that the result of Fudenberg *et al.* depends on their assumption that only a very restrictive form of information sampling is possible. Our more flexible assumptions about information sampling make it *possible* for the DM's posterior to evolve in many more ways than are allowed by the process assumed by Fudenberg *et al.*, but in our case the way in which it is *optimal* for the DM's posterior to evolve is actually restricted to a lower-dimensional subset of the probability simplex than in their case, allowing a simpler characterization of the optimal stopping rule.

In addition to characterizing optimal sequential information sampling in our dynamic model, we show that the model's predictions with regard to choice frequencies conditional on the state of the world are the same as those of a static rational-inattention model with an appropriately chosen information-cost function for the choice of a single signal. (The finite set of possible signal values in the equivalent static model correspond to the different possible *terminal* information states in the dynamic model.) We show how to derive the static information-cost function corresponding to a given flow information-cost function in the dynamic model. While we also consider this problem in the more general case of a flow information-cost function that may depend on the prior (as in the case of the flow information cost based on mutual information assumed in Woodford [2014]), we are primarily interested in the case in which the flow information-cost function is prior-invariant, for the reasons mentioned above. We call this the case of *sequential prior-invariance.*

Even in this case, the equivalent static information-cost function is *not* prior-invariant. It is true that a given information-sampling strategy (specifying whether to continue sampling, and if so which experiment to perform next, as a function of the information received to that point) will result in a probability distribution of possible final information states (described by the relative likelihoods of the different possible states of nature, given the evidence), conditional on the true state, that is independent of the prior; and given a prior-invariant flow information-cost function, the expected cumulative information costs associated with the strategy, conditional on the true state, will also be independent of the prior. But these costs will generally differ depending on the true state (for example, the probability of a sequence of outcomes that leads to a longer period of additional sampling before decision will be greater in some states than in others); hence the *expected infor-*

mation costs associated with the strategy will generally depend on the prior. Moreover, in the dynamic model, it is not generally the case that a given specification of the probabilities of reaching different final information states conditional on the true state can only be achieved using a single dynamic information-sampling strategy; and which among the alternative feasible strategies that reach the same final information states with the same probabilities will have the lowest expected information cost will generally depend on the prior. For this reason, even when the flow information-cost function is prior-invariant, the equivalent static information-cost function is not, though it can depend on the prior only in a very specific way that we explain.

In fact, we further show that the complete family of sequentially prior-invariant static information-cost functions can be fully specified by the values of a finite matrix of coefficients that we call the "information-cost matrix." Different specifications of the information-cost matrix correspond to different assumptions about the degree to which it is costly to differentiate among different ones of the possible states of the world. The fact that it is possible to assume different matrices means that our theory does not make such sharp predictions as that of Sims, which has only a single information-cost parameter; but the greater flexibility of our theory in this particular respect is important, as we regard it as incorrect to assume in all cases that each of the possible states of the world is equally easy or difficult to distinguish from each of the others. Subject to the proviso that we allow flexibility in specifying which states are easy or difficult to tell apart, our theory still offers an extremely strong characterization of the form of the static information-cost function, and hence much more specific predictions than the generalized theories of rational inattention proposed by authors such as De Oliveira et al. [2013] and Caplin and Dean [2015].

The class of static rational-inattention problems defined by these information-cost functions are relatively straightforward to solve, and we compare the predictions of our model to those of the standard (static) rational inattention model that assumes a cost function based on mutual information. Properties such as the invariant likelihood ratio property and the locally invariant posteriors property described in Caplin and Dean [2013]) are not general implications of our alternative model. At the same time, it should in principle be possible to identify the information-cost matrix in experimental settings, as in Caplin and Dean [2015], and test the very precise predictions of the theory of stochastic choice implied by our model. By identifying the static rational-inattention problem with a particular kind of dynamic information-sampling problem, we obtain additional testable predictions, if we suppose, as in studies such as Clithero and Rangel [2013] and Krajbich et al. [2014], that the observable data include decision times as well as the frequencies with which different decisions are made.

Our paper obviously builds upon the rational inattention literature, surveyed in Sims [2010]. In its use of axioms to characterize the assumed form of the flow information-cost function, it is particularly close to Caplin and Dean [2013], Caplin and Dean [2015], and De Oliveira et al. [2013]. The Chentsov [1982] theorems used to characterize the properties of general rational inattention cost functions were also used by Hébert [2014], in a different context. We also use techniques developed by Kamenica and Gentzkow [2011] and Matejka and McKay [2011] in characterizing the solution to our problem.

Section 2 discusses general properties of information-cost functions, introduces some terminology, and summarizes our main results. Section 3 discusses our technical assumptions regarding the nature of flow information costs, and presents a key theorem, characterizing the local structure of flow information-cost functions. Section 4 begins our analysis of the consequences of general information-cost functions by solving a static rational-inattention problem. Section 5 then discusses the dynamic information-sampling problem, which is a repeated version of the static rational-inattention model, and shows that it is equivalent to a static model with a particular information-cost function. Section 6 concludes.

2 A New Cost Function for Rational Inattention Problems

We begin by describing the rational inattention framework of Sims [2010]. Let $x \in X$ be the underlying state of the nature, and $a \in A$ be the action taken by the agent. The agent's utility from taking action a in state x is $u(x,a)$. However, the agent does not perfectly observe the state x. Instead, the agent will receive a signal, $s \in S$, which can convey information about the state. For simplicity, we assume that X, A, and S are finite sets. We define $\mathscr{P}(\Omega)$ as the probability simplex over finite set Ω.

Let $q(x)$ denote the agent's prior belief (before receiving a signal) about the probability of state x. Define $p(s|x)$ as the probability of receiving signal s in state x. The set of conditional probability distributions $\{p(s|x)\}$ define a "signal structure." After receiving signal s, the agent will hold a posterior, $q(x|s)$, consistent with Bayes' rule:

$$q(x|s) = \frac{q(x)p(s|x)}{\sum_{x' \in X} q(x')p(s|x')}.$$

The agent will take an action consistent with this posterior:

$$a^*(s;\{p(\cdot|\cdot)\}) \in \arg\max_a \sum_x q(x|s)u(x,a). \tag{1}$$

The rational inattention problem maximizes the expected utility of the agent over possible signal structures, taking into account the cost of each signal structure. We take the alphabet of signals, S, as given.[3] When the cost of a particular signal structure is proportional to mutual information, with constant of proportionality $\theta > 0$, the agent solves

$$\max_{\{p(\cdot|x) \in \mathscr{P}(S)\}_{x \in X}} \sum_{x \in X} q(x) \sum_{s \in S} p(s|x)u(x,a^*(s;\{p(\cdot|\cdot)\})) - \theta I(\{p(\cdot|\cdot)\},q). \tag{2}$$

[3] A standard result in this literature is that it is without loss of generality to define S as equal to the set of possible actions, A.

Here, $I(\cdot)$ denotes the mutual information between the signal and the state.

There are several equivalent definitions of mutual information. For our purposes, the most convenient formulation uses the Kullback-Leibler divergence, also known as relative entropy. The KL divergence between a probability distribution p and another probability distribution p' is defined as

$$D_{KL}(p||p') = \sum_{s\in S} p(s)\ln(\frac{p(s)}{p'(s)}).$$

The mutual information between the signal s and the state x can be defined as

$$I(\{p(\cdot|\cdot)\},q) = \sum_{s\in S}(\sum_{x'\in X} q(x')p(s|x'))D_{KL}(q(\cdot|s)||q(\cdot)).$$

Mutual information has several notable properties. First, the cost of a particular signal structure $\{p(\cdot|\cdot)\}$ depends on the prior. Two rationally inattentive agents with different priors about the state would generically face different costs when observing the same signal structure. Second, the cost does not depend on the economic meaning of the states X, only the cardinality of the set. If two states x and x' are equally likely under the prior, swapping the conditional distribution of signals $p(\cdot|x)$ and $p(\cdot|x')$ does not change the mutual information.

More generally, an information cost function is a mapping from a signal structure $\{p(\cdot|\cdot)\}$ and prior q to a cost,

$$C : \mathscr{P}(S)^{|X|} \times \mathscr{P}(X) \to \mathbb{R}.$$

In this paper, we will discuss several classes of cost functions.

One class of cost functions of interest is the "prior-invariant" cost functions. A cost function is "prior-invariant" if it depends only on the signal structure, and not on the prior.

In the standard rational inattention problem, with the mutual information cost function, the cost function depends on the prior. But as noted by Kamenica and Gentzkow [2011], any particular information-acquisition technology (specifying the procedures to be performed) should have a cost of use that is independent of one's prior, and indeed it is standard in the literature on optimal experimentation in statistical decision theory to assume that the cost of an experiment is independent of one's prior at the time that one does or does not choose to undertake it. Of course even under this assumption, two DMs with different priors might choose to acquire different information, and pay different amounts for that information.

A second class of cost functions of interest is that of "posterior-separable" cost functions.[4] These cost functions are based on divergences. A divergence is a function $D : \mathscr{P}(X) \times \mathscr{P}(X) \to \mathbb{R}^+$ such that $D(p||q) = 0$ if and only if $p = q$, and $D(p||q) > 0$ otherwise. One can think of a divergence as a distance between two probability distributions, except that it does not need to be symmetric and does not necessarily obey the triangle inequality. A posterior-separable cost function can be written as

$$C_D(\{p(\cdot|\cdot)\}, q) = \sum_{s \in S} (\sum_{x \in X} p(s|x)q(x)) D(q(\cdot|s)||q). \tag{3}$$

Mutual information is one example of a posterior-separable cost function; the associated divergence is the Kullback-Leibler divergence. The rational inattention problem with a posterior-separable cost function can be analyzed using the "concavification" techniques described by Kamenica and Gentzkow [2011] and Caplin and Dean [2013].

We argue that a certain type of posterior-separable cost function should be of particular

[4]The term "posterior-separable" is taken from Caplin and Dean [2013]. However, we use it for a larger class of cost functions; Caplin and Dean [2013] apply it only to the case in which the divergence in question is a Bregman divergence.

interest. We define the "sequentially prior-invariant" class of cost functions as

$$C_{I,k}(\{p(\cdot|\cdot)\},q) = \sum_{s\in S}(\sum_{x'\in X} q(x')p(s|x'))D_{I,k}(q(\cdot|s)||q(\cdot)),$$

where $D_{I,k}$ is a divergence defined by a matrix k:

$$D_{I,k}(p||q) \quad = \quad \sum_{x\in X}\sum_{x'\in X} k(x,x')f(\frac{p(x)}{q(x)},\frac{p(x')}{q(x')}), \tag{4}$$

with

$$f(u,u') = \begin{cases} \frac{(u'-1)u\ln u-(u-1)u'\ln u'}{u-u'} & \text{if } u \neq u' \\ u-1-\ln u & \text{if } u = u' \end{cases}.$$

The essence of the paper is explaining why we believe this class of cost functions is often an appropriate one.

Our argument relies upon an analysis of optimal sequential information sampling in a dynamic model. We show that the cumulative information acquired in the dynamic model is the same as in a static rational-inattention model with a particular specification of the information-cost function for the static problem. In the static problem, we suppose that the DM simply arranges to acquire a signal that will, depending on its realization, take him to one or another of the possible final belief states under the optimal information-sampling policy in the dynamic model, and the cost of such a signal is assumed to depend on what the set of possible final belief states are, and the probability of reaching each of them conditional on the true state of the world. We further show that if the flow information-cost function in the dynamic model is any *prior-invariant* function satisfying certain additional regularity properties, the information-cost function of the equivalent static problem is a member of the *sequentially prior-invariant* class just defined, where the matrix of coefficients $k(x,x')$ depends on local properties of the flow information-cost function in the dynamic model.

It may seem surprising that the cost function of the equivalent static problem is no longer prior-invariant, even when the flow cost function is. The reason is that in the dynamic model, an optimal information-sampling strategy will generally involve collecting different amounts of information before a decision is made, depending on the signals received from earlier experiments; and the probability of receiving the sort of signals that result in more information being collected will differ depending on the state of the world. Consider any strategy that specifies after any possible partial history of signals whether the belief state thus reached belongs to the set of terminal states (so that sampling will cease and an action will be taken), and if not, which of the set of feasible experiments should be conducted next. In the case of prior-invariant flow information-cost function, there will be an expected cumulative information cost $c(x)$ before a decision is reached in the case of state x, that is independent of the prior probabilities assigned to different states. But $c(x)$ will generally depend on the true state of the world x, and hence the expected cost of choosing this strategy, $\sum_x q(x)c(x)$, will depend on the prior.

In fact, the class of sequentially prior-invariant cost functions can alternatively be written in the form

$$C_{I,k}(\{p(\cdot|\cdot)\},q) \;=\; \min_{\hat{p}} \sum_x q(x) c_x(\{p(\cdot|\cdot)\};\hat{p}), \tag{5}$$

where for each state x we define

$$c_x(\{p(\cdot|\cdot)\};\hat{p}) \;=\; \sum_{x',x''} k(x',x'') \sum_s \left[p(s|x) g\left(\frac{p(s|x')}{\hat{p}(s)}, \frac{p(s|x'')}{\hat{p}(s)}\right) + \hat{p}(s) h\left(\frac{p(s|x')}{\hat{p}(s)}, \frac{p(s|x'')}{\hat{p}(s)}\right) \right],$$

and the functions g and h are defined by

$$g(u,u') \equiv f(u,u') - u f_1(u,u') - u' f_2(u,u'),$$

$$h(u,u') \equiv u f_1(u,u') + u' f_2(u,u'),$$

and $f(u,u')$ is the function defined above. Here each choice of $\hat{p}$ (a probability measure over the set of final belief states s) indexes a possible information-sampling strategy consistent with the specification $\{p(\cdot|\cdot)\}$ of the conditional probabilities of reaching the alternative possible final belief states. This set of possible strategies is independent of the prior, and the state-contingent expected information cost $c_x(\{p(\cdot|\cdot)\};\hat{p})$ of using such a strategy is independent of the prior as well. However, the minimum expected cost for a strategy that is consistent with final belief probabilities $\{p(\cdot|\cdot)\}$, given by (5), depends on the prior.

The characterization (5) is useful because it gives an expression c_x for the expected cumulative information costs conditional on each possible state of the world x. We further show below, in our characterization of optimal sequential information sampling, that these cumulative flow information costs are always proportional to the length of time that sampling continues before a decision. Hence in the dynamic model with prior-invariant flow information costs, we have a precise prediction not only for the probabilities of choosing different actions conditional on the state of the world, but also for the mean time that should be required for a decision conditional on the state. This provides a further set of testable implications of the model, as illustrated in Woodford [2014].

The sequentially prior-invariant cost functions depend on a matrix, $k(x,x')$, that we call the "information-cost matrix." This matrix describes the relative difficulty of learning about different states of the world. In many economic applications, there is an ordering of the states of the world. As an example, suppose the states of the world represent possible returns of the stock market, relative to a bank account, and the decision problem is a portfolio choice problem about how much to invest in the stock market. In this case, it seems natural to assume that, when the agent learns about the likelihood of stock returns being 10%, she also learns some information about the likelihood of stock returns being 9% or 11%, but does not learn much about the likelihood of stock returns being 50% or -50%. Put another way, there are complementarities with respect to learning about the 10% stock market return state and the 9% state. In the matrix $k(x,x')$ that we describe, these complementarities are represented by negative off-diagonal elements. Positive off-diagonal elements would represent substitutabilities. These notions of complementarity and substitutability are not present in the standard rational inattention cost function (mutual information).

Additionally, there may be some states that are simply easier or harder to learn about. Continuing with the stock market example, it may be easy to acquire information about the relative likelihood of a 10% stock return and a 5% stock return, but very difficult to acquire information about the relative likelihood of a 10% stock return and a -50% stock return. In the matrix $k(x,x')$, the relative difficult of learning about a particular state $x \in X$ is represented by the value of $k(x,x)$. Again, the idea that certain states can be easier or harder to learn about is not present in the standard rational inattention cost function.

Below, we show an example of the matrix $k(x,x')$. For reasons that will be explained in the next section, for all matrices $k(x,x')$, each row and column of the matrix sums to zero. This example matrix embeds some notion of "distance" between different states of the world:

$$k(x,x') = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 & \dots & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & \ddots & \vdots \\ 0 & -\frac{1}{2} & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 & -\frac{1}{2} \\ 0 & \dots & 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Each adjacent state in the matrix is "close," in there sense that there are complementarities (negative off-diagonal elements) when learning about whether this state occurs and whether another close state has occurred. Except for the "tail" states, this matrix does not exhibit smaller or larger costs of learning about particular states, although it is easy to construct examples that do exhibit this feature.

In the next section, we discuss the conditions we impose on the (static or "flow") information-cost functions that we study.

3 Information Costs

In this section, we discuss a large class of information costs, which include mutual information. These information costs are characterized by a set of conditions that are relevant or useful in economic applications. We show that all of the information costs in this class are approximately the same, in a sense, for signals that convey very little information. In the subsequent sections, we will analyze dynamic problems, in which very little information is gathered each period. We will rely on these approximations to generate predictions, regardless of which information cost from this class is relevant.

We assume four conditions that characterize the family of information cost functions we consider. All of these conditions are satisfied by mutual information, but also by many other cost functions. The first three conditions characterize "canonical" rational inattention cost functions, in the terminology of De Oliveira et al. [2013].

Condition 1. Signal structures that convey no information ($p(s|x) = p(s|x')$ for all $s \in S$ and $x, x' \in X$) have zero cost. All other signal structures have a cost greater than zero.

This condition ensures that the least costly strategy for the agent is to acquire no information, and make her decision based on the prior. The requirement that gathering no information has zero utility cost is a normalization.

Condition 2. The cost of information is convex in $p(\cdot|x)$ for each $x \in X$. That is, for all $\lambda \in (0,1)$ and all $q \in \mathscr{P}(X)$,

$$C(\{\lambda p_1(\cdot|\cdot) + (1-\lambda)p_2(\cdot|\cdot)\},q) \leq \lambda C(\{p_1(\cdot|\cdot)\},q) + (1-\lambda)C(\{p_2(\cdot|\cdot)\},q).$$

This condition is useful as an economic assumption because it encourages the agent to minimize the number of distinct signals employed. It also helps ensure that there is a unique signal structure that solves the rational inattention problem.

The next condition uses Blackwell's ordering. Consider two signal structures, $\{p_1(\cdot|\cdot)\}$, with signal alphabet S, and $\{p_2(\cdot|\cdot)\}$, with alphabet S'. The first signal structure Blackwell dominates the second signal structure if, for all utility functions $u(a,x)$ and all priors $q(x)$,

$$\sup_{a(s)} \sum_{x\in X} \sum_{s\in S} q(x)p_1(s|x)u(a(s),x) \geq \sup_{a(s')} \sum_{x\in X} \sum_{s'\in S'} q(x)p_2(s'|x)u(a(s'),x).$$

If signal structure $\{p_1(\cdot|\cdot)\}$ Blackwell dominates $\{p_2(\cdot|\cdot)\}$, 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 signal structures do not dominate each other in this sense. However, when a signal structure does Blackwell dominate another signal structure, we assume the dominant signal structure is more costly.

Condition 3. If the signal structure $\{p_1(\cdot|\cdot)\}$ with signal alphabet S is more informative, in the Blackwell sense, than $\{p_2(\cdot|\cdot)\}$, with signal alphabet S', then, for all $q \in \mathscr{P}(X)$,

$$C(\{p_1(\cdot|\cdot)\},q) \geq C(\{p_2(\cdot|\cdot)\},q).$$

The fourth condition we assume, which is not imposed by De Oliveira et al. [2013], is a differentiability condition that will allow us to characterize the local properties of our cost functions.

Condition 4. The information cost function is continuously twice-differentiable in $p(\cdot|x)$, for each $x \in X$ and $q \in \mathscr{P}(X)$, in the neighborhood of an uninformative signal structure.

From these four conditions, we derive a result about the second-order properties of the cost function. The first three conditions are innocuous, in the sense that, for any stochastic choice data, there is a cost function satisfying those properties that is the consistent with conditions 1-3 (theorem 2 of Caplin and Dean [2015]). Condition 4 is not completely general; for example, it rules out the case in which the agent is constrained to use only signals in a parametric family of probability distributions, and the cost of other signal distributions is infinite. Condition 4 also rules out other proposed alternatives, such as the channel capacity constraint suggested by Woodford [2012].[5] Mutual information, as mentioned above, satisfies each of these four conditions. However, it is not the only cost function to do so. To introduce these other cost functions, it is first useful to recall Blackwell's theorem.

Theorem 1. *(Blackwell [1953]) If, and only if, the signal structure $\{p_1(\cdot|\cdot)\}$, with signal alphabet S, is more informative, in the Blackwell sense, than $\{p_2(\cdot|\cdot)\}$, with signal alphabet S', then there exists a Markov transition matrix $\Pi : S \to S'$ such that, for all $s' \in S'$ and $x \in X$,*

$$p_2(s'|x) = \sum_{s \in S} \Pi(s', s) p_1(s|x). \tag{6}$$

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 really garble the signals. These garbling matrices have left inverses that are also Markov transition matrices. If we define a signal structure $\{p(\cdot|\cdot)\}$, with signal alphabet S, and another signal structure $\{p'(\cdot|\cdot)\}$, with signal alphabet S', using one of these matrices, via equation (6), then $\{p(\cdot|\cdot)\}$ is more informative than $\{p'(\cdot|\cdot)\}$, but $\{p'(\cdot|\cdot)\}$ is also more informative than $\{p(\cdot|\cdot)\}$. These two

[5] We speculate that it may be possible to apply our methods to generalized versions of the channel capacity.

signal structures are called "Blackwell-equivalent," and it follows that the cost of these two signal structures must be equal, by condition 3.

The left-invertible Markov transition matrices associated with Blackwell-equivalent signal 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). The KL divergence, used earlier to define mutual information, is invariant. Let Π be a Markov congruent embedding from S to S'. For any probability distributions p and r over $\mathscr{P}(S)$,

$$D_{KL}(p||r) = D_{KL}(\Pi p||\Pi r).$$

There are many other invariant divergences. Let $D(\cdot)$ be an arbitrary invariant divergence that is convex in its arguments. We can define an alternative version of mutual information,

$$I_D(\{p(\cdot|\cdot)\}, q) = \sum_{x\in X} q(x) D(p(\cdot|x)|| \sum_{x'\in X} q(x')p(\cdot|x')).$$

This alternative version satisfies conditions 1-3 above, and is therefore also a "canonical" rational inattention cost function. More generally, there are cost functions that satisfy the conditions above, but are not based on invariant divergences. Examples of such costs functions have been proposed, for the purpose of studying rational inattention problems, by Caplin and Dean [2013].

We proceed by considering properties that are common to all of these information costs. By condition 3, all canonical information costs functions are invariant to Markov congruent embeddings. Let Π be a Markov congruent embedding. It necessarily follows that

$$C(\{p(\cdot|\cdot)\}, q) = C(\{\Pi p(\cdot|\cdot)\}, q).$$

The argument is that the signal structure $\{p(\cdot|\cdot)\}$ is more informative than $\{\Pi p(\cdot|\cdot)\}$, but the reverse is also true, due to the existence of a Markov left inverse for Π.

All convex, invariant divergences with continuous Hessian matrices have the property that, at the point $p = r$, their Hessian matrix is proportional to the Fisher information matrix, with some constant of proportionality $c > 0$ (Chentsov [1982]). The KL divergence has this property, and it can be interpreted as a statement about the costs of a small amount of information. In states x and x', if the conditional distribution of signals is close to the unconditional distribution of signals, very little information has been gained about the relative likelihood of x versus x'.

This property of invariant divergences is a corollary of Chentsov's more general results. Chentsov proved the following results:[6]

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.

These results will allow us to characterize the local properties of rational inattention cost functions, via a Taylor expansion. Consider a signal structure $p(s|x;\varepsilon,\nu) = r(s) + \varepsilon\tau(s|x) + \nu\omega(s|x)$. Here, $\tau(s|x)$ and $\omega(s|x)$ represent informative signal structures such that $\tau(s|x) \neq 0$ and $\omega(s|x) \neq 0$ only if $r(s) > 0$, and ε and ν are perturbations in those directions, away from an uninformative signal structure. By condition 1, $C(\{p(s|x;0,0)\};q) = 0$. The first order term is

[6]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.

$$\frac{\partial}{\partial \varepsilon} C(\{p(\cdot|\cdot;\varepsilon,\nu)\};q)|_{\varepsilon=\nu=0} = \sum_{x\in X} C_x(\{p(\cdot|\cdot;0,0)\};q)\cdot\tau(\cdot|x),$$

where C_x denotes the derivative with respect to $p(\cdot|x)$. $C_x(\{p(\cdot|\cdot;0,0)\};q)$, for $r \in \mathscr{P}(S)$, forms a continuous 1-form tensor field over the probability simplex $\mathscr{P}(S)$. By the invariance of $C(\cdot)$, it also follows that C_x is invariant, meaning that

$$C_x(\{p(\cdot|\cdot;0,0)\};q)\cdot\tau(\cdot|x) = C_x(\{\Pi p(\cdot|\cdot;0,0)\};q)\cdot\Pi\tau(\cdot|x).$$

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

$$\frac{\partial}{\partial \nu}\frac{\partial}{\partial \varepsilon} C(\{p(\cdot|\cdot;\varepsilon,\nu)\},q)|_{\varepsilon=\nu=0} = \sum_{x'\in X}\sum_{x\in X} < \omega(\cdot|x')|C_{xx'}(\{p(\cdot|\cdot;0,0)\},q)|\tau(\cdot|x)>,$$

where the notation $< V_1|h|V_2 >$ denotes the inner product of the tangent vectors V_1 and V_2 with respect to the quadratic form tensor h. 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 the matrix $k(x,x';q)$ as the constants of proportionality associated with each $x,x' \in X$. That is,

$$\frac{\partial}{\partial \nu}\frac{\partial}{\partial \varepsilon} C(\{p(\cdot|\cdot;\varepsilon,\nu)\},q)|_{\varepsilon=\nu=0} = \sum_{x'\in X}\sum_{x\in X} k(x,x';q) < \omega(\cdot|x')|g(r)|\tau(\cdot|x)>,$$

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

In the case of mutual information, the matrix $k(x,x';q)$ is itself the inverse Fisher information matrix,

$$k(x,x';q) = \delta(x,x')q(x) - q(x)q(x'),$$

where $\delta(x,x')$ is the Kronecker delta function.

We are now in position to discuss our approximation of the information cost function. We use Taylor's theorem to approximate the cost function and its gradient up to order Δ (we use Δ because in future sections, we will be looking at small time intervals).

Theorem 2. *Suppose that a signal structure $\{p(\cdot|\cdot)\}$, with signal alphabet S, is described by the equation*

$$p(s|x) = r(s) + \Delta^{\frac{1}{2}}\tau(s|x) + o(\Delta^{\frac{1}{2}}),$$

where, for any $x \in X$, $\tau(s|x) \neq 0 \Rightarrow r(s) > 0$. Let $C(\{p(\cdot|\cdot)\};q)$ be a rational inattention cost function that satisfies conditions 1-4. Then

1. *The cost of this signal structure is, for some matrix $k(x,x';q)$,*

$$C(\{p(\cdot|\cdot)\};q) = \frac{1}{2}\Delta \sum_{x' \in X} \sum_{x \in X} k(x,x';q) < \tau(\cdot|x')|g(r)|\tau(\cdot|x) > + o(\Delta).$$

2. *The gradient of this cost function with respect to $\tau(\cdot|x)$, for a given $x \in X$,*

$$\nabla_x C(\cdot) = \Delta^{0.5} \sum_{x' \in X} k(x,x';q) < \tau(\cdot|x')|g(r)| + o(\Delta^{\frac{1}{2}}).$$

3. *The matrix $k(x,x';q)$ is positive semi-definite and symmetric, with a single eigenvector in its null space, and satisfies $\sum_{x'} k(x,x';q) = 0$.*

Proof. See appendix, section A.1. □

The results of theorem 2 characterize the cost of a small amount of information, for any rational inattention cost function satisfying our conditions. The theorem substantially restricts the local structure of the cost function, relative to the most general possible alternatives (which would not satisfy our conditions). Potential information structures $\{p(s|x)\}$ can be represented as vectors of dimension $N = (|S|-1) \times |X|$. Under the assumptions of conditions 1, 2, and 4 (but not the Blackwell's ordering condition, condition 3), the cost function would locally resemble an inner product with respect to a positive semi-definite, $N \times N$ matrix. By imposing condition 3, the results of theorem 2 show that we can restrict this matrix to the $k(x,x';q)$ matrix, an $|X| \times |X|$ matrix. If the agent were only allowed binary signals ($|S| = 2$), this restriction would be trivial. When the agent is allowed to contemplate more general signal structures, the restriction is non-trivial.

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. The results are described in the corollary below.

Corollary 1. *Under the assumptions of theorem 2, the posterior beliefs can be written as*

$$q(x|s) = q(x) + \Delta^{\frac{1}{2}} \frac{q(x)}{r(s)} (\tau(s|x) - \sum_{x' \in X} \tau(s|x')q(x')) + o(\Delta^{\frac{1}{2}}).$$

Define the matrix

$$\bar{k}(x,x';q) = \begin{cases} \frac{k(x,x';q)}{q(x)q(x')} & if\ q(x), q(x') > 0 \\ 0 & otherwise. \end{cases}$$

The cost function can be written as

$$C(\{p(\cdot|\cdot)\},q) = \frac{1}{2}\Delta \sum_{s\in S}\sum_{x'\in X}\sum_{x\in X} r(s)[q(x|s)-q(x)][q(x'|s)-q(x')]\bar{k}(x,x';q)+o(\Delta).$$

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

In the next section, we analyze a standard rational inattention problem, using these results.

4 Static Problems of Rational Inattention: Approximations

We begin with the standard rational inattention problem, described in equation (2). Our plan is to use the approximation we have derived in the previous section to consider a problem in which the utility consequences of the decision are small (or, equivalently, the cost of acquiring information is large). We parametrize the scale of the utility function by the parameter $\Delta^{\frac{1}{2}}$, and consider the limit as $\Delta \to 0^+$. Our choice of the notation Δ, and the scaling of the utility by $\Delta^{\frac{1}{2}}$, is motivated by the problem in the next section, in which Δ will be the length of time period. For the static problem, scaling the utility by $\Delta^{\frac{1}{2}}$, as opposed to Δ or some other alternative, is arbitrary and without loss of generality.

The agent solves

$$\max_{\{p(s|x)\in\mathscr{P}(S)\}} \Delta^{\frac{1}{2}} \sum_{x\in X} q(x) \sum_{s\in S} p(s|x) u(x,a^*(s;\{p(s|x)\})) - C(\{p(\cdot|\cdot)\},q), \tag{7}$$

with the optimal action $a^*(\cdot)$ defined as in equation (1). We will use the standard simplification[7] that it is without loss of generality to consider a signal alphabet S that is the one-to-one

[7] Sims [2010], De Oliveira et al. [2013]

with the set of possible actions A. That is, instead of optimizing over conditional probability distributions $\{p(s|x)\}$, we will optimize over the conditional probabilities of actions, $\{p(a|x)\}$. We assume that the utility function $u(x,a)$, over states $x \in X$ and actions $a \in A$, has full row rank, meaning that no action's payoffs can be perfectly replicated by a linear combination of the other actions.[8] We also assume that the prior, $q(x)$, has full support over the set of states, X, and that the utility function is bounded.

We consider an approximation as Δ becomes small. Let p^*_Δ denote the solution to the static rational inattention problem (equation (7)), given the parameter Δ. In the lemma below, we show that the optimal policy converges to an uninformative signal structure.

Lemma 1. *Define* $r^*_\Delta(a) = \sum_{x\in X} q(x) p^*_\Delta(a|x)$. *Under an arbitrary norm on the tangent space of signal structures,* $||\cdot|| : \mathbb{R}^{|A|\times|X|} \to \mathbb{R}^+$, *there exists a* $\Delta_B > 0$ *and constant* B *such that, for all* $\Delta < \Delta_B$,

$$\lim_{\Delta\to 0^+} \Delta^{\frac{1}{2}} B \geq ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||.$$

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

This lemma demonstrates that the optimal policy converges to an uninformative one as the parameter Δ converges to zero. We can interpret this limit as studying a situation in which the utility benefits of choosing the "right" action instead of the "wrong" action shrink towards zero, or equivalently as a situation in which the cost of acquiring information becomes increasingly large.

A straightforward corollary of this lemma is that the optimal policy can be expressed in a manner similar to the assumption in made in theorem 2.

[8] This assumption is made for mathematical convenience. We believe our results would hold in the general case, without this restriction.

Corollary 2. *Let* Δ_m*,* $m = \{0, 1, \ldots\}$*, denote a sequence such that* $\lim_{m\to\infty} \Delta_m = 0$*. The optimal policy* $p^*_{\Delta_m}$ *has a convergent sub-sequence, denoted* p^*_n*, that can be expressed as*

$$p^*_n(a|x) = r^*(a) + \phi^*_n(a) + \Delta_n^{\frac{1}{2}} \tau^*(a|x) + o(\Delta_n^{\frac{1}{2}}), \tag{8}$$

where $r^* \in \mathscr{P}(A)$*,* $\lim_{n\to\infty} \phi^*_n(a) = 0$ *for all* $a \in A$*,* $\sum_{a\in A} \phi^*_n(a) = \sum_{a\in A} \tau^*(a|x) = 0$ *for all* $x \in X$*,* $\sum_{x\in X} \tau^*(a|x)q(x) = 0$ *for all* $a \in A$*, and* $||q(\cdot)\tau^*(\cdot|\cdot)||$ *is bounded.*

For all a such that $r^*(a) > 0$*, the posterior associated with taking action a satisfies*

$$q^*_n(x|a) = q(x) + \Delta_n^{\frac{1}{2}} q(x) \frac{\tau^*(a|x)}{r^*(a)} + o(\Delta_n^{\frac{1}{2}}).$$

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

The existence of a convergent sub-sequence follows immediately from the boundedness of the optimal policy. At this point, not clear that there is a unique limit to which all optimal policies must converge. Intuitively, as the agent's cost of acquiring information increases, an optimal should converge to one that randomizes over set of ex-ante optimal actions. We denote this set as $A_+ = \{a \in \arg\max_{a'\in A} \sum_{x\in X} q(x)u(x, a')\}$.

One might think that any $r^* \in \mathscr{P}(A_+)$ can be the limit of a sequence of optimal policies. However, we find that the optimal policy converges to a unique probability distribution. The uniqueness of r^* is driven by the need to maximize information-gathering opportunities. We reconcile these apparently conflicting intuitions by noting that the ϕ^*_n term can converge to zero arbitrarily slowly. In the theorem below, we characterize the unique values of r^* and $\tau^*(a|x)$ that characterize all optimal policies.

Theorem 3. *The set of optimal policies in the rational inattention problem, as* $\Delta \to 0$*,* *satisfy*

$$p^*_\Delta(a|x) = r^*(a) + \phi_\Delta(a) + \Delta^{\frac{1}{2}} r^*(a) \sum_{x' \in X_0} q(x') k^+(x,x';q)[u(x',a) - \sum_{a' \in A} r(a')u(x',a')] + o(\Delta^{\frac{1}{2}}),$$

where $r^*(a) \in \mathscr{P}(A_+)$ *and* $k^+(x,x';q)$ *is the pseudo-inverse of* $k(x,x';q)$*.*

The optimal posteriors $q^*_\Delta(x|a)$ *satisfy*

$$q^*_\Delta(x|a) = q(x) + \Delta^{\frac{1}{2}} \sum_{x' \in X_0} \bar{k}^+(x,x';q)[u(x',a) - \sum_{a' \in A} r(a')u(x',a')] + o(\Delta^{\frac{1}{2}}),$$

where $\bar{k}^+(x,x';q)$ *is the pseudo-inverse of* $\bar{k}(x,x';q)$*.*

The function $r^*(a)$*, which is the limit of every sequence of optimal policies* $p^*_\Delta(a|x)$*, is the unique solution to the problem*

$$\max_{r(a) \in \mathscr{P}(A_+)} \sum_{a \in A_+} r(a) m(a,a)$$
$$- \sum_{a \in A_+} \sum_{a' \in A_+} r(a') m(a,a') r(a),$$

where the matrix $m(a,a')$ *is defined, for* $a,a' \in A_+$*, as*

$$m(a,a') = \sum_{x \in X_0} \sum_{x' \in X_0} \bar{k}^+(x',x;q) u(x',a) u(x,a').$$

Proof. See appendix, section A.5. □

This theorem generalizes common results from a static rational inattention problem with the mutual information cost function. Under the mutual information cost function, the matrix $\bar{k}^+(x,x';q)$ is the inverse Fisher information matrix, evaluated at the prior $q(x)$, and the matrix $m(a,a')$ is the covariance matrix of the ex-ante utility of the various ex-ante optimal actions.

For the prior-invariant cost functions, $\bar{k}^{+}(x,x';q) = q(x)q(x')k^{+}(x,x')$. Given any prior-invariant cost function, we can compute the associated information cost matrix, k, and define a sequentially prior-invariant cost function using this matrix (see equation (4)). Up to order $\Delta^{\frac{1}{2}}$, the behavior of the agent under the prior-invariant cost function and the sequentially prior-invariant cost function would be identical. This is our first justification for the use of the sequentially prior-invariant cost function.

Up to order $\Delta^{\frac{1}{2}}$, the agent always chooses an action that is ex-ante optimal. Among this set of actions, the agent prefers actions whose payoffs differ from the payoffs of the other ex-ante optimal actions. In the case of the mutual information cost function, it is desirable to sometimes choose actions whose utility varies the most across states (because one can choose the action more frequently when the utility is high than when it is low). In the general case, a similar intuition holds, but something other than the variance is the relevant statistic.

Because the matrix $\bar{k}^{+}(x,x';q)$ is not (in general) the inverse Fisher information matrix, the general cost functions we derive differ in their predictions from the mutual information cost function in several respects. Neither the invariant likelihood ratio property or the locally invariant posteriors property described in Caplin and Dean [2013] hold, consistent with those authors' rejection of the invariant likelihood ratio property in a laboratory setting.

In the next section, we explore a repeated version of the static problem rational inattention problem.

5 Sequential Evidence Accumulation

In this section, we study a dynamic problem in which the agent has repeated opportunities to gather information before making a decision. The state of the world, $x \in X$, remains constant over time. At each time t, the agent can either stop and take an action $a \in A$, or continue and receive a signal $p_t(s|x)$, for some signal $s \in S$. We assume that the number of potential actions is weakly less than the number of states, $|A| \leq |X|$. We also assume that the signal alphabet S is finite and fixed over time, with $|S| > |X|$. However, signal structure $p_t(s|x)$ is a choice variable that can be state- and time-dependent. Fixing the signal structure S has no economic meaning, because the information content of receiving a particular signal $s \in S$ can change between periods. The assumption allows us to assume a finite signal structure and invoke the results from the previous sections.[9]

The agent's prior beliefs at time t, before receiving the signal, are denoted q_t. Each time period has a length Δ. If the agent always stopped after a single period of learning, this problem would be virtually identical to the static rational inattention problem studied in the previous section. Let τ denote the time at which the agent stops and makes a decision, with $\tau = 0$ corresponding to making a decision without acquiring any information. At this time, the agent receives utility $u(x,a) - \kappa\tau$ if she takes action a at time τ and the true state of the world is x. The parameter κ governs the penalty the agent faces from delaying his decision. The reason the agent does not make a decision immediately is that she is able to gather information, and make a more-informed decision.

The agent can choose a signal 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 agent needs to design the

[9] As mentioned previously, the work of Ay et al. [2014] discusses how to extend the Chentsov [1982] theorems to infinite dimensional structures. We speculate that their results would allow us to extend our theorems to infinite dimensional signal spaces.

signal structure. The agent is constrained to satisfy

$$E_0[\frac{\Delta}{\tau\rho}\sum_{j=0}^{\tau\Delta^{-1}-1} C(\{p_{\Delta j}(\cdot|x')\}_{x'\in X}, q_{\Delta j})^{\rho}]^{\frac{1}{\rho}} \leq \Delta c, \tag{9}$$

if the agent choose to acquire any information at all ($\tau > 0$). 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. In the limit as $\rho \to \infty$, this would approach a per-period constraint on the amount of information the agent can obtain. For finite values of ρ, the agent can allocate more information gathering to states and times in which it is more advantageous to gather more information. We will assume, however, that $\rho > 1$.

We will also impose two additional assumptions on the cost function. We will assume that the cost function exhibits strong convexity with respect to signal 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 $q \in \mathscr{P}(X)$, and all signal structures that are sufficiently close to uninformative ($C(\{p(\cdot|\cdot)\}, q) < B$),

$$C(\{p(\cdot|\cdot)\}, q) \geq \frac{m}{2}\sum_{s\in S}(\sum_{x'\in X} p(s|x')q(x'))||q(\cdot|s) - q(\cdot)||_X^2,$$

where $q(x|s) = \frac{q(x)p(s|x)}{\sum_{x'\in X} p(s|x')q(x'))}$, and $||\cdot||_X$ is an arbitrary norm on the tangent space of $\mathscr{P}(X)$.

This condition is slightly stronger than 1; it it essentially "local strong convexity" instead of local convexity. It implies that the second smallest eigenvalue of the information cost matrix is bounded away from zero. The mutual information cost function satisfies this condition; it is also satisfied by all prior-invariant cost functions satisfying our other conditions.

We also impose a differentiability assumption on the matrix $k(x,x',q)$ associated with the cost function.

Condition 6. Each element of the matrix $k(x,x',q)$ that is associated with the cost function $C(\{p(\cdot|\cdot)\},q)$ is differentiable with respect to the prior, q.

This condition ensures that the cost of acquiring a small amount of information varies smoothly over the space of prior beliefs. Armed with these two additional conditions, we discuss our solution to the sequential evidence accumulation problem.

Let $V(q_0;\Delta)$ denote the value of the solution to the sequence problem for an agent with prior beliefs q_0, and let q_τ denote the agent's beliefs when stopping to make a decision.

$$V(q_0;\Delta) = \max_{\{p_{\Delta j}\},\tau} E_0[\sum_{x\in X} q_\tau(x)(u(a^*(q_\tau),x) - \kappa\tau)],$$

subject to the information cost constraint, equation (9). The dual version of this problem can be written, assuming the agent acquires some information, as

$$\begin{aligned} W(q_0,\lambda;\Delta) = \max_{\{p_{\Delta j}\},\tau} & E_0[\sum_{x\in X} q_\tau(x)(u(a^*(q_\tau),x) - \kappa\tau)] - \\ & \lambda E_0[\frac{\Delta^{1-\rho}}{\tau}\sum_{j=0}^{\tau\Delta^{-1}-1}\{\frac{1}{\rho}C(\{p_{\Delta j}(\cdot|\cdot)\},q_{\Delta j}(\cdot))^\rho - \Delta^\rho c^\rho\}]. \end{aligned} \quad (10)$$

Here, the function $W(q_0,\lambda;\Delta)$ 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 refer to the function W as the value function, bearing in mind that λ is not actually exogenous to the problem. We will proceed under the assumption that $\lambda \in (0,\kappa c^{-\rho})$. Below, we demonstrate that there is no duality gap in the continuous time limit of this problem, and that this assumption is without loss of generality.

5.1 The Value Function

We begin by describing the recursive representation for the value function $W(q_t, \lambda; \Delta)$, and discussing certain technical lemmas that are necessary to establish our main results. The value function has a recursive representation:

$$
\begin{aligned}
W(q_t, \lambda; \Delta) \quad = \quad & \max\{\max_{\{p(\cdot)\}} -\kappa\Delta + \lambda\Delta^{1-\rho}(\Delta^{\rho}c^{\rho} - \frac{1}{\rho}C(\cdot)^{\rho}) + \\
& \sum_{x \in X}\sum_{s \in S} p(s|x)q_t(x)W(q_{t+\Delta}(\cdot, s), \lambda; \Delta, x), \\
& \max_{a \in A}\sum_{x \in X} q_t(x)u(a, x)\},
\end{aligned}
$$

where $q_{t+\Delta}(\cdot, s)$ is pinned down by Bayes' rule and $W(q_t, \lambda; \Delta, x)$ is the "state-specific" value function (the value function conditional on the true state being x). The state-specific value function also has a recursive representation, in the region in which the agent continues to gather information:

$$
\begin{aligned}
W(q_t, \lambda; \Delta, x) \quad = \quad & -\kappa\Delta + \lambda\Delta^{1-\rho}(\Delta^{\rho}c^{\rho} - \frac{1}{\rho}C(\cdot)^{\rho}) + \\
& \sum_{s \in S_{\Delta j}} p_t^*(s|x)W(q_{t+\Delta}^*(\cdot, s), \lambda; \Delta, x).
\end{aligned}
$$

In this equation, we take the optimal signal structure as given. Note that, by construction, wherever the agent does not choose to stop,

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

To being our analysis, we note that the value function $W(q_t, \lambda; \Delta)$ is well-behaved:

Lemma 2. *The value function $W(q_t, \lambda; \Delta)$ is bounded on $q_t \in \mathscr{P}(X)$, and convex in q.*

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

The boundedness of the value function follows from the setup of the problem: ultimately, the agent will make a decision, and the utility from making the best possible decision in the best possible state of the world is finite. The convexity of the value function is what motivates the agent to acquire information. By updating her beliefs from q to either q' or q'', with $q = \alpha q'' + (1-\alpha)q'$ for some $\alpha \in (0,1)$, the agent improves her welfare by enabling better decision making.

The boundedness and convexity of the value function are sufficient to establish that a second-order Taylor expansion of the value function exists almost everywhere. Using this result, we can characterize the optimal information gathering policy (or policies, as there is not necessarily a unique optimum) as the time period shrinks. In particular, we establish that an optimal signal structure $p^*_{t,\Delta}(\cdot|\cdot)$ converges to an uninformative signal structure as $\Delta \to 0^+$.

Lemma 3. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that $\lim_{m\to\infty} \Delta_m = 0$. At each time t and prior $q_t \in \mathscr{P}(X)$, the optimal policy p^*_{t,Δ_m} has a convergent sub-sequence, denoted $p^*_{t,n}$, that can be expressed as*

$$p^*_{t,n}(s|x) = r^*_t(s) + \phi^*_{t,n}(s) + \Delta_n^{\frac{1}{2}} \tau^*_t(s|x) + o(\Delta_n^{\frac{1}{2}}),$$

where $r^*_t \in \mathscr{P}(S)$, $\lim_{n\to\infty} \phi^*_{t,n}(s) = 0$ for all $s \in S$, $\sum_{s\in S} \phi^*_{t,n}(s) = \sum_{s\in S} \tau^*_t(s|x) = 0$ for all $x \in X$, $\sum_{x\in X} \tau^*_t(s|x) q_t(x) = 0$ for all $s \in S$, and $||q_t(\cdot)\tau^*_t(\cdot|\cdot)||$ is bounded.

For all s such that $r^*_t(s) > 0$, the posterior associated with receiving signal s satisfies

$$q^*_{t,n}(x|s) = q_t(x) + \Delta_n^{\frac{1}{2}} q_t(x) \frac{\tau^*_t(s|x)}{r^*_t(s)} + o(\Delta_n^{\frac{1}{2}}).$$

Proof. See appendix, section A.8. □

This lemma is virtually identical to corollary 2 in the static problem. The key step is proving something analogous to lemma 1 in the static problem: that, as the time period shrinks, the optimal quantity of information acquired vanishes at a sufficiently fast rate. The convergence of the signal structure to an uninformative one, as the time period shrinks, allows us to use the approximation described in the previous sections of this paper to study the continuous time limit of the sequential evidence accumulation model.

In the standard rational inattention problems, it is without loss of generality to equate signals and actions. In this problem, when the agent does not stop and make a decision, the "action" is updating one's beliefs. Rather than consider a probability distribution over signals, and then an updating of beliefs by Bayes' rule, one can consider the agent to be choosing a probability distribution over posteriors, subject to the constraint that the expectation of the posterior is equal to the prior.[10] For any convergent sub-sequence of optimal policies, we can define the revision to the posterior (up to order $\Delta^{\frac{1}{2}}$) as

$$z_t^*(x|s) = q_t(x)\frac{\tau_t^*(s|x)}{r_t^*(s)}.$$

The almost-everywhere differentiability of the value function invites us to consider approximations to the Bellman equations described previously.

Proposition 1. *For any $q_t \in \mathscr{P}(X)$ and an associated convergent sequence of optimal policies, if the value function is twice-differentiable with respect to q at q_t, then for all s such that $r^*(s) > 0$,*

$$\theta\Delta \sum_{x'\in X}\sum_{x\in X} \bar{k}(x,x',q_t) z_t^*(x|s) z_t^*(x'|s) = \Delta z_t^*(\cdot|s)^T \cdot W_{qq}(q_t,\lambda;\Delta)\cdot z_t^*(\cdot|s) + o(\Delta),$$

[10]The notion of choosing a probability distribution over posteriors appears in Kamenica and Gentzkow [2011] and Caplin and Dean [2015], among other papers.

where $\theta = \lambda(\rho\frac{\kappa-\lambda c^{\rho}}{\lambda(\rho-1)})^{\frac{\rho-1}{\rho}}$. The scale of the vector $z_t^*(\cdot|s)$ is such that, up to order Δ, the flow information cost is constant everywhere:

$$\frac{1}{\rho}\Delta^{-\rho}C(\cdot)^{\rho} = \frac{\kappa-\lambda c^{\rho}}{\lambda(\rho-1)} + o(1).$$

Proof. See appendix, section A.9. □

The results of this proposition describe an approximate difference equation that the value function must satisfy, for any update to the posterior belief, $z_t^*(x|s)$, that would be observed under the optimal policy. In the case of the mutual information cost function, in which $k(x,x',q_t)$ is the inverse Fisher information matrix, the matrix $\bar{k}(x,x',q_t)$ is the Fisher information matrix.

5.2 Continuous Time

The results of proposition 1 make it tempting to neglect the higher-order terms in that proposition and attempt to solve the resulting continuous time problem. In this section, we establish that the value function associated with the discrete time problem does in fact converge to a function characterized by a version of this equation that neglects higher-order terms. We will call this function "the" continuous time value function, although we have not yet proved it is unique. It is important to emphasize that this function is not necessarily the solution to a continuous-time rational inattention problem; it is instead a function which is very close to the solution of the discrete time rational inattention problem, as the time interval becomes small.

We begin by defining the conditional covariance matrix of the posterior beliefs, given some time interval Δ. Define

$$\Omega^*_{t,\Delta}(x,x') = \sum_{s\in S} r^*_{t,\Delta}(s)(q^*_{t+\Delta,\Delta}(x'|s) - q_t(x'))(q^*_{t+\Delta,\Delta}(x|s) - q_t(x)).$$

As mentioned previously, given some beliefs q_t, there may be multiple optimal policies, and therefore multiple matrices $\Omega_{t,\Delta}$ that characterize the covariance matrix of updates to beliefs under an optimal policy. We begin by showing that there exists a convergent sub-sequence of time intervals such that a limiting value function and stochastic process for beliefs exist.

Lemma 4. *Let Δ_m, $m \in \mathbb{N}$, denote a sequence such that* $\lim_{m\to\infty}\Delta_m = 0$. *Let $q_{t,m}$ denote the stochastic process for the agent's beliefs at time t, under the optimal policy, given $\Delta = \Delta_m$. There exists a sub-sequence $n \in \mathbb{N}$ such that*

1. *The value function $W(q,\lambda;\Delta_n)$ converges uniformly on $q \in \mathscr{P}(X)$ to a bounded and convex function, $W(q,\lambda)$.*

2. *The law of the stochastic process $q_{t,n}$ converges, in the sense of weak convergence, to*

$$q_t = q_0 + \int_0^t \sigma^*_s dB_s,$$

where B_s is an $|X|$-dimensional Brownian motion, each element of $\sigma^*_s\sigma^T_s$ is a uniformly integrable stochastic process adapted to the filtration generated by q_t, and

$$\sigma^*_s\sigma^{*T}_s = \lim_{n\to\infty} \Delta_n^{-1}\Omega^*_{t,\Delta_n}.$$

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

A key implication of lemma 4 is the existence of a limit for $\Delta_n^{-1}\Omega_{t,\Delta_n}$. Thus far, we have avoided claiming that there is a unique optimal policy; the results in the previous section consider only convergent sub-sequences of policies at a particular point. The results in lemma 4 prove something stronger– that under a sub-sequence of optimal policies, the stochastic process for beliefs converges to a particular martingale. As a result, we can speak of a convergent sub-sequence of policies as a function of beliefs, as opposed to a point-wise convergent sub-sequence.

We next show that the value function that is the limit of the discrete time problem indeed satisfies the differential equation suggested by proposition 1.

Lemma 5. *Let $n \in \mathbb{N}$ index the sub-sequence of policies described in lemma 4. The function $W(q_t,\lambda) = \lim_{n\to\infty} W(q_t,\lambda;\Delta_n)$ satisfies, for all $q_t \in \mathscr{P}(X)$,*

$$\theta tr[\bar{k}(\cdot,\cdot,q_t)\sigma_t\sigma_t^T] = tr[W_{qq}(q_t,\lambda)\sigma_t\sigma_t^T],$$

where $tr[\cdot]$ denotes the trace of a matrix.

Proof. See the appendix, section A.11. The proof relies substantially on the results of Amin and Khanna [1994]. □

We have shown that the continuous time value function satisfies the differential equation suggested by our discrete time approximation. In the next theorem, we provide a partial solution for the continuous time value function. We show that it coincides with the solution to a particular kind of static rational inattention problem. We also resolve a lingering issues regarding uniqueness and duality; we show that the limiting value function is unique and that the value function for the constrained problem, $V(q_0)$, is equal to the value function we have been discussing, $W(q_0,\lambda)$.

Theorem 4. *Consider the continuous time limit of the sequential evidence accumulation model. There exists a unique solution to the sequence problem, $V(q_0) = \lim_{\Delta\to 0^+} V(q_0;\Delta)$,* *that satisfies*

$$V(q_0) = \max_{\pi \in \mathscr{P}(A), \{q_a(x) \in \mathscr{P}(X)\}_{a \in A}} \sum_{a \in A} \sum_{x \in X} \pi(a) q_a(x) u(a,x) - \theta^* \sum_{a \in A} \pi(a) D_k(q_a || q_0),$$

subject to the constraint that $\sum_{a \in A} \pi(a) q_a(x) = q_0(x)$ *for all* $x \in X$.

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

The constant θ^* *is*

$$\theta^* = \rho^{\frac{1}{\rho}} \frac{\kappa}{c}.$$

The divergence D_k *is defined as*

$$D_k(p||q) = \sum_{x \in X} \sum_{x' \in X} (p(x) - q(x))(p(x') - q(x')) \int_0^1 (1-s) \bar{k}(x, x', sp(x) + (1-s) q(x)) ds.$$

Proof. See appendix, section A.12. □

The sequential evidence accumulation problem can be thought of as a static rational inattention problem, with a separable cost function constructed from a particular divergence. The divergence embodies the expected time cost required for the agent to decide on a particular action. The probability distribution $q_a^* \in \mathscr{P}(X)$ is the agent's belief conditional on taking action $a \in A$. The vector q_a^* is unique, given a particular action a, meaning that there is only one belief the agent can reach before choosing to stop and take a particular action. The further this belief is from the agent's prior, q_0, the more time it will take (in expectation) for the agent to arrive at this belief before acting. The divergence encodes this notion of the belief q_a^* being "far" or "close to" q_0.

The interpretation of the divergence as a measure of the expected time to make a decision, conditional on choosing a particular action, is a novel aspect of the dynamic model. The case of the mutual information cost function illuminates the issue. Suppose that the original cost function $C(\cdot)$ in the static problem was mutual information. The matrix $k(x,x',q)$ would be the inverse Fisher information matrix, the divergence D_H would be the Kullback-Leibler divergence, and the expectation of the KL divergence over different actions would itself be mutual information. That is, mutual information exhibits a "self-similarity" property, meaning that as a cost function in the static model, it leads in the dynamic model to the same static problem. As a cost function for the static problem, we could interpret the cost function as relating to cognitive constraints, whereas in the dynamic problem, it represents the expected time cost of making a decision given information processing constraints.

In fact, any cost function $C(\cdot)$ that is invariant with respect to Markov congruent embeddings of the states $x \in X$ will yield the inverse Fisher information matrix as the matrix $k(x,x',q)$, and therefore the mutual information cost function in the dynamic problem. That is, one does not need need to believe in the particular functional form of mutual information in the static model to use it; one could instead appeal to the dynamic model for motivation, with any invariant cost function.

However, we have argued that invariance is not a natural property with respect to the states $x \in X$. These states have economic meaning, and it may be more or less costly to gather information about particular states of the world. In this case, the divergence D_k will not be the Kullback-Leibler divergence, but some other divergence. In these cases, "self-similarity" still holds; any cost function $C(\cdot)$ in the static model that has a particular information cost matrix $k(x,x';q)$ will result in a solution to the dynamic model that is characterized by the divergence D_k associated with that matrix. In particular, using the information cost function based on this divergence in the static model will result in the same cost function emerging from the dynamic model.

The constant multiplying the divergence has a simple interpretation (ignoring the constant $\rho^{\rho^{-1}}$, which is some number in $(1,1.45)$). The costlier it is for the agent to delay his action (κ), the costlier it is to make a well-informed choice. The more flow capacity the agent has to acquire information (c), the less costly it is to make a well-informed choice. The ratio of these two parameters, along with the scale of the matrix $k(x,x';q)$, determines how much time is required for the agent to, in expectation, acquire a particular quantity of information.

The value function is the solution to the sequence problem, and therefore is equal to the expected payoff minus the expected time cost. This is true unconditionally, and conditional on choosing a particular action. From this, it follows that the divergence represents that expected time cost associated with choosing that particular action. In laboratory settings, such as the one employed by Caplin and Dean [2015], it would be feasible to measure the average time it takes for an agent to reach a particular decision. This raises the hope that some properties of these divergences are measurable in choice data.

We have previously emphasized the "prior-invariant" class of cost functions. For these cost functions, two agents with different beliefs face the same cost of the acquiring the information. To embed this in our framework, we assume that the matrix $k(x,x',q)$ is identical for all q. It would follow that

$$\bar{k}(x,x';q) = \frac{k(x,x')}{q(x)q(x')},$$

where we have omitted the usual dependence of the k matrix on q to indicate the assumption of prior-independence.

Corollary 3. *Suppose that an information cost function, and therefore the associated information cost matrix $k(x,x',q)$, does not depend on the prior q. Let $D_{I,k}$ denote the divergence associated with the solution to the sequential evidence accumulation problem (theorem 4)* *in the case of prior independence. This divergence can be written as*

$$\begin{aligned} D_{I,k}(p||q) &= \sum_{x\in X}\sum_{x'\in X} k(x,x')\int_0^1 (1-s)\frac{(p(x)-q(x))(p(x')-q(x'))}{(sp(x)+(1-s)q(x))(sp(x')+(1-s)q(x'))}ds. \\ &= \sum_{x\in X}\sum_{x'\in X} k(x,x') f(\frac{p(x)}{q(x)},\frac{p(x')}{q(x')}), \end{aligned}$$

with

$$f(u,u') = \begin{cases} \frac{(u'-1)u\ln u-(u-1)u'\ln u'}{u-u'} & \text{if } u \neq u' \\ u-1-\ln u & \text{if } u = u' \end{cases}.$$

Proof. This follows by integration. □

5.3 The Dynamics of Sequential Evidence Accumulation

In this subsection, we discuss the dynamics of the agent's beliefs and information gathering strategy in the continuous time limit of the dynamic model. First, note that, conditional on choosing a particular action a, there is a unique belief q_a the agent will hold when making that decision. Intuitively, if the agent's beliefs could evolve such that there were two different beliefs q_a and q'_a such that the agent would stop and choose action a, then the agent has gathered too much information. The agent would be better off converging to some belief that was a linear combination of those two beliefs. From this logic, we can also observe that the number of "stopping beliefs" is at most the number of actions, $|A|$, weakly less than the number of states $|X|$ and signals $|S|$.

Given the uniqueness of these stopping beliefs, it follows by the martingale property of beliefs that at any time and state, the current belief is a linear combination of these stopping beliefs. This linear subspace has at most dimension $|A|-1<|S|$, and therefore the agent needs only $|A|-1$ signals in his signal alphabet. The agent's beliefs are a random walk in this linear subspace, evolving until they reach one of its boundaries, at which point the agent makes her decision. Each period, the agent gathers a constant flow amount of information, as measured by the information cost function. Note that the information cost matrix, $k(\cdot)$, can depend on the current belief, q_t. This implies that the volatility of beliefs can change, as the agent's beliefs change, even though the amount of information does not change.

These dynamics, for the most part, do not depend on the nature of the information cost function. However, there is one way in which the mutual information and prior-invariant cost functions differ in the dynamics of their belief evolution. For mutual information, at each moment in time, the agent receives one of many possible signals. Roughly, each of these signals can be interpreted as "make this action more likely." In contrast, in the prior-invariant case, the agent receives only a binary signal almost everywhere. In the prior-invariant case, the agents' beliefs move along a line, until they reach a "junction," at which point they may (or may not) change direction and move along a different line. In some sense, the agent in this case considers her options serially, whereas with the mutual information cost function, the agent considers her options in parallel. We leave a more detailed exploration of this phenomena for future work.

6 Conclusion

In this paper, we have introduced a new generalization of rational inattention models. This generalization is built from conditions that are appropriate for many economic problems. Despite the generality of our framework, we derive concrete predictions about behavior. We show that repeated versions of the static rational information problem, which we call sequential information accumulation, lead themselves to static rational inattention problems that are tractable. Moreover, the "information cost" in such problems in fact represents the expected time require to make a particular decision. We derive an expression for the divergence that corresponds to information costs that do not depend on the agent's prior, which is appealing in many applications. Together, these results provide a new perspective on which cost functions are appropriate for use in rational inattention problems and how those cost functions might be estimated in data.

References

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

Kaushik Amin and Ajay Khanna. Convergence of american option values from discrete-to continuous-time financial models1. *Mathematical Finance*, 4(4):289–304, 1994.

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

Lawrence M Benveniste and Jose A Scheinkman. On the differentiability of the value function in dynamic models of economics. *Econometrica: Journal of the Econometric Society*, pages 727–732, 1979.

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

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

Mark Caplin and Andrew Dean. The behavioral implications of rational inattention with shannon entropy. 2013.

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

John A. Clithero and Antonio Rangel. Combining response times and choice data using a neuroeconomic model of the decision process improves out-of-sample predictions. 2013.

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

Henrique De Oliveira, Tommaso Denti, Maximilian Mihm, and M Kemal Ozbek. Rationally inattentive preferences. *Available at SSRN 2274286*, 2013.

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. Stochastic choice and optimal sequential sampling. *Available at SSRN 2602927*, 2015.

Benjamin Hébert. Moral hazard and the optimality of debt. *Available at SSRN 2185610*, 2014.

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.

Filip Matejka and Alisdair McKay. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *CERGE-EI Working Papers wp442, The Center for Economic Research and Graduate Education-Economic Institute, Prague*, 2011.

Goran Peskir and Albert Shiryaev. *Optimal stopping and free-boundary problems*. Springer, 2006.

R Tyrrell Rockafellar. Convex analysis. 1970.

R Tyrrell Rockafellar. Second-order convex analysis. *J. Nonlinear Convex Anal*, 1:1–16, 1999.

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

Abraham Wald. *Sequential Analysis*. John Wiley & Sons, 1947.

Michael Woodford. Inattentive valuation and reference-dependent choice. *Unpublished Manuscript, Columbia University*, 2012.

Michael Woodford. An optimizing neuroeconomic model of discrete choice. Technical report, National Bureau of Economic Research, 2014.

A Proofs

A.1 Proof of theorem 2

Parts 1 and 2 of the theorem follow from a Taylor expansion of the cost function. Using the lemmas and theorem of Chentsov [1982], cited in the text, we know that for any invariant cost function with continuous second derivatives,

$$C(\{p(\cdot|\cdot)\},q) = \frac{1}{2}\Delta \sum_{x'\in X}\sum_{x\in X} k(x,x',q) < \tau(\cdot|x')|g(r)|\tau(\cdot|x) > +O(\Delta^{1.5}).$$

The second claim follows by a similar argument.

We next demonstrate the claimed properties of $k(x,x',q)$. First, $k(x,x',q)$ is symmetric, by the symmetry of partial derivatives and the assumption of continuous second derivatives (condition 4). Recall the assumption that

$$p(s|x) = r(s) + \Delta^{0.5}\tau(s|x) + O(\Delta),$$

which implies that $\sum_{s\in S} r(s) = 1$ and $\sum_{s\in S}\tau(s|x) = 0$ for all $x \in X$. Consider a signal structure, for which $\tau(s|x) = \phi(s)v(x)$, with $\sum_{s\in S}\phi(s) = 0$. For this signal structure,

$$C(\{p(\cdot|\cdot)\},q) = \frac{1}{2}\Delta\bar{g} \sum_{x'\in X}\sum_{x\in X} v(x)v(x')k(x,x',q) + O(\Delta^{1.5}),$$

where $< \phi|g(r)|\phi >= \bar{g} > 0$ is the Fisher information of $\phi(s)$. The posteriors associated with this signal structure are

$$p(x|s) = \frac{p(s|x)q(x)}{\sum_{x'\in X} q(x')p(s|x')} = q(x)\frac{r(s) + \Delta^{0.5}\phi(s)v(x)}{r(s) + \Delta^{0.5}\phi(s)\sum_{x'\in X} q(x')v(x')}.$$

If $v(x)$ is constant, then $p(x|s) = q(x)$ for all $s \in S$, the signal structure is uninformative, and $C(\{p(\cdot|\cdot)\},q) = 0$ by 1. In this case,

$$\sum_{x'\in X}\sum_{x\in X} k(x,x',q) = 0.$$

If $v(x)$ is not constant, then $p(x|s)$ is not equal to $q(x)$ for all s. In this case, the signal structure is informative, and $C(\{p(\cdot|\cdot)\},q) > 0$. Therefore,

$$\sum_{x'\in X}\sum_{x\in X} v(x)v(x')k(x,x',q) > 0.$$

It follows that $k(x,x)$ is positive semi-definite. It has a single eigenvector (a vector of constants) in its null space, and all of its other eigenvectors are associated with positive eigenvalues.

A.2 Proof of corollary 1

Under the stated assumptions,

$$p(s|x) = r(s) + \Delta^{\frac{1}{2}}\tau(s|x) + o(\Delta^{\frac{1}{2}}).$$

By Bayes' rule,

$$q(x|s) = \frac{p(s|x)q(x)}{\sum_{x'\in X} p(s|x')q(x')}.$$

It follows immediately that

$$\lim_{\Delta\to 0^+} q(x|s) = q(x)\frac{r(s)}{r(s)} = q(x).$$

Next,

$$\begin{aligned}\Delta^{-\frac{1}{2}}[q(x|s) - q(x)] &= \Delta^{-\frac{1}{2}} q(x)\frac{p(s|x) - \sum_{x'\in X} p(s|x')q(x')}{\sum_{x'\in X} p(s|x')q(x')}\\ &= q(x)\frac{\tau(s|x) - \sum_{x'\in X}\tau(s|x')q(x') + o(1)}{\sum_{x'\in X} p(s|x')q(x')}.\end{aligned}$$

For any s such that $r(s) > 0$,

$$\lim_{\Delta \to 0^+} \Delta^{-\frac{1}{2}}[q(x|s) - q(x)] = q(x)\frac{\tau(s|x) - \sum_{x' \in X} \tau(s|x')q(x')}{r(s)}.$$

By theorem 2,

$$C(\{p(\cdot|\cdot)\}, q) = \frac{1}{2}\Delta \sum_{x' \in X} \sum_{x \in X} k(x, x', q) < \tau(\cdot|x')|g(r)|\tau(\cdot|x) > +O(\Delta^{1.5}).$$

By the assumption that $\sum_{x \in X} k(x, x', q) = 0$, we have

$$C(\{p(\cdot|\cdot)\}, q) = \frac{1}{2}\Delta \sum_{x' \in X} \sum_{x \in X} k(x, x', q) \cdot$$
$$< \tau(\cdot|x') - \sum_{x' \in X} \tau(\cdot|x')q(x')|g(r)|\tau(\cdot|x) - \sum_{x' \in X} \tau(\cdot|x')q(x') > +$$
$$+ O(\Delta^{1.5}).$$

By the definition of the Fisher matrix,

$$C(\{p(\cdot|\cdot)\}, q) = \frac{1}{2}\Delta \sum_{s \in S: r(s) > 0} \sum_{x' \in X} \sum_{x \in X} r(s)k(x, x', q) \cdot$$
$$\frac{\tau(s|x') - \sum_{x' \in X} \tau(s|x')q(x')}{r(s)} \frac{\tau(s|x) - \sum_{x' \in X} \tau(s|x')q(x')}{r(s)} +$$
$$O(\Delta^{1.5}).$$

Substituting in the result regarding the posterior, assuming that $q(x) > 0$ for all $x \in X$,

$$C(\{p(\cdot|\cdot)\}, q) = \frac{1}{2}\Delta \sum_{s \in S: r(s) > 0} \sum_{x' \in X} \sum_{x \in X} r(s)\frac{k(x, x', q)}{q(x)q(x')}(q(x|s) - q(x))(q(x'|s) - q(x')) + O(\Delta^{1.5}),$$

which is the result.

A.3 Proof of lemma 1

Let $n = \{0, 1, \ldots\}$ index a sequence of time intervals Δ_n such that $\lim_{n\to\infty} \Delta_n = 0$. Define $p_n^*(a|x) = p_{\Delta_n}^*(a|x)$ and $r_n^*(a) = r_{\Delta_n}^*(a)$.

By optimality, we must have

$$\Delta^{\frac{1}{2}} \sum_{x\in X} q(x) \sum_{a\in A} [p_\Delta^*(a|x) - r_\Delta^*(a)] u(x,a) \geq C(\{p_\Delta^*\}; q).$$

By the Cauchy-Schwarz inequality, under Euclidean norm $||\cdot||_2$,

$$||u(\cdot,\cdot)||_2 \times ||q(x)p_\Delta^*(\cdot|\cdot) - q(x)r_\Delta^*(\cdot)||_2 \geq \sum_{x\in X} q(x) \sum_{a\in A} [p_\Delta^*(a|x) - r_\Delta^*(a)] u(x,a).$$

Note that by the full rank assumption, $||u(\cdot,\cdot)||_2 > 0$.

By Taylor's theorem and the continuous second-differentiability of $C(\cdot)$,

$$C(\{p_\Delta^*\}; q) = \frac{1}{2} \sum_{x\in X} \sum_{x'\in X} k(x,x';q) < p_\Delta^*(\cdot|x) - r_\Delta^*(\cdot) | g(r_\Delta^*) | p_\Delta^*(\cdot|x') - r_\Delta^*(\cdot) > \tag{11}$$
$$+ o(||q(x)(p_\Delta^*(\cdot|\cdot) - r_\Delta^*(\cdot))||^2).$$

Define a norm $||\cdot||_\xi$ by

$$||q(\cdot)(p(\cdot|\cdot) - \hat{p}(\cdot|\cdot))||_\xi^2 = \frac{1}{2} \sum_{x\in X} \sum_{x'\in X} \frac{k(x,x';q) + \xi}{q(x)q(x')} \times$$
$$< q(x)(p(\cdot|x) - \hat{p}(\cdot|x)) | g(\frac{1}{2} \sum_{x''\in X} (p(\cdot|x'') + \hat{p}(\cdot|x''))) | q(x')(p(\cdot|x') - \hat{p}(\cdot|x')) >, \tag{12}$$

which is a norm by the positive-definiteness of $[k(x,x';q) + \xi]$ for any $\xi > 0$ and the full support assumption on q. By construction, for all $a \in A$,

$$\sum_{x \in X} q(x)[p^*_\Delta(a|x) - r^*_\Delta(a)] = 0,$$

and therefore

$$C(\{p^*_\Delta\}; q) = ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi + o(||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi).$$

By the equivalence of norms, there exists a positive constant χ such that

$$\chi\Delta^{\frac{1}{2}}||u(\cdot,\cdot)||_\xi \times ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||_\xi \geq \\ ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi + o(||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi).$$

Taking the absolute value of the higher order term,

$$\chi\Delta^{\frac{1}{2}}||u(\cdot,\cdot)||_\xi \times ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||_\xi \geq \\ ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi - |o(||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi)|.$$

By the definition of the little-o notation, for every $\varepsilon > 0$, there exists a $\Delta_\varepsilon > 0$ such that for all $\Delta < \Delta_\varepsilon$,

$$\chi\Delta^{\frac{1}{2}}||u(\cdot,\cdot)||_\xi \times ||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||_\xi \geq (1-\varepsilon)||q(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||^2_\xi.$$

It follows that for an arbitrary $\varepsilon \in (0,1)$ and all $\Delta < \Delta_\varepsilon$, either $||q_t(\cdot)(p^*_\Delta(\cdot|\cdot) - r^*_\Delta(\cdot))||_\xi = 0$

or

$$\Delta^{\frac{1}{2}}\chi\frac{||u(\cdot,\cdot)||_{\xi}}{(1-\varepsilon)} \geq ||q(\cdot)(p^{*}_{\Delta}(\cdot|\cdot) - r^{*}_{\Delta}(\cdot))||_{\xi},$$

and therefore in all cases

$$\Delta^{\frac{1}{2}}\chi\frac{||u(\cdot,\cdot)||_{\xi}}{(1-\varepsilon)} \geq ||q(\cdot)(p^{*}_{\Delta}(\cdot|\cdot) - r^{*}_{\Delta}(\cdot))||_{\xi}.$$

By the equivalence of norms, the result holds for all norms, providing a construction of the constant $B > 0$ and associated parameter $\Delta_B (= \Delta_\varepsilon)$.

A.4 Proof of corollary 2

A convergent sub-sequence exists by the Bolzano-Weierstrauss theorem. Denote this sequence $l = \{0, 1, \ldots\}$. For such a sequence, define

$$r^{*}(a) = \lim_{l\to\infty}\sum_{x\in X} p^{*}_{l}(a|x)q(x).$$

By lemma 1, it follows that

$$\lim_{l\to\infty} p^{*}_{l}(a|x) = r^{*}(a)$$

for all $x \in X$.

Define

$$\tau_l(a|x) = \Delta_l^{-\frac{1}{2}}(p(a|x) - \sum_{x\in X} p^{*}_{l}(a|x')q(x')).$$

Again by lemma 1, $||q(x)\tau_l(a|x)|| \leq B$, and therefore a convergent sub-sequence exists. Denote this sequence $n = \{0, 1, \ldots\}$. Define

$$\tau^{*}(a|x) = \lim_{n\to\infty}\Delta_n^{-\frac{1}{2}}(p^{*}_{n}(a|x) - \sum_{x\in X} p^{*}_{n}(a|x')q(x')).$$

It follows that $\sum_{a\in A}\tau^*(a|x)=0$, that $\sum_{x\in X}\tau^*(a|x)q(x)=0$, and that $||q(x)\tau^*(a|x)||$ is bounded.

Define $\phi_n^*(a)=\sum_{x\in X}p_n^*(a|x')q(x')-r^*(a)$. It immediately follows that $\lim_{n\to\infty}\phi_n^*(a)=0$ for all $a\in A$ and that $\sum_{a\in A}\phi_n^*(a)=0$. Finally, note that

$$\lim_{n\to\infty}\frac{p_n^*(a|x)-r^*(a)-\phi_n^*(a)-\Delta_n^{\frac{1}{2}}\tau^*(a|x)}{\Delta_n^{\frac{1}{2}}}=0.$$

Finally, we demonstrate the claim that

$$q_n^*(x|a)=q(x)+\Delta_n^{\frac{1}{2}}q(x)\frac{\tau^*(a|x)}{r^*(a)}+o(\Delta_n^{\frac{1}{2}}).$$

By Bayes' rule,

$$q_n^*(x|a)=\frac{p_n^*(a|x)q(x)}{\sum_{x'\in X}p_n^*(a|x')q(x')}.$$

It follows from the convergence of $p_n^*(a|x)$ to $r^*(a)$ that, for all a such that $r^*(a)>0$,

$$\lim_{n\to\infty}q_n^*(x|a)=q(x).$$

Next, note that

$$\begin{aligned}\Delta_n^{-\frac{1}{2}}(q_n^*(x|a)-q(x))&=\Delta_n^{-\frac{1}{2}}q(x)\frac{p_n^*(a|x)-\sum_{x'\in X}p_n^*(a|x')q(x')}{\sum_{x'\in X}p_n^*(a|x')q(x')}\\&=q(x)\frac{\tau^*(a|x)+o(1)}{\sum_{x'\in X}p_n^*(a|x')q(x')}.\end{aligned}$$

It follows that, for all a such that $r^*(a)>0$,

$$\lim_{n\to\infty}\Delta_n^{-\frac{1}{2}}(q_n^*(x|a)-q(x))=q(x)\frac{\tau^*(a|x)}{r^*(a)},$$

proving the claim.

A.5 Proof of theorem 3

Consider the sequence m and convergent sub-sequence n defined in corollary 2. Define

$$r_n^*(a) = r^*(a) + \phi_n^*(a) + o(\Delta_n^{\frac{1}{2}}).$$

The Lagrangian version of the optimization problem is

$$\begin{aligned}\max_{\{p(a|x)\in\mathbb{R}^+_{|A|}\},\kappa(x),\nu(a|x)} \Delta_n^{0.5}\sum_{x\in X} q(x)\sum_{a\in A} p(a|x)u(x,a) - C(\{p(a|x)\}) \\ +\Delta_n^{0.5}\sum_{x\in X} q(x)\kappa_n(x)(1-\sum_{a\in A} p(a|x)) \\ +\Delta_n^{0.5}\sum_{x\in X} q(x)\sum_{a\in A} p(a|x)\nu_n(a|x).\end{aligned}$$

The concavity of the problem and the feasibility of gathering no information imply that Slater's condition holds, and the first-order conditions are necessary.The first order condition for $p(a|x)$ is

$$\frac{\partial C(\{p_n^*(a'|x')\})}{\partial p(a|x)} = q(x)\Delta_n^{\frac{1}{2}}[u(x,a) - \kappa_n(x) + \nu_n(a|x)].$$

Using Taylor's theorem, by the continuous second-differentiability of $C(\cdot)$,

$$\begin{aligned}\frac{\partial C(\{p_n^*(a'|x')\})}{\partial p(a|x)} = \sum_{x'\in X}\sum_{a'\in A}[p_n^*(a'|x') - r_n^*(a')]\frac{\partial^2 C(\{p(a''|x'')\})}{\partial p(a|x)\partial p(a'|x')}|_{\{p(a''|x'')\}=r_n^*(a)} \\ + o(||q(x)(p_n^*(a'|x') - r_n^*(a'))||).\end{aligned}$$

By lemma 1, if for all $\varepsilon > 0$, there exists an n_ε such that for all $n > n_\varepsilon$,

$$|x_n| \leq \varepsilon ||p_n^*(a'|x') - r_n^*(a')||,$$

then it follows that

$$|x_n| \leq \varepsilon B \Delta_n^{\frac{1}{2}},$$

and therefore for all $n > n_{\varepsilon B^{-1}}$,

$$|x_n| \leq \varepsilon \Delta_n^{\frac{1}{2}}.$$

It follows that

$$o(||p_n^*(a'|x') - r_n^*(a')||) \subseteq o(\Delta_n^{\frac{1}{2}}).$$

Using equation (8) and the first order condition,

$$\sum_{x' \in X} \sum_{a' \in A} [\Delta_n^{\frac{1}{2}} \tau(a'|x') + o(\Delta_n^{\frac{1}{2}})] \frac{\partial^2 C(\{p(a''|x'')\})}{\partial p(a|x) \partial p(a'|x')}|_{\{p(a''|x'')\} = r_n^*(a)} + o(\Delta_n^{\frac{1}{2}})$$
$$= q(x) \Delta_n^{\frac{1}{2}} [u(x,a) - \kappa_n(x) + \nu_n(a|x)].$$

We have

$$\frac{\partial^2 C(\{p(a''|x'')\})}{\partial p(a|x) \partial p(a'|x')}|_{\{p(a''|x'')\} = r_n^*(a)} = k(x,x') g_{a,a'}(r_n^*(a'')),$$

where $g_{a,a'}(r_n^*(a''))$ is the Fisher information matrix evaluated at r_n^*. Returning to the FOC,

$$\sum_{x' \in X} [\Delta_n^{\frac{1}{2}} \tau(a|x') + o(\Delta_n^{\frac{1}{2}})] \frac{k(x,x';q)}{r^*(a) + \phi_n^*(a) + o(\Delta_n^{\frac{1}{2}})} + o(\Delta_n^{\frac{1}{2}}) = q(x) \Delta_n^{\frac{1}{2}} [u(x,a) - \kappa_n(x) + \nu_n(a|x)].$$

Multiplying by $r_n^*(a)$ and summing over a,

$$\kappa_n(x) = \sum_{a \in A} r_n^*(a) u(x,a) + o(1).$$

Summing over x,

$$o(1) = \sum_{x \in X} q(x)[u(x,a) - \sum_{a' \in A} r_n^*(a') u(x,a')] + o(1) + \sum_{x \in X} q(x) \nu_n(a|x).$$

Taking the limit,

$$\lim_{n \to \infty} \sum_{x \in X} q(x) \nu_n(a|x) = \sum_{x \in X} q(x)[(\sum_{a' \in A} r^*(a') u(x,a') - u(x,a)].$$

By complementary slackness, it follows that

$$supp(r^*) \subseteq (A_+)$$

and that $v_n(a|x) = 0$ for all $a \in A_+$. It also follows that $\lim_{n \to \infty} \nu_n(a|x)$ is bounded above.

Returning to the FOC and rescaling,

$$\sum_{x' \in X} [\tau(a|x') + o(1)] k(x,x';q) + o(1) = q(x) r_n^*(a)[u(x,a) - \kappa_n(x) + \nu_n(a|x)].$$

Consider some $a \notin supp(r^*)$. Taking limits, it would follow that

$$\sum_{x' \in X} \tau(a|x') k(x,x';q) = q(x) \lim_{n \to \infty} r_n^*(a) \nu_n(a|x),$$

which by the boundedness of $\lim_{n \to \infty} \nu_n(a|x)$ is zero. Therefore,

$$supp(\tau(a|x)) = supp(r^*).$$

Now consider an $a \in supp(r^*)$. It follows that

$$\sum_{x' \in X} \tau(a|x')k(x,x';q) = q(x)r^*(a)[u(x,a) - \sum_{a' \in A} r^*(a')u(x,a')].$$

The only vectors in the null space of $k(x,x';q)$ are proportional to ι, the vector of ones. It follows that this equation has no solution if

$$\sum_{x \in X} q(x)r^*(a)[u(x,a) - \sum_{a' \in A} r^*(a')u(x,a')] \neq 0.$$

However, $r^* \in \mathscr{P}(A_+)$, and therefore $\sum_{x \in X} q(x)u(x,a) = \sum_{x \in X} q(x)u(x,a')$ for all a, a' such that $r(a) > 0$ and $r(a') > 0$. Therefore, a solution exists.

Let $k^+(x,x';q)$ denote the pseudo-inverse of $k(x,x';q)$. Any function of the form

$$\tau(a|x') = \sum_{x \in X} k^+(x,x';q)q(x)r^*(a)[u(x,a) - \sum_{a' \in A} r^*(a')u(x,a')] + \lambda(a)$$

is a solution. However, by the condition that $\sum_{x \in X} \tau(a|x)q(x) = 0$, we have

$$\sum_{x' \in X} \sum_{x \in X} q(x')k^+(x,x';q)q(x)r^*(a)[u(x,a) - \sum_{a' \in A} r^*(a')u(x,a')] + \lambda(a) = 0.$$

Therefore,

$$\tau(a|x') = \sum_{x\in X} k^{+}(x,x';q)q(x)r^{*}(a)[u(x,a) - \sum_{a'\in A} r^{*}(a')u(x,a')] -$$

$$\sum_{x'\in X}\sum_{x\in X} q(x')k^{+}(x,x';q)q(x)r^{*}(a)[u(x,a) - \sum_{a'\in A} r^{*}(a')u(x,a')].$$

Next, we return to the original problem and plug in the results derived thus far. First, Taylor-expanding the cost function around r_n^*,

$$C(\{p_n^*(a|x)\}) = \frac{1}{2}\sum_{a\in A}\sum_{x\in X}\sum_{x'\in X}[\Delta_n^{\frac{1}{2}}\tau(a|x') + o(\Delta_n^{\frac{1}{2}})]\frac{k(x,x';q)}{r^*(a) + \phi_n^*(a) + o(\Delta_n^{\frac{1}{2}})}[\Delta_n^{\frac{1}{2}}\tau(a|x) + o(\Delta_n^{\frac{1}{2}})] + o(\Delta).$$

This can rewritten as

$$\Delta_n^{-1}C(\{p_n^*(a|x)\}) = \frac{1}{2}\sum_{a\in supp(r^*)}\sum_{x\in X}\sum_{x'\in X}\tau(a|x')\frac{k(x,x';q)}{r^*(a)}\tau(a|x) + \frac{1}{2}\sum_{a\in A\setminus supp(r^*)}\sum_{x\in X}\sum_{x'\in X} o(1)\frac{k(x,x';q)}{r_n^*(a)} + o(1).$$

Now consider the utility benefit:

$$\Delta_n^{0.5}\sum_{x\in X} q(x)\sum_{a\in A} p_n^*(a|x)u(x,a) = \Delta_n^{0.5}\sum_{x\in X} q(x)\sum_{a\in A} r^*(a)u(x,a) + \Delta_n^{0.5}\sum_{x\in X} q(x)\sum_{a\in A}\phi_n^*(a)u(x,a) + \Delta_n\sum_{x\in X} q(x)\sum_{a\in A}\tau(a|x)u(x,a) + o(\Delta_n).$$

Because $r^* \in \mathscr{P}(A_+)$, we must have

$$\Delta_n^{0.5} \sum_{x \in X} q(x) \sum_{a \in A} \phi_n^*(a) u(x,a) \leq 0,$$

and by the positive-semidefiniteness of $k(x,x';q)$, it must be the case that

$$\sum_{a \in A \setminus supp(r^*)} \sum_{x \in X} \sum_{x' \in X} o(1) \frac{k(x,x';q)}{r_n^*(a)} \geq 0.$$

Now consider the policy

$$\bar{p}_n(a|x) = r^*(a) + \Delta_n^{\frac{1}{2}} \tau(a|x).$$

For such a policy,

$$\begin{aligned} U(\bar{p}_n(a|x)) \quad = \quad & \Delta_n^{0.5} \sum_{x \in X} q(x) \sum_{a \in A} r^*(a) u(x,a) + \\ & \Delta_n \sum_{x \in X} q(x) \sum_{a \in A} \tau(a|x) u(x,a) - \\ & \frac{1}{2} \Delta_n \sum_{a \in supp(r^*)} \sum_{x \in X} \sum_{x' \in X} \tau(a|x') \frac{k(x,x';q)}{r^*(a)} \tau(a|x) + \\ & o(\Delta_n). \end{aligned}$$

By assumption,

$$\Delta_n^{-1} [U(p_n^*(a,x)) - U(\bar{p}_n(a|x))] \geq 0,$$

and therefore

$$\begin{aligned}
\lim_{n\to\infty} \Delta_n^{-\frac{1}{2}} \sum_{x\in X} q(x) \sum_{a\in A} \phi_n^*(a) u(x,a) &= 0, \\
\lim_{n\to\infty} \Delta_n^{-1} C(\{p_n^*(a|x)\}) - \frac{1}{2} \sum_{a\in supp(r^*)} \sum_{x\in X} \sum_{x'\in X} \tau(a|x') \frac{k(x,x';q)}{r^*(a)} \tau(a|x) &= 0.
\end{aligned}$$

It follows that

$$\begin{aligned}
U(p_n^*(a|x)) = \; & \Delta_n^{0.5} \sum_{x\in X} q(x) \sum_{a\in A} r^*(a) u(x,a) + \\
& \Delta_n \sum_{x\in X} q(x) \sum_{a\in A} \tau(a|x) u(x,a) - \\
& \frac{1}{2}\Delta_n \sum_{a\in supp(r^*)} \sum_{x\in X} \sum_{x'\in X} \tau(a|x') \frac{k(x,x';q)}{r^*(a)} \tau(a|x) + \\
& o(\Delta_n).
\end{aligned}$$

Plugging in the solution for $\tau(a|x)$,

$$\begin{aligned}
U(p_n^*(a|x)) = \; & \Delta_n^{0.5} \sum_{x\in X} q(x) \sum_{a\in A} r^*(a) u(x,a) + \\
& \frac{1}{2}\Delta_n \sum_{a\in A} \sum_{x\in X} \sum_{x'\in X} q(x) k^+(x',x;q) q(x') r^*(a) [u(x',a) - \sum_{a'\in A} r^*(a') u(x',a')] u(x,a) \\
& -o(\Delta_n).
\end{aligned}$$

Because $r^*(a) \in \mathscr{P}(A_+)$, the first term is invariant to the choice of $r^*(a)$. Therefore,

$$r^*(a) \in \arg\max_{r(a)\in\mathscr{P}(A_+)} \sum_{a\in A} m(a,a) r(a) - \sum_{a\in A} \sum_{a'\in A} m(a,a') r(a) r(a'),$$

where

$$m(a,a') = \sum_{x\in X}\sum_{x'\in X} q(x)k^{+}(x',x;q)[q(x')u(x',a) - \sum_{x''\in X} q(x'')u(x'',a)][u(x,a') - \sum_{x''\in X} q(x'')u(x'',a')]$$
$$= \sum_{x\in X}\sum_{x'\in X} q(x)k^{+}(x',x;q)q(x')u(x',a)u(x,a').$$

By the assumption that $u(x,a)$ has full row rank, $m(a,a')$ is positive definite. It follows that $r^{*}(a)$ is unique.

We have proven that every convergent sub-sequence of $p^{*}_{m}(a|x)$ converges to a unique $r^{*}(a)$. It follows that $p^{*}_{m}(a|x)$ converges to $r^{*}(a)$. The claim regarding the posteriors follows from corollary 2, the definition

$$\bar{k}(x,x';q) = \frac{k(x,x';q)}{q(x)q(x')},$$

and the definition of the pseudo-inverse.

A.6 Proof of lemma 2

Write the value function in sequence-problem form:

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

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

$$W(q_0,\lambda;\Delta) \leq \bar{u} + \max_{\tau} -\kappa\tau + \Delta \sum_{j=0}^{\tau\Delta^{-1}-1} \lambda c^{\rho}.$$

Because $\lambda \in (0, \kappa c^{-\rho})$, the expression

$$-\kappa\tau + \Delta \sum_{j=0}^{\tau\Delta^{-1}-1} \lambda c^{\rho} = (\lambda c^{\rho} - \kappa)\tau$$

is weakly negative, and therefore

$$W(q_0,\lambda;\Delta) \leq \bar{u}.$$

By a similar argument, there is a smallest possible decision utility $\underline{\mathrm{u}}$, and because stopping now and deciding is always feasible,

$$W(q_0,\lambda;\Delta) \geq \underline{\mathrm{u}}.$$

Therefore, $W(q_0,\lambda;\Delta)$ is bounded for all $\lambda \in (0, \kappa c^{-\rho})$ and all Δ.

Next, recall that

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

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 $\mathscr{P}(X)$. Suppose not; then the agent could simply adopt the signal structures associated with beliefs q' and achieve higher utility, contradicting the optimality of the policy.

The convexity of the value function follows from this observation:

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

$$
W(\alpha q+(1-\alpha)q',\lambda;\Delta) \leq \alpha \sum_{x\in X} q(x)W(q,\lambda;\Delta,x)+(1-\alpha)\sum_{x\in X} q'(x)W(q',\lambda;\Delta,x),
$$

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

A.7 Additional Lemma

Lemma 6. *In the sequential evidence accumulation problem, for any norm $||\cdot||$ on the tangent space of signal structures, there exist constants B and $\bar{\Delta}$ such that, for all $\Delta<\bar{\Delta}$ and $q_t \in \mathscr{P}(X)$,*

$$
||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot)-r^*_{t,\Delta}(s))|| \leq B\Delta^{\frac{\rho-1}{2\rho-1}},
$$

where $p^*_{t,\Delta}(s|x)$ denotes an optimal policy given q_t and the time interval Δ, and $r^*_{t,\Delta}(s) = \sum_{x\in X} p^*_{t,\Delta}(s|x)q_t(x)$.

Proof. The agent's problem, conditional on not stopping at the current time, is

$$
\begin{aligned}
W(q_t,\lambda;\Delta) &= \max_{\{p(\cdot)\}} -\kappa\Delta+\lambda\Delta^{1-\rho}(\Delta^{\rho}c^{\rho}-\frac{1}{\rho}C(\cdot)^{\rho})+ \\
&\quad \sum_{x\in X}\sum_{s\in S} p(s|x)q_t(x)W(q_{t+\Delta}(\cdot,s),\lambda;\Delta,x),
\end{aligned} \tag{13}
$$

where $q_{t+\Delta}(\cdot,s)$ is the posterior associated with receiving signal s, and is determined using Bayes' rule, the prior q_t, and $p(s|x)$. Let $q^*_{t+\Delta,\Delta}(x,s)$ denote the posteriors associated with the optimal policy. The optimal signal structure must achieve weakly higher utility than any other signal structure. Consider, in particular, an uninformative signal structure. We must have

$$\sum_{s\in S} r^*_{t,\Delta}(s)[W(q^*_{t+\Delta,\Delta}(\cdot,s),\lambda;\Delta) - W(q_t,\lambda;\Delta)] \geq \lambda\Delta^{1-\rho}\frac{1}{\rho}C(\cdot)^{\rho}.$$

By the boundedness and convexity of W, it is Lipschitz-continuous, and therefore

$$K\sum_{s\in S} r^*_{t,\Delta}(s)||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t(\cdot)||_{X,2} \geq \sum_{s\in S} r^*_{t,\Delta}(s)|W(q^*_{t+\Delta,\Delta}(\cdot,s),\lambda;\Delta) - W(q_t,\lambda;\Delta)|,$$

where K is the associated Lipschitz constant and $||\cdot||_{X,2} : \mathbb{R}^{|X|-1} \to \mathbb{R}^+$ is the Euclidean norm, defined on the tangent space of posteriors. By the concavity of the square root function,

$$\sum_{s\in S} r^*_{t,\Delta}(s)||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t(\cdot)||_{X,2} \leq ||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(s))||_2, \tag{14}$$

where $||\cdot||_2 : \mathbb{R}^{|X|\times|A|-1} \to \mathbb{R}^+$ denotes the Euclidean norm on the tangent space of joint distributions over signals and states. From this argument, we observe that

$$K^{\rho^{-1}}\Delta^{1-\rho^{-1}}||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(s))||_2^{\rho^{-1}} \geq (\frac{\lambda}{\rho})^{\rho^{-1}}C(\cdot).$$

By the finiteness of the simplex and the assumption that $\rho > 1$, it follows that for Δ sufficiently small, $C(\cdot)$ is bounded above by any positive constant. Therefore, by 5, for all $\Delta < \bar{\Delta}$ such that 5 applies uniformly,

$$C(\cdot) \geq m\sum_{s\in S} r^*_{t,\Delta}(s)||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t(\cdot)||^2_{X,2},$$

for some positive constant m that does not depend on $p^*_{t,\Delta}$. By the fact that $r^*_{t,\Delta} \in \mathscr{P}(S)$,

$$\sum_{s \in S} r^*_{t,\Delta}(s)||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t(\cdot)||^2_{X,2} \geq \sum_{s \in S} r^*_{t,\Delta}(s)^2||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t||^2_{X,2},$$

and therefore

$$\sum_{s \in S} r^*_{t,\Delta}(s)||q^*_{t+\Delta,\Delta}(\cdot,s) - q_t||^2_{X,2} \geq ||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(\cdot))||^2_2. \tag{15}$$

Putting this together,

$$(\frac{K\rho}{\lambda})^{\rho^{-1}}\Delta^{1-\rho^{-1}}||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(s))||_2^{\rho^{-1}} \geq m||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(s))||^2_2.$$

Therefore,

$$m^{-1}(\frac{K\rho}{\lambda})^{\rho^{-1}}\Delta^{1-\rho^{-1}} \geq ||q_t(\cdot)(p^*_{t,\Delta}(\cdot|\cdot) - r^*_{t,\Delta}(s))||_2^{2-\rho^{-1}},$$

which proves the lemma for $B = m^{\frac{-\rho}{2\rho-1}}(\frac{K\rho}{\lambda})^{\frac{1}{2\rho-1}}$ and the Euclidean norm. The result holds for all norms by the equivalence of norms. □

A.8 Proof of lemma 3

By lemma 6, any convergent sub-sequence of optimal policies converges uniformly to an uninformative signal structure. However, the rate of convergence implied by this lemma is not sufficiently tight for our purposes, so we proceed with an additional argument.

Consider a sequence of Δ_n such that $p^*_{t,n}(s|x) = p^*_{t,\Delta_n}(s|x)$ converges (such a sequence exists by the boundedness of $p^*_{t,\Delta}(s|x)$). Define $r^*_t(s) = \lim_{n\to\infty}\sum_{x\in X} q_t(x)p^*_{t,n}(s|x)$. Define the function

$$f(\{p(s|x)\};q_t,\Delta,\lambda) = -\sum_{s\in S}(\sum_{x\in X} q_t(x)p(s|x))[W(q_{t+\Delta}(\cdot,s),\lambda;\Delta) - W(q_t(\cdot),\lambda;\Delta)],$$

and the function

$$g(\{p(s|x)\};q_t,\Delta,\lambda) = -\kappa\Delta + \lambda\Delta^{1-\rho}(\Delta^\rho c^\rho - \frac{1}{\rho}C(\{p(s|x)\};q_t)^\rho).$$

Note that both functions are concave, and that the function $g(\cdot)$ is differentiable with respect to the signal structure. The Bellman equation and optimality conditions require that, for an optimal signal structure,

$$g(\{p^*_{t,n}(s|x)\};q_t,\Delta_n,\lambda) - f(\{p^*_{t,n}(s|x)\};q_t,\Delta_n,\lambda) = 0$$

and for all signal structures,

$$g(\{p(s|x)\};q_t,\Delta_n,\lambda) - f(\{p(s|x)\};q_t,\Delta_n,\lambda) \leq 0.$$

Define $S_n \subseteq S$ as the support of $r^*_{t,n}(\cdot)$. By the convergence of $||q_t(\cdot)(p^*_{t,n}(\cdot|\cdot) - r^*_{t,n}(\cdot))||$ to zero, for each state $x \in X$, $p^*_{t,n}(\cdot|x)$ is in the interior of S_n. By Lemma 1 of Benveniste and Scheinkman [1979], the function f is differentiable, at $\{p^*_{t,n}(s|x)\}$, with respect to signal structures for which the support of $\sum_{x\in X} p(s|x)q_t(x)$ remains in S_n.

By theorem 23.8 of Rockafellar [1970], it follows that at each posterior $q^*_{t+\Delta_n,n}(\cdot,s)$ that arises from an optimal signal structure, the function $W(q,\lambda;\Delta)$ is differentiable with respect to q. (Suppose not: then for some signal realization, the sub-gradient of $W(q,\lambda;\Delta)$ would contain multiple vectors, and it would follow by theorem 23.8 that the sub-gradient of the function f also contained multiple vectors, contradicting the differentiability result derived previously).

Denote the derivative as $W_q(\cdot)$. We can write the optimality (first-order) condition as

$$(\sum_{s\in S} < W_q(q^*_{t+\Delta_n,n}(\cdot,s),\lambda;\Delta_n),\omega(s|\cdot)q_t(\cdot) >) = \lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1} < \nabla C(\cdot),\omega(\cdot|\cdot) >,$$

for any $\omega(s|x)$ such that, for all $x\in X$, $supp(\omega(s|x))\subseteq S_n$. Specializing this result to $\omega(s|x) = p^*_{t,n}(s|x) - r^*_{t,n}(s)$,

$$(\sum_{s\in S} < W_q(q^*_{t+\Delta_n,n}(\cdot,s),\lambda;\Delta_n),(p^*_{t,n}(s|\cdot) - r^*_{t,n}(s))q_t(\cdot) >) = \lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1} < \nabla C(\cdot),p^*_{t,n}(\cdot|\cdot) - r^*_{t,n}(\cdot) > . \quad (16)$$

Next, observe that, for all $\lambda\in[0,\kappa c^{-\rho})$,

$$\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s) - q_t||^2_{X,2} > 0.$$

Otherwise, $W(q^*_{t+\Delta_n,n}(\cdot,s),\lambda;\Delta_n) = W(q_t,\lambda;\Delta_n)$ for all $s\in S$, $C(\cdot)=0$, and the Bellman equation (equation (13)) could not be satisfied. It follows that the set

$$\hat{Q}_{t,n} = \{q\in\mathscr{P}(X): \sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s) - q||^2_{X,2} < 2\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s) - q_t||^2_{X,2} \& ||q-q_t||_{X,2} < \Delta_n\}$$

forms a non-empty open set containing q_t.

By the boundedness and convexity of the value function, Alexandrov's theorem holds, and, almost everywhere, the value function is first and second-order differentiable. Therefore, almost everywhere, and in particular for some $\hat{q}_{t,n} \in \hat{Q}_{t,n}$,

$$W(q^*_{t+\Delta_n,n}(\cdot,s),\lambda;\Delta_n) = W(\hat{q}_{t,n},\lambda;\Delta_n) + W_q(\hat{q}_t,\lambda;\Delta_n)(q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n})$$
$$+\frac{1}{2} < (q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n})|A|(q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n}) >$$
$$+o(||q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n}||^2_{X,2}),$$

for some symmetric Hessian matrix A. By the results of theorem 2.3 in Rockafellar [1999], we can write

$$W_q(q^*_{t+\Delta_n,n}(\cdot,s),\lambda;\Delta) = W_q(\hat{q}_{t,n},\lambda;\Delta_n) + < (q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n})|A| +$$
$$o(||q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n}||_{X,2}).$$

That is, the points of second-differentiability of the value function are also the points at which the gradient has a Taylor expansion. Plugging this into equation (16),

$$\sum_{s\in S} < W_q(\hat{q}_{t,n},\lambda;\Delta_n), q_t(\cdot)(p^*_{t,n}(s|x) - r^*_{t,n}(s)) > +$$
$$\sum_{s\in S} < (q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n})|A|q_t(\cdot)(p^*_{t,n}(s|\cdot) - r^*_{t,n}(s)) > +$$
$$\sum_{s\in S} < o(||q^*_{t+\Delta_n,n}(\cdot,s) - \hat{q}_{t,n}||_{X,2}), q_t(\cdot)(p^*_{t,n}(s|x) - r^*_{t,n}(s)) >=$$
$$\lambda\Delta^{1-\rho}C(\cdot)^{\rho-1} < \nabla C(\cdot), p^*_{t,n}(s|x) - r^*_{t,n}(s) >$$

Using Bayes' rule, this simplifies to

$$\sum_{s\in S} r^*_{t,n}(s) < (q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n})|A|(q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}) > +$$

$$\sum_{s\in S} r^*_{t,n}(s) o(||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,2}) =$$

$$\lambda \Delta_n^{1-\rho} C(\cdot)^{\rho-1} < \nabla C(\cdot), p^*_{t,n}(s|x) - r^*_{t,n}(s) > . \quad (17)$$

Using the Alexandrov formula and the Bellman equation,

$$W(\hat{q}_{t,n},\lambda;\Delta_n) - W(q_t,\lambda;\Delta_n) + \frac{1}{2}\sum_{s\in S} r^*_{t,n}(s) < (q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_t)|A|(q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_t) > +$$

$$\sum_{s\in S} r^*_{t,n}(s) o(||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,2}) = \frac{\lambda}{\rho}\Delta_n^{1-\rho} C(\cdot)^{\rho} + (\kappa - \lambda c^{\rho})\Delta_n.$$

By the Lipschitz-Continuity of $W(\cdot)$,

$$|W(\hat{q}_{t,n},\lambda;\Delta_n) - W(q_t,\lambda;\Delta_n)| \leq K||\hat{q}_{t,n} - q_t||_{X,2},$$

and therefore $W(\hat{q}_{t,n},\lambda;\Delta_n) - W(q_t,\lambda;\Delta_n) = o(\Delta_n)$. Moreover, note that by the definition of $\hat{Q}_{t,n}$,

$$\sum_{s\in S} r^*_{t,n}(s) o(||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,2}) = o(\sum_{s\in S} r^*_{t,n}(s) o(||q^*_{t+\Delta_n,n}(\cdot,s)-q_t||^2_{X,2}).$$

Substituting out the Hessian term using equation (17),

$$\lambda \Delta_n^{1-\rho} C(\cdot)^{\rho-1} (\frac{1}{2} < \nabla C(\cdot), p^*_{t,n}(s|x) - r^*_{t,n}(s) > -\frac{1}{\rho} C(\cdot)) +$$

$$o(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_t||^2_{X,2}) = (\kappa - \lambda c^{\rho})\Delta_n + o(\Delta_n).$$

By Taylor's theorem,

$$\lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1}\{(1-\rho^{-1})\frac{1}{2} < p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot)|\nabla^2 C(\cdot)|_{p_n^*=r_n^*}|p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot) > +$$
$$o(||q_t(\cdot)(p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot))||^2)\}+o(\sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s)-q_t||_{X,2}^2)=(\kappa-\lambda c^\rho)\Delta_n+o(\Delta_n).$$

By the results in the proof of corollary 1,

$$(< p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot)|\nabla^2 C(\cdot)|_{p_n^*=r_n^*}|p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot) >) =$$
$$\sum_{s\in S}\sum_{x'\in X}\sum_{x\in X} r_{t,n}^*(s)[q_{t+\Delta_n,n}^*(x,s)-q_t(x)][q_{t+\Delta_n,n}^*(x',s)-q_t(x)]\bar{k}(x,x';q_t).$$

By the properties of the k matrix, $\bar{k}(x,x';q_t)$ defines a norm $||\cdot||_{X,k}:\mathbb{R}^{|X|-1}\to\mathbb{R}^+$ on the tangent space of posteriors. That is,

$$(< p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot)|\nabla^2 C(\cdot)|_{p_n^*=r_n^*}|p_{t,n}^*(\cdot|\cdot)-r_{t,n}^*(\cdot) >) = \sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s)-q_t||_{X,k}^2.$$

Using these results,

$$\frac{1}{2}(1-\rho^{-1})\lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1}(\sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s)-q_t||_{X,k}^2+o(\sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s)-q_t||_{X,k}^2))$$
$$+o(\sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s)-q_t||_{X,k}^2)=(\kappa-\lambda c^\rho)\Delta_n+o(\Delta_n). \quad (18)$$

By the definition of little-o notation, this can be rewritten, given an arbitrary $\varepsilon \in (0,1)$, as

$$\frac{1}{2}(1-\rho^{-1})(1-\varepsilon)\lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1}\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-q_t||^2_{X,k}+$$

$$+o(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-q_t||^2_{X,k}) \leq (\kappa-\lambda c^\rho+\varepsilon)\Delta_n,$$

for all $n > n_\varepsilon$. Rescaling,

$$\frac{1}{2}(1-\rho^{-1})(1-\varepsilon)\lambda C(\cdot)^{\rho-1}\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}+$$

$$+\Delta_n^{\rho-1}o(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}) \leq (\kappa-\lambda c^\rho+\varepsilon)\Delta_n^\rho.$$

Using the Taylor expansion of the cost function again, for some $\hat{n}_\varepsilon \geq n_\varepsilon$, for all $n > \hat{n}_\varepsilon$,

$$C(\cdot) \geq (1-\varepsilon)\frac{1}{2}\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}.$$

Therefore,

$$(\frac{1}{2}(1-\varepsilon))^\rho(1-\rho^{-1})\lambda(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k})^\rho+$$

$$+\Delta_n^{\rho-1}o(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}) \leq (\kappa-\lambda c^\rho+\varepsilon)\Delta_n^\rho.$$

It follows that, for any $\xi > 0$, for n sufficiently large,

$$(\frac{1}{2}(1-\varepsilon))^{\rho}(1-\rho^{-1})\lambda(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k})^{\rho} \le$$

$$(\kappa-\lambda c^{\rho}+\varepsilon)(\Delta_n^{\rho}+\xi\Delta_n^{\rho-1}\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}.$$

Therefore, either

$$(\frac{1}{2}(1-\varepsilon))^{\rho}(1-\rho^{-1})\lambda(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k})^{\rho} \le 2(\kappa-\lambda c^{\rho}+\varepsilon)\Delta_n^{\rho}$$

or

$$(\frac{1}{2}(1-\varepsilon))^{\rho}(1-\rho^{-1})\lambda(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k})^{\rho} \le$$

$$2\xi\Delta_n^{\rho-1}\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k}.$$

Therefore,

$$\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-\hat{q}_{t,n}||^2_{X,k} \le B\Delta_n,$$

where

$$B = max((\frac{2(\kappa-\lambda c^{\rho}+\varepsilon)}{(\frac{1}{2}(1-\varepsilon))^{\rho}(1-\rho^{-1})\lambda})^{\rho^{-1}}, (\frac{2\xi}{(\frac{1}{2}(1-\varepsilon))^{\rho}(1-\rho^{-1})\lambda})^{\rho-1}) > 0.$$

By equation (15),

$$||q_t(\cdot)(p^*_{t,n}(\cdot|\cdot)-r^*_{t,n}(\cdot))||^2_2 \le B_2\Delta_n, \tag{19}$$

where B_2 is a positive constant.

This is exactly the result of lemma 1 for the static problem. The proof of corollary 2, which relies only on that lemma, shows that

$$p^*_{t,n}(s|x) = r^*_t(s) + \phi^*_{t,n}(s) + \Delta_{t,n}^{\frac{1}{2}} \tau^*_t(s|x) + o(\Delta_n^{\frac{1}{2}}),$$

and that

$$q^*_{t,n}(x|s) = q_t(x) + \Delta_n^{\frac{1}{2}} q_t(x) \frac{\tau^*_t(s|x)}{r^*_t(s)} + o(\Delta_n^{\frac{1}{2}}).$$

A.9 Proof of proposition 1

Let $n \in \mathbb{N}$ index the convergent sequence of optimal policies. Using the Taylor-expansion for the first order condition (as inequation (17) in the proof of lemma 3), at any point q_t at which W is twice-differentiable,

$$\sum_{s \in S} r^*_{t,n}(s) < (q^*_{t+\Delta_n,n}(\cdot,s) - q_t) | \nabla^2 W(q_t, \lambda; \Delta_n) | (q^*_{t+\Delta_n,n}(\cdot,s) - q_t) > +$$

$$o(\sum_{s \in S} r^*_{t,n}(s) || q^*_{t+\Delta_n,n}(\cdot,s) - q_t ||^2_{X,k}) =$$

$$\lambda \Delta_n^{1-\rho} C(\cdot)^{\rho-1} < \nabla C(\cdot), p^*_{t,n}(s|x) - r^*_{t,n}(s) > .$$

By the convergence results lemma 3, for all s such that $r^*_t(s) > 0$,

$$q^*_{t+\Delta_n,n}(x,s) - q_t(x) = \Delta_n^{\frac{1}{2}} z^*_t(x|s) + o(\Delta_n^{\frac{1}{2}}),$$

and therefore

$$\sum_{s \in S} r^*_{t,n}(s) || q^*_{t+\Delta_n,n}(\cdot,s) - q_t ||^2_{X,k} = \Delta_n \sum_{s \in S} r^*_t(s) || z^*_t(x|s) ||^2_{X,k} + o(\Delta_n)$$

and

$$o(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-q_t||^2_{X,k}) \subseteq o(\Delta_n).$$

By the Taylor expansion of $C(\cdot)$,

$$C(\{p^*_{t,n}\},q_t) = \frac{1}{2}\Delta_n \sum_{s\in S} r^*_t(s)||z^*_t(x|s)||^2_{X,k} + o(\Delta_n).$$

Rewriting this first-order condition,

$$\Delta_n(\sum_{s\in S} r^*_t(s) < z^*_t(\cdot|s)|\nabla^2 W(q_t,\lambda;\Delta_n)|z^*_t(\cdot|s) >) =$$

$$2\lambda\Delta_n(\frac{1}{2}\sum_{s\in S} r^*_t(s)||z^*_t(x|s)||^2_{X,k})^{\rho} + o(\Delta_n).$$

By equation (18) in the proof of lemma 3,

$$\frac{1}{2}(1-\rho^{-1})\lambda\Delta_n^{1-\rho}C(\cdot)^{\rho-1}(\sum_{s\in S} r^*_{t,n}(s)||q^*_{t+\Delta_n,n}(\cdot,s)-q_t||^2_{X,k}+o(\Delta_n)) = (\kappa-\lambda c^{\rho})\Delta_n + o(\Delta_n).$$

It follows that

$$(1-\rho^{-1})\lambda\Delta_n(\frac{1}{2}\sum_{s\in S} r^*_t(s)||z^*_t(x|s)||^2_{X,k})^{\rho} = (\kappa-\lambda c^{\rho})\Delta_n + o(\Delta_n),$$

which can be written as

$$\frac{\lambda}{\rho}\Delta_n^{-\rho}C(\cdot)^{\rho} = \frac{\kappa-\lambda c^{\rho}}{\rho-1} + o(1).$$

Define

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

Putting this together with the first-order condition,

$$(\sum_{s\in S} r_t^*(s) < z_t^*(\cdot|s)|\nabla^2 W(q_t,\lambda;\Delta_n)|z_t^*(\cdot|s) >) =$$

$$\theta \sum_{s\in S} r_t^*(s)||z_t^*(x|s)||_{X,k}^2 + o(\Delta_n),$$

which proves the result for the given convergent sequence of policies.

A.10 Proof of lemma 4

We being by assuming, without loss of generality, that the optimal policies are Markovian with respect to the state variable q_t. This implies that the covariance matrices $\Omega_{t,\Delta}$ are measurable with respect to the filtration generated by the state variable q_t.

By the uniform boundedness and convexity of the family of value functions $W(q,\lambda;\Delta_m)$, this family of value functions is uniformly Lipschitz-continuous. The existence of a convergent sub-sequence follows from the Arzela-Ascoli theorem (see also Rockafellar [1970] theorem 10.9).

Pass to this sub-sequence, which (for simplicity) we also index by m. The beliefs $q_{t,m}$ are a family of $\mathbb{R}^{|X|}$-valued stochastic processes, with $q_{t,m} \in \mathscr{P}(X)$ for all $t \in [0,\infty)$ and $m \in \mathbb{N}$. Construct them as RCLL processes by assuming that $q_{\Delta_m j+\varepsilon,m} = q_{\Delta_m j,m}$ for all m, $\varepsilon \in [0,\Delta_m)$, and $j \in \mathbb{N}$. By chapter 6, theorem 3.21, condition 1 in Jacod and Shiryaev [2013], and the boundedness of $q_{t,m}$, it follows that the laws of $q_{t,n}$ are tight. By Prokhorov's theorem (chapter 6, theorem 3.9 in Jacod and Shiryaev [2013]) , it follows that there exists a convergent sub-sequence. Index this sub-sequence by $n \in \mathbb{N}$.

By lemma 6 and equation (14), for any $q_{t,n} \in \mathscr{P}(X)$, we must have

$$\sum_{s\in S} r_{t,n}^*(s)||q_{t+\Delta_n,n}^*(\cdot,s) - q_{t,n}(\cdot)||_{X,2} \leq B\Delta^{\frac{\rho-1}{2\rho-1}}.$$

Therefore, for any $\varepsilon > 0$,

$$P^n(||q_{t+\Delta_n,n} - q_{t,n}||_{X,2} > \varepsilon) \leq \frac{B}{\varepsilon}\Delta_n^{\frac{\rho-1}{2\rho-1}},$$

where P^n denotes the probability measure associated with the stochastic process $q_{t,n}$. It follows from chapter 6, theorem 3.26, condition 3 in Jacod and Shiryaev [2013] that the processes $q_{t,n}$ are "C-tight," meaning that the laws of $q_{t,n}$ converge in measure to the law of some continuous stochastic process, which we denote q_t.

Under the optimal policy, the stochastic process $\Delta_n^{-1}C(\{p^*_{t,n}\}, q_{t,n})$ is L^ρ-integrable (otherwise, the agent would achieve negatively infinite utility). It follows from 5 that $\Delta_n^{-1}\Omega_{t,n}$ is L^ρ-integrable.

By Bayes' rule, the processes $q_{t,n}$ are martingales, and therefore q_t is also a martingale. By chapter 7, theorem 3.4 in Jacod and Shiryaev [2013], convergence in law implies that

$$\lim_{n\to\infty} \Delta_n^{-1}\Omega_{s,n} = \sigma_s\sigma_s^T,$$

where $\Delta_n^{-1}\Omega_{n,s}$ and $\sigma_s\sigma_s^T$ are the second modified characteristics of the semi-martingales (in this case, martingales) $q_{s,n}$ and q_s. By the closedness of L^ρ, σ_s is $L^{2\rho}$-integrable, and therefore square-integrable.

We construct a martingale with unit variance for its increments from the innovations in $q_{t,n}$, following Amin and Khanna [1994] (lemmas 3.1, 3.2, and 3.3, noting that the statement of lemma 3.3 contains a typo). The essence of the issue is that the conditional covariance matrix of $q_{t,n}$, $\Omega_{t,n}$, is not necessarily full rank. By the results in that paper, this martingale and $q_{t,n}$ jointly converge in measure to a Brownian motion B_t and q_t, where the Brownian motion is measurable with respect to the filtration generated by q_t.

Fixing q_0, the converse also holds– q_t is measurable with respect to the filtration generated by the Brownian motion (this follows from the Markov property of $\Omega_{s,n}$). By the martingale representation theorem,

$$q_t = q_0 + \int_0^t \sigma_s dB_s.$$

A.11 Proof of lemma 5

We begin by discussing the convergence of stopping times. Let τ_n^* denote the sequence of stopping times associated with the optimal policies. Note that, by the boundedness of the value function, we must have

$$E_0[\tau_n^*] \leq \bar{\tau},$$

for some weakly positive constant $\bar{\tau}$ and all n. It follows by the positivity of τ_n^* that 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.

Following the arguments of Amin and Khanna [1994], lemmas 3.1 through 3.5, we can construct the martingale described in the proof of lemma 4 so that the martingale, stopping times τ_n^*, and beliefs $q_{t,n}$ converge in measure to τ^*, B_t, and q_t. By the continuous mapping theorem, it follows that

$$\lim_{n\to\infty} E_0[\sum_{x\in X} q_{\tau_n^*,n}(x)(u(a^*(q_{\tau_n^*,n}),x) - \kappa\tau_n^*)] = E_0[\sum_{x\in X} q_{\tau^*}(x)(u(a^*(q_{\tau^*}),x) - \kappa\tau^*)].$$

We now construct a possibly sub-optimal policy by rescaling the optimal Markov policy,

$$\bar{p}_{t,n}(s|x) = \alpha_{t,n} p_{t,n}^*(s|x) + (1-\alpha_{t,n}) r_{t,n}^*(s),$$

to satisfy $C(\{\bar{p}_{t,n}\}, q) = \Delta_n(\rho \frac{\kappa - \lambda c^\rho}{\lambda(\rho-1)})^{\rho^{-1}}$. Such a policy exists by the strong convexity condition, the continuity of the cost function, and the strict positivity of $\Delta_n^{-1}C(\cdot)$ for the optimal policy. Note that $\lim_{n\to\infty}\alpha_{t,n}=1$.

Define stopping boundaries using the value functions. That is, define the closed set

$$Q_n = \{q \in \mathscr{P}(X) : W(q,\lambda;\Delta_n) \leq \sum_{x\in X} q(x)(u(a^*(q),x)\}.$$

Define the stopping times $\bar{\tau}_n$ as the first hitting time for the set Q_n. Consider the value function $\bar{W}(q_0,\lambda;\Delta_n)$ associated with this policy. By construction,

$$\lim_{n\to\infty} E_0[\sum_{x\in X} \bar{q}_{\bar{\tau}_n,n}(x)(u(a^*(\bar{q}_{\bar{\tau}_n,n}),x)-\kappa\bar{\tau}_n)] = E_0[\sum_{x\in X} q_{\tau^*}(x)(u(a^*(q_{\tau^*}),x)-\kappa\tau^*)].$$

Adopt the convention that after the stopping time, both the optimal signals $p^*_{t,n}$ and the possibly-suboptimal signals $\bar{p}_{t,n}$ are uninformative. It follows by optimality that we must have

$$\frac{\lambda}{\rho}E_0[\int_0^\infty \Delta_n^{-\rho}C(\{p^*_{s,n}\},q_{s,n})^\rho ds] \leq \frac{\lambda}{\rho}E_0[\int_0^\infty \Delta_n^{-\rho}C(\{\bar{p}_{s,n}\},\bar{q}_{s,n})^\rho ds].$$

By construction,

$$\frac{\lambda}{\rho}E_0[\int_0^\infty \Delta_n^{-\rho}C(\{\bar{p}_{s,n}\},\bar{q}_{s,n})^\rho ds] = \frac{\lambda}{\rho}E_0[\int_0^{\tau_n^*} \rho\frac{\kappa-\lambda c^\rho}{\lambda(\rho-1)}ds].$$

By Fatou's lemma and proposition 1,

$$\frac{\lambda}{\rho}E_0[\int_0^{\tau^*} \rho\frac{\kappa-\lambda c^\rho}{\lambda(\rho-1)}ds] \leq \lim_{n\to\infty}\frac{\lambda}{\rho}E_0[\int_0^\infty \Delta_n^{-\rho}C(\{p^*_{s,n}\},q_{s,n})^\rho ds].$$

It follows by the convergence of τ_n^* to τ^* that

$$\frac{\lambda}{\rho}E_0[\int_0^{\tau^*} \rho\frac{\kappa-\lambda c^\rho}{\lambda(\rho-1)}ds] = \lim_{n\to\infty}\frac{\lambda}{\rho}E_0[\int_0^\infty \Delta_n^{-\rho}C(\{p^*_{s,n}\},q_{s,n})^\rho ds].$$

Therefore,

$$W(q_0,\lambda) = \lim_{n\to\infty} W(q_0,\lambda;\Delta_n)$$
$$= E_0[\sum_{x\in X} q_{\tau^*}(x)(u(a^*(q_{\tau^*}),x) - \kappa\tau^*)] -$$
$$\lambda E_0[\int_0^{\tau^*} (\frac{\kappa - \lambda c^\rho}{\lambda(\rho-1)} - c^\rho)ds]].$$

It immediately follows that the value function is twice-differentiable anywhere the agent does not stop, and that

$$\frac{1}{2} tr[\sigma_t \sigma_t^T W_{qq}(q_t,\lambda)] = (\kappa - \lambda c^\rho)\frac{\rho}{\rho-1}.$$

By proposition 1 and lemma 4,

$$\frac{1}{2} tr[\sigma_t \sigma_t^T \bar{k}(\cdot,\cdot,q_t)] = (\rho\frac{\kappa - \lambda c^\rho}{\lambda(\rho-1)})^{\rho^{-1}}$$

For

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

the result follows.

A.12 Proof of theorem 4

By the proof of section A.11,

$$\frac{1}{2} tr[\sigma_t \sigma_t^T \bar{k}(\cdot,\cdot,q_t)] = (\rho\frac{\kappa - \lambda c^\rho}{\lambda(\rho-1)})^{\rho^{-1}}$$

for all $q_t \in \mathscr{P}(X)$, and

$$tr[\sigma_t \sigma_t^T W_{qq}(q_t,\lambda)] = tr[\sigma_t \sigma_t^T \bar{k}(\cdot,\cdot,q_t)].$$

Moreover, by the HJB equation, for any (possibly sub-optimal) policy that results in some covariance matrix $\hat{\sigma}_t \hat{\sigma}_t^T$,

$$tr[\hat{\sigma}_t \hat{\sigma}_t^T W_{qq}(q_t,\lambda)] \leq tr[\hat{\sigma}_t \hat{\sigma}_t^T \bar{k}(\cdot,\cdot,q_t)].$$

Therefore,

$$\bar{k}(\cdot,\cdot,q_t) \succeq W_{qq}(q_t,\lambda), \tag{20}$$

in the sense that the difference of those two matrices is positive semi-definite.

We define a function

$$\psi(q,q_0,\lambda) = W(q,\lambda) - W(q_0,\lambda) - W_q(q_0,\lambda)^T \cdot (q - q_0),$$

that is convex in q. Following the discussion in chapter 8.1 of Amari and Nagaoka [2007], we can use this convex function to define a metric g and a flat, affine connection $\nabla^{(1)}$, as well as a dually flat connection, $\nabla^{(-1)}$. The probability simplex and metric, $(\mathscr{P}(X),g)$, together form a statistical manifold, and with the aforementioned connections form a dually flat space. The canonical divergence (a Bregman divergence) associated with this dually flat space is

$$\begin{aligned} D(p||q) &= \psi(p,q_0,\lambda) - \psi(q,q_0,\lambda) - \psi_q(q,q_0,\lambda)^T(p-q) \\ &= W(p,\lambda) - W(q,\lambda) - W_q(q,\lambda)^T \cdot (p-q). \end{aligned} \tag{21}$$

We can define a dual coordinate system,

$$\eta(q) = \psi_q(q) = W_q(q,\lambda) - W_q(q_0,\lambda). \tag{22}$$

Define $M(q_0) \subseteq \mathscr{P}(X)$ as the set of beliefs reachable under the stochastic process q_t induced by the optimal policy, and stopping times τ^*. Let g_m be the metric on M induced by g. The pair $(M(q_0), g_m)$ form a sub-manifold of the aforementioned statistical manifold, $(\mathscr{P}(X), g)$. By the martingale property of beliefs, $(M(q_0), g_m)$ must be auto-parallel with respect to $(\mathscr{P}(X), g)$ and the connection $\nabla^{(1)}$; otherwise, $(M(q_0), g_m)$ would have non-zero embedding curvature and beliefs would not be martingales. It follows from theorem 1.1 of Amari and Nagaoka [2007] that $M(q_0)$ is an affine subspace of $\mathscr{P}(X)$.

By theorem 3.5 of Amari and Nagaoka [2007], $M(q_0)$ is a dually flat space with respect to g_m and the connections $\nabla_M^{(1)}$ and $\nabla_M^{(-1)}$, where those connections are the connections induced by $\nabla^{(1)}$ and $\nabla^{(-1)}$, respectively. By theorem 3.6 of Amari and Nagaoka [2007], there exists a canonical divergence defined on $(M(q_0), g_M, \nabla_M^{(1)}, \nabla_M^{(-1)})$, which we denote, for any $p,q \in M(q_0)$, $D_M(p||q)$. By equation 3.47 of Amari and Nagaoka [2007], for all $p,q \in M(q_0)$,

$$D_M(p||q) = D(p||q).$$

By lemma 5, for any z,z' in the tangent space of $M(q_0)$,

$$< z,z' >_{g_M} = \theta \sum_{x\in X} \sum_{x'\in X} \bar{k}(x,x',q) z(x) z'(x').$$

Consider a boundary point, $q_B \in M(q_0)$, satisfying

$$W(q_B,\lambda) = \sum_{x\in X} q_B(x) u(a^*(q_B),x).$$

Such a point exists by the integrability of τ^* (otherwise, the agent would never stop).

Let $\gamma : [0,1] \to M(q_0)$ define a $\nabla_M^{(1)}$-geodesic in $M(q_0)$ that connects q_0 to q_B. By theorem 1.1 of Amari and Nagaoka [2007] and the flatness of $M(q_0)$, this geodesic is a straight line. Moreover, it is also a geodesic with respect to $\nabla^{(1)}$, by the fact that $M(q_0)$ is auto-parallel. By equations 3.47 and 3.54 in Amari and Nagaoka [2007],

$$\begin{aligned}
D_M(q_B||q_0) &= \theta \sum_{x\in X}\sum_{x'\in X}(q_B(x)-q_0(x))(q_B(x')-q_0(x'))\int_0^1 (1-s)\bar{k}(x,x',(1-s)q_0+sq_B)ds. \\
&= \theta D_k(q_B||q_0) \\
&= D(q_B||q_0).
\end{aligned}$$

By the second proof method for the smooth-pasting condition in Peskir and Shiryaev [2006], that condition is necessary in our context. We have, for any boundary point q_B, the value matching condition,

$$W(q_B,\lambda) = \sum_x u(a^*(q_B),x)q_B(x),$$

and the smooth-pasting condition,

$$W_q(q_B,\lambda) = u(a^*(q_B),\cdot) + \omega(q_B).$$

By the definition of the dual coordinates (equation (22)),

$$\begin{aligned}
\eta(q_B) &= W_q(q_B,\lambda) - W_q(q_0,\lambda) \\
&= u(a^*(q_B),\cdot) + \omega(q_B) - W_q(q_0,\lambda).
\end{aligned}$$

and

$$\eta(q_0) = 0.$$

Note that, for any two boundary points q_B and q'_B such that $a^*(q_B) = a^*(q'_B)$,

$$(\eta(q_B) - \eta(q'_B))^T \cdot (q_B - q'_B) = 0,$$

and therefore (using the triangular relation, defined below), for each action $a \in A$, given q_0, at most one point q_a satisfying the smooth-pasting condition. Define

$$\eta_a = u(a,\cdot) - W_q(q_0,\lambda).$$

Pick an action a such that a boundary point q_B with $a^*(q_B) = a$ exists. Such an action exists by the finiteness of the stopping time. Using the definition of the divergence (equation (21)), for

$$W(q_0,\lambda) = \sum_x \{u(a,x)q_0(x) + (q_B(x) - q_0(x))(\eta_a(x,q_B) - \eta(x,q_0))\} - D(q_B||q_0).$$

By the triangular relation (theorem 3.7 in Amari and Nagaoka [2007]), for any p,

$$\sum_x \{(p(x) - q_0(x))(\eta(x,q_B) - \eta(x,q_0))\} - D(p||q_0) = D(q_0||q_B) - D(p||q_B).$$

By definition, $q_B = \arg\min_p D(p||q_B)$, and therefore

$$q_B = \arg\max_p \sum_x \{(p(x) - q_0(x))(\eta_a(x) - \eta(x,q_0))\} - D(p||q_0).$$

We can rewrite the value function as

$$W(q_0,\lambda) = \max_p W_q(q_0,\lambda)^T \cdot q_0 + [u(a,\cdot) - W_q(q_0,\lambda)]^T \cdot p - D(p||q_0).$$

For any point p, the geodesic between p and q_0 is a straight line, by the $\nabla^{(1)}$-flatness of space. Therefore, by equation 3.47 in Amari and Nagaoka [2007], using the summation convention,

$$D(p||q_0) = \int_0^1 (1-s)(p^i - q_0^i)(p^j - q_0^j) g_{ij}(sp + (1-s)q_0) ds.$$

By equation (20),

$$D(p||q_0) \leq D_k(p||q_0).$$

For any $p \in M(q_0)$, and in particular any point q_B that is an optimal stopping boundary, equality holds. Therefore,

$$q_B \in \arg\max_p \sum_x \{(p(x) - q_0(x))(\eta_a(x) - \eta(x,q_0))\} - D_k(p||q_0).$$

That is, suppose some other q' was the (strict) maximizer. Then for that q', we would have (using the summation convention)

$$\begin{aligned}
(q'^i - q_0^i)(\eta_i(q_B) - \eta_i(q_0)) - D_k(q'||q_0) &> \\
(q_B^i - q_0^i)(\eta_i(q_B) - \eta_i(q_0)) - D_k(q_B||q_0) &= \\
(q_B^i - q_0^i)(\eta_i(q_B) - \eta_i(q_0)) - D(q_B||q_0) &\geq \\
(q'^i - q_0^i)(\eta_i(q_B) - \eta_i(q_0)) - D(q'||q_0) &\geq \\
(q'^i - q_0^i)(\eta_i(q_B) - \eta_i(q_0)) - D(q'||q_0), &
\end{aligned}$$

a contradiction. Moreover, by the equality of the divergences on the manifold $M(q_0)$,

$$W(q_0) = \max_{p \in \mathscr{P}(X)} W_q(q_0,\lambda)^T \cdot q_0 + [u(a,\cdot) - W_q(q_0,\lambda)]^T \cdot p - \theta D_k(p||q_0).$$

Let p_a^* denote the maximizer of the right-hand side of this equation, given a. Because this equation must hold for all a such that a stopping boundary associated with that action exists, it follows that for any action a, if there exists another action a' with

$$[u(a',\cdot) - W_q(q_0,\lambda)]^T \cdot p_{a'}^* - \theta D_k(p_{a'}^*||q_0) > [u(a,\cdot) - W_q(q_0,\lambda)]^T \cdot p_a^* - \theta D_k(p_a^*||q_0), \tag{23}$$

then such an action a cannot be taken at any optimal stopping boundary.

Define $A(q_0) \subseteq A$ as the set of actions which an optimal stopping boundary associated with that action exists. By the requirement that beliefs are martingales, the integrability of the stopping time, and the optional stopping theorem, there exists a $\pi^*(a)$ with support on $A(q_0)$ such that

$$\sum_{a \in A} \pi^*(a) p_a^*(x) = q(x)$$

for all $x \in X$. By the envelope theorem, for any $a \in A(q_0)$,

$$0 = W_{qq}(q_0,\lambda) \cdot (q_0 - p_a^*) - D_{k,q}(p_a^*||q_0),$$

where $D_{k,q}(p||q)$ is the Frechet derivative of $D_k(p||q)$ with respect to q. Such a derivative exists by the differentiability of the matrix k. Therefore,

$$\sum_{a \in A} \pi^*(a) D_{k,q}(p_a^*||q_0) = 0.$$

Consider the problem

$$\tilde{W}(q_0,\lambda) = \min_{\omega\in\mathbb{R}^{|X|}} \sum_{a\in A} \pi^*(a)\{\omega^T \cdot q_0 + [u(a,\cdot)-\omega]^T \cdot p_a - \theta D_k(p_a^*||q_0)\}.$$

By the envelope theorem,

$$\tilde{W}_q(q_0,\lambda) = \omega,$$

and therefore

$$W(q_0,\lambda) = \max_{\{p_a\in\mathscr{P}(X)\}} \min_{\omega\in\mathbb{R}^{|X|}} \sum_a \pi^*(a)\{\omega^T \cdot q_0 + [u(a,\cdot)-\omega]^T \cdot p_a - \theta D_k(p_a||q_0)\}.$$

By the equality of this function for each action in $A(q_0)$, and equation (23),

$$\pi^* \in \arg\max_{\pi\in\mathscr{P}(A)} \max_{\{p_a\in\mathscr{P}(X)\}} \min_{\omega\in\mathbb{R}^{|X|}} \sum_a \pi(a)\{\omega^T \cdot q_0 + [u(a,\cdot)-\omega]^T \cdot p_a - \theta D_k(p_a||q_0)\},$$

and therefore

$$W(q_0,\lambda) = \max_{\pi\in\mathscr{P}(A)} \max_{\{p_a\in\mathscr{P}(X)\}} \min_{\omega\in\mathbb{R}^{|X|}} \sum_a \pi(a)\{\omega^T \cdot q_0 + [u(a,\cdot)-\omega]^T \cdot p_a - \theta D_k(p_a||q_0)\}. \tag{24}$$

The problem

$$\bar{W}(q_0,\lambda) = \max_{\pi\in\mathscr{P}(A)} \max_{\{p_a\in\mathscr{P}(X)\}} \sum_a \pi(a)\{\sum_{x\in X} u(a,x)p_a(x) - \theta D_k(p_a||q_0)\}$$

subject to

$$\sum_{a\in A} \pi(a)(q_0(x)-p_a(x)) = 0$$

has entirely affine constraints, and therefore the KKT conditions are necessary. It follows that the solution to this problem is the solution to its Lagrangian, equation (24).

We next demonstrate equality of the primal and dual. Define λ^* by

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

or equivalently

$$\lambda^* = \frac{\kappa}{\rho c^\rho}.$$

For this value of λ,

$$W(q_0, \lambda^*) = E_0[\sum_{x \in X} q_{\tau^*}(x) u(a^*(q_{\tau^*}), x) - \kappa \tau^*],$$

and the constraint is satisfied.

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 by n). By the standard duality inequalities,

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

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 $V(q_0) = W(q_0, \lambda^*)$.

We can define

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

Note that every convergent sub-sequence of $V(q_0;\Delta_n)$ converges to the same function. By the boundedness of value function, it follows that

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