Online Appendix of Hébert and Woodford, “Neighborhood-Based Information Costs”

A Neighborhoods with Constant Marginal Costs **2**

B Security Design with Fisher Information **5**

B.1 Security Design with Neighborhood-Based Cost Functions 9

C Proofs **19**

C.1 Proof of Proposition 1 19

C.2 Proof of Proposition 2 24

C.3 Proof of Lemma 1 25

C.4 Proof of Proposition 3 25

C.5 Proof of Lemma 2 29

C.6 Proof of Corollary 2 33

C.7 Proof of Proposition 4 35

C.8 Proof of Corollary 3 41

C.9 Proof of Corollary 4 42

C.10 Proof of Lemma 3 45

C.11 Proof of Proposition 5 47

C.12 Proof of Proposition 6 48

D Technical Appendix **52**

D.1 Convergence to the Continuous State Model 52

D.2 Additional Technical Lemmas for Proposition 1 and Corollary 4 . . 56

D.3 Additional Technical Lemmas for Binary Choice 61

D.4 Additional Definition and Lemmas for Convergence 67

A Neighborhoods with Constant Marginal Costs

In this appendix section, we discuss an alternative assumption, described by Pomatto, Strack and Tamuz (2020) as "constant marginal costs," the leads to a somewhat different local information cost, a version of the "total information" measure of Bloedel and Zhong (2020). We first define what it means for a local information cost to exhibit constant, increasing, or decreasing marginal costs. Our definition follows Pomatto, Strack and Tamuz (2020).

Consider two signal structures, p^1 and p^2. Define $p^1 \otimes p^2$ as the signal structure associated with receiving both signals, under the assumption that the signal realizations are independent conditional on the state $x \in X$. That is, $(p^1 \otimes p^2)_{s_1 s_2, x} = p^1_{s_1,x} p^2_{s_2,x}$. A DM who receives the signal structure $p^1 \otimes p^2$ can be thought of as observing both signals simultaneously or as sequentially receiving one signal and then the other (the equivalence of these two interpretations follows from uniform posterior separability).

We will say that a cost function exhibits increasing/constant/decreasing marginal costs if receiving both signals is more/equally/less costly than the sum of the costs of receiving the signals separately.

Definition 3. A cost function exhibits increasing marginal costs if, for all signal structures p^1, p^2 and all priors q_0,

$$C(p^1 \otimes p^2, q_0; S, X) \geq C(p^1, q_0; S, X) + C(p^2, q_0; S, X).$$

A cost function exhibits decreasing marginal costs if, for all signal structures p^1, p^2 and all priors q_0,

$$C(p^1 \otimes p^2, q_0; S, X) \leq C(p^1, q_0; S, X) + C(p^2, q_0; S, X).$$

A cost function exhibits constant marginal costs if it exhibits both increasing and decreasing marginal costs.

Note that neighborhood-based cost functions, by Assumption 2, always exhibit constant marginal costs with respect to signal structures p^1 and p^2 that provide information about states without any neighborhoods in common. When instead p^1 and p^2 both discriminate between states within some neighborhood, whether the cost function exhibits increasing, decreasing, or constant marginal costs is governed by the nature of the local information cost in that neighborhood. Note also that a cost function might not exhibit decreasing, increasing, or constant marginal costs, if none of the above inequalities holds for all p^1, p^2, and q_0.

Intuitively, there is a connection between whether marginal costs are increasing or decreasing and the curvature of the information cost function. It is well-known that using Shannon's entropy leads to decreasing marginal costs, and in some applications of rational inattention this can lead to non-concavities and problems with equilibrium existence (Myatt and Wallace (2011)). Perhaps unsurprisingly, more curved cost functions can lead instead to increasing marginal costs. In the lemma below, we restate the familiar result for Shannon's entropy ($\rho = 1$ using the generalized entropy index), and show that with $\rho = 2$ the generalized entropy neighborhood-based cost functions exhibit increasing marginal costs.

Lemma 3. *The generalized entropy neighborhood-based cost function with $\rho = 1$, $H_{NG}(q;1,X,\mathscr{N})$, exhibits decreasing marginal costs. The generalized entropy neighborhood-based cost function with $\rho = 2$, $H_{NG}(q;2,X,\mathscr{N})$, exhibits increasing marginal costs.*

Proof. See the Appendix, Section C.10. $\square$

Based on these results, it is tempting to speculate that $H_{NG}(q;\rho,X,\mathscr{N})$ will exhibit constant marginal costs for some $\rho \in (1,2)$. Instead, we show that it is the sum of the $\rho = 1$ and $\rho = 2$ generalized entropy neighborhood-based cost functions that exhibits constant marginal costs. This property arises because information costs in this case are linear in the prior, implying (under the UPS assumption) that the expected cost of the receiving the signals sequentially is exactly equal to the cost of receiving them simultaneously.

Proposition 5. *Suppose a neighborhood-based cost function $H(q;X,\mathscr{N})$ exhibits constant marginal costs. Then the local information costs are proportional to, for* *all q_i in the interior of the simplex,*

$$H^{CM}(q_i;|X_i|) = H^G(q_i;1,|X_i|) + H^G(q_i;2,|X_i|),$$

which simplifies to

$$H^{CM}(q_i;|X_i|) = \frac{1}{|X_i|}\sum_{j=1}^{|X_i|}\sum_{k=1}^{|X_i|} q_{i,j}\ln(\frac{q_{i,j}}{q_{i,k}}).$$

The static rational inattention problem can be written as

$$\begin{aligned} V_{CM}(q;X,\mathscr{N}) = \max_{\{p_x\in\mathscr{P}(S)\}_{x\in X}} & \sum_{s\in S}\pi_s(p,q_0)\hat{u}(q_s(p,q)) \\ & - \theta\sum_{i\in\mathscr{I}} c_i\bar{q}_i \sum_{x\in X_i}\sum_{x'\in X_i\setminus\{x\}} q_x D_{KL}(p_x||p_{x'}), \end{aligned}$$

where $\{c_i\in\mathbb{R}_+\}_{i\in I}$ are positive constants.

Proof. See the Appendix, Section C.11. The proof builds on results in Bloedel and Zhong (2020). □

Bloedel and Zhong (2020) characterize the set of UPS cost functions with constant marginal costs, which they call total information costs. This family includes the neighborhood-based cost functions with constant marginal costs (because the neighborhood based costs functions are uniformly posterior-separable). They show that any UPS cost function with constant marginal costs must satisfy

$$H^{TI}(q;X) = \sum_{x\in X}\sum_{x'\in X\setminus\{x\}} \gamma_{x,x'} q_x \ln(\frac{q_x}{q_{x'}}) \tag{19}$$

for some non-negative constants $\gamma_{x,x'}$.[28]

It is immediately apparent that any total information costs with $\gamma_{x,x'} = \gamma_{x',x}$ can

[28]This class of cost functions is also a special case of the class of LLR cost functions defined by Pomatto, Strack and Tamuz (2020), a larger family of cost functions that exhibit constant marginal costs. (See their Proposition 8.) Thus there is a non-empty intersection between our neighborhood-based costs and the class of LLR cost functions defined by Pomatto, Strack and Tamuz (2020), though neither class is entirely contained in the other.

be interpreted as a neighborhood-based information cost. The simplest way to do this is by defining the set of neighborhoods $\{X_i\}$ to be the set of all pairwise combinations of states in X, in which case $c_i = 2\gamma_{x,x'}$ for the states $\{x, x'\} = X_i$. These c_i constants are by definition non-negative, and hence satisfy the only restrictions required for the neighborhood-based cost functions.

Constant marginal costs is an appealing assumption if the signal structures p^1 and p^2 are interpreted as experiments (as in Pomatto, Strack and Tamuz (2020)), because it seems natural to assume that if each experiment has a cost, doing both experiments should have a cost equal to the sum of the two costs. Under this interpretation, the constant marginal costs property can be thought of as "constant returns to scale."

However, the constant marginal costs assumption also pins down the curvature of the local information costs. An immediately corollary of Proposition 5 and (8) is that the elasticity of choice probabilities to incentives assuming constant marginal costs will be a weighted average of the elasticity with the $\rho = 2$ generalized entropy index and the Shannon entropy case ($\rho = 1$). Although this elasticity will be lower than the elasticity in the Shannon entropy case, it will be higher than the elasticity for the generalized entropy index with $\rho \approx 13$, which was found by Dean and Neligh (2019) to best describe their data. Moreover, it is not a priori obvious that marginal costs should be constant in the context of discriminating between neighboring states. It is intuitive that a DM might find it difficult to make sharp distinctions between neighboring states, and that the marginal difficulty of such distinctions might increase the more precisely the DM attempts to discriminate between the states. That is, local information costs might be characterized by increasing, and not constant, marginal costs.

B Security Design with Fisher Information

In this appendix section we consider the security design model with adverse selection in Yang (2020),[29] which builds on the results concerning binary choice de-

[29]Our neighborhood cost function could also be applied in the same fashion to the model of security design with moral hazard in attention described in the appendix of Hébert (2018).

scribed in the main text. The purpose of this application is two-fold. First, we illustrate another distinction between mutual information and Fisher information, concerning kinked payoffs rather than discrete jumps in payoffs. Second, we illustrate how our results on binary choice described previously can be incorporated into other problems.

Let $X = [0, \bar{x}]$ be the value of some assets, and let $s : X \to \mathbb{R}_+$ be a security that offers a payoff of $s(x)$ when the underlying assets value is x. Consider the problem of a risk-neutral buyer (the DM) who is presented with a take-it-or-leave-it offer of this security at a price K. The actions available to the buyer are to accept or decline this offer, $A = \{accept, decline\}$. Normalizing the utility of declining the offer to zero, the utility of accepting the security s at price K is

$$u_{accept}(x; s, K) = s(x) - K.$$

The seller and buyer share a common prior $q \in \mathscr{P}_{LipG}(X)$.

The problem facing this buyer fits exactly into the binary choice framework. As a result, we can determine whether the buyer will always decline, always accept, or pursue an interior strategy using the results of Proposition 3 (for Fisher information) and Woodford (2008)/Yang (2020) (for mutual information and variants thereof).

Let us suppose the security being offered is a debt security,

$$s(x; x^*) = \min\{x, x^*\}$$

for some $x^* \in (0, \bar{x})$. This security is continuous; by the results of Proposition 3, with the Fisher information cost the buyer will optimally choose a continuously twice-differentiable (and hence continuously differentiable) probability of acceptance $p^*_{accept,FI}(x)$, assuming the buyer chooses to gather some information. In contrast, by the results of Woodford (2008), with mutual information and again assuming the buyer gathers some information,

$$p^*_{accept,MI}(x) = \frac{\pi \exp(\theta^{-1} \min\{x, x^*\} - \theta^{-1} K)}{1 - \pi + \pi \exp(\theta^{-1} \min\{x, x^*\} - \theta^{-1} K)}$$

for some $\pi \in (0, 1)$. This probability of acceptance is kinked, and is in fact constant for all $x \geq x^*$, a property it inherits from the payoff function. The particular prediction that $p^*_{accept,MI}(x)$ is constant on $x \geq x^*$ is both testable and familiar from the regime change and perceptual experiment examples discussed in the main text, and is rejected in experimental evidence. More generally, mutual information and Fisher information will generate different predictions in all binary choice problems with kinked but continuous payoffs– this difference is not specific to security design.

Let us now turn to the problem of a security designer/seller who wishes to design s and then offer it to the buyer described above at the price K. This problem is analyzed by Yang (2020), who shows that when the buyer's information cost is proportional to mutual information,[30] it is optimal for the seller to offer a debt security. This result holds regardless of whether it is optimal for the seller to induce the buyer to always accept the security or induce the buyer to gather information.

We will first discuss the case in which the seller wishes to avoid information gathering by the buyer. To motivate trade, we follow Yang (2020) and assume that the seller retains whatever asset value is not sold to the buyer, $x - s(x)$, and discounts these cashflows at a rate of $\beta \in (0,1)$. Let S be the class of feasible security designs; for this application, we will require that securities satisfy limited liability and be "doubly monotone," meaning that $s(x)$ and $x - s(x)$ are both non-negative, non-decreasing functions of x. Note that this assumption is common in the security design literature but is not imposed by Yang (2020).

The problem of the seller is to choose s and K to maximize her payoff, subject to the constraint that the buyer does not acquire information (characterized in Proposition 3):

$$\max_{s \in S, K \in \mathbb{R}} \int_X q(x)(K - \beta s(x))dx$$

[30] Yang (2020) proves this result for a general class of state-separable information costs that includes mutual information but does not include Fisher information.

subject to

$$\inf_{p_L \in \{p \in C^1(X,(0,\infty)) : \int_X q(x)p(x)dx = 1\}} \int_X q(x) p_L(x)(s(x) - K)dx + \frac{\theta}{4} \int_X q(x) \frac{(p_L'(x))^2}{p_L(x)} dx \geq 0.$$

It is immediately apparent that the constraint will bind; if it did not, the seller could increase the offering price K until the constraint was binding. Consequently, we can use the results of Lemma 2 to define a necessary and sufficient condition for the buyer to choose not to acquire information.

The resulting security design problem can be analyzed using Hamiltonian methods. Because the s must be doubly monotone and $s(0) = 0$ by limited liability, we can think of $s'(x) = v(x)$ as the control variable. Using the results of Lemma 2, the state vector $(s(x), \psi(x))$ must evolve as

$$\frac{d}{dx}(\begin{bmatrix} s(x) \\ \psi(x) \end{bmatrix}) = \begin{bmatrix} v(x) \\ \frac{1}{2\theta}(s(x) - K) + \frac{1}{4}\psi(x)^2 + \frac{q'(x)}{q(x)}\psi(x) \end{bmatrix}.$$

The Hamiltonian, treating the price K as given, is

$$H(s, \psi, v, \lambda_1, \lambda_2, x; K) = q(x)(K - \beta s) + \lambda_1 v + \lambda_2 (\frac{1}{2\theta}(s - K) + \frac{1}{4}\psi^2 + \frac{q'(x)}{q(x)}\psi),$$

noting that the choice of v is restricted to the interval $[0, 1]$. The relevant boundary conditions are $\psi(0) = \psi(\bar{x}) = 0$ (from Lemma 2), $s(0) = 0$ (from limited liability), and the free boundary condition associated with $s(\bar{x})$, $\lambda_1(\bar{x}) = 0$. The full problem must also consider the optimal price,

$$K^* \in \arg\max_{K \in \mathbb{R}} \int_X H(s^*(x; K), \psi^*(x; K), v^*(x; K), \lambda_1^*(x; K), \lambda_2^*(x; K); K)dx.$$

Given this description of the problem, it is straightforward to numerically compute the optimal security design. However, for this particular Hamiltonian system, we are able to analytically characterize the optimal security design.

Proposition 6. *Suppose the buyer's cost of information acquisition is proportional to the Fisher information cost function. The optimal doubly-monotonic, limited liability security design of seller who wishes to avoid information acquisition is a debt security,* $s^*(x) = \min\{x, x^*\}$ *for some* $x^* \in (0, \bar{x}]$.

Proof. See the Appendix, section C.12. □

This proposition demonstrates that the results of Yang (2020) for mutual information are robust to using the Fisher information cost under the additional restriction that the security be doubly-monotonic.

The Hamiltonian approach outlined here can be readily modified to cover the case without the double-monotonicity requirement (in which $s(x)$ is a control variable instead state variable) as well as the case in which the buyer acquires information (in which $p(x)$ and $p'(x)$ replace $\psi(x)$ as state variables). The security design problem can also be analyzed on a discrete state space, using one of the neighborhood-based cost functions described in the text. In the next appendix subsection, we quantitatively analyze the some of the other cases discussed in Yang (2020) (when the seller induces information acquisition by the buyer, and without monotonicity constraints) in the discrete state case.

B.1 Security Design with Neighborhood-Based Cost Functions

In this appendix section, we numerically analyze the security design problem described above. For this section, we will use a finite state space, instead of a continuous one. The purpose of this section is to show that neighborhood-based cost functions remain tractable (at least computationally) in this application. We will briefly summarize the environment for the discrete state case.

Let X be a finite set of states. A seller offers a security $s \in \mathbb{R}_+^{|X|}$, whose payoffs are contingent on the realized value of the assets backing the security, $x \in X \subset \mathbb{R}_+$, to a buyer at a price K. The buyer's problem is to gather information about which asset values $x \in X$ are most likely and then accept ("like," L) or reject (R) this take-it-or-leave it offer. Both parties are risk-neutral, and the seller discounts the cashflows by a factor $\beta \in (0,1)$, relative to the buyer. The security is constrained by limited liability, $0 \leq s_x \leq x$. Let S_{LL} be the set of limited liability securities, and let $S \subset S_{LL}$ be the set of limited liability that are doubly monotone (as described above). The seller designs the security and offers a price,

$$\max_{s \in S_{LL}, K \in \mathbb{R}} \pi_L(s,K) q_L(s,K)^T (K\iota - \beta s),$$

where ι is a vector of ones, possibly subject to the monotonicity constraint $s \in S$. In this expression, $\pi_L(s,K)$ and $q_L(s,K)$ are the optimal policies of the buyer who solves the rational inattention problem of (1), with $A = \{L, R\}$,

$$\begin{aligned} V(q_0; s, K) = & \max_{\pi_L \in [0,1], q_L, q_R \in \mathscr{P}(X)} \pi_L q_L^T \cdot (s - K\iota) \\ & - \theta \pi_L D_H(q_L || q_0) - \theta (1 - \pi_L) D_H(q_R || q_0), \end{aligned}$$

subject to the constraint that $\pi_L q_L + (1 - \pi_L) q_R = q_0$.

We explore, numerically, how the result of Yang (2020) on the optimality of debt with the mutual information cost function changes with alternative Bregman divergence cost functions (which are defined by the $H(\cdot)$ functions). We consider three alternatives, a generalized entropy index neighborhood-based function (Definition 2) with a pairwise neighborhood structure (as in section 2.3), a generalized entropy index cost function (i.e. a neighborhood cost function with only one neighborhood), and a "weighted" Shannon's entropy. Weighted Shannon's entropy is

$$H_w(q) = \sum_{x \in X} (e_x^T w)(e_x^T q) \ln(\frac{e_x^T q}{\iota^T q}),$$

where w is a vector of weights. Constant weights correspond to Shannon's entropy.

Summarizing our results, we replicate numerically the proof of Yang (2020) that, with mutual information, the optimal security design is always a debt. In contrast, for weighted mutual information and the generalized entropy index, the shape of the security design depends on the weights and the prior, respectively. The neighborhood cost function, on the other hand, appears to always generate the same shape irrespective of the prior.

Below, we describe our calculation procedure, and the parameters we use to generate figures 6 and 7 below, which show the optimal securities when s is not and is required to be doubly monotone, respectively. Our choice of parameters is guided by a desire to illustrate the differences between the cost functions, and to ensure that acceptance is not certain ($\pi_L < 1$). Our numerical calculation uses the first-order approach,[31] solving

$$\max_{s\in S_{LL}, K\in\mathbb{R}, \pi_L\in[0,1], q_L\in\mathscr{P}(X)} \pi_L q_L^T (K\iota - \beta s)$$

subject to the buyer's first order condition and that beliefs remain in the simplex,

$$\begin{aligned} s - K\iota + \theta H_q(q_0 - \pi_L q_L) &= \theta H_q(\pi_L q_L), \\ e_x^T (q_0 - \pi_L q_L) &\geq 0, \forall x \in X. \end{aligned}$$

and the monotonicity constraints (if applicable). Combining the first-order conditions of this security design problem and the limited liability constraints,

$$\begin{aligned} (1-\beta)s^* = {} & \theta H_q(q - \pi_L^* q_L^*) - \theta H_q(\pi_L^* q_L^*) + \\ & + \theta[H_{qq}(q - \pi_L^* q_L^*) + H_{qq}(\pi_L^* q_L^*)](\beta \pi_L^* q_L^* - \lambda + \nu), \end{aligned}$$

where λ and ν are the multipliers on the limited liability constraints. This illustrates that the optimal security design is determined by the H function, subject to the caveat that $\pi_L^* q_L^*$ is endogenous.

Our numerical experiment uses an X with twenty-one states, with values of x evenly spaced from 0 to 10. We use a seller β of 0.5, and prior q that is an equal-weighted mixture of a uniform and binomial (21 outcomes of a 50-50 coin flip) distribution. We have chosen these parameters to help illustrate the differences between the cost functions.[32]

[31] We conjecture, but have not proven, that the first-order approach is valid in this context.

[32] In particular, the effects of weighted vs. standard Shannon's entropy are proportional to $\ln(\beta)$, so we choose a value of β significantly different from one. The differences between the generalized entropy index and Shannon's entropy disappear with a uniform prior, so we use the binomial part of the prior to highlight those differences. At the same time, it is helpful for numerical purposes to ensure the prior is significantly different from zero in each state, which is why we have the uniform

For the generalized entropy and neighborhood-based cost functions, we use $\rho = 13$. This value is close to the estimated parameter of Dean and Neligh (2019) for these two cost functions, although there is no particular reason to apply parameters estimated for perceptual experiments to security design. The various cost functions are not of the same "scale," so the same values of θ do not necessarily result in the securities of the same scale. We have chosen $\theta = \frac{1}{2}$ for Shannon's entropy, $\theta = 1$ for weighted Shannon's entropy and the neighborhood cost function, and $\theta = \frac{1}{50}$ for the generalized entropy function, which results in securities that are of the same scale but distinct in our graphs. For our weighted Shannon's entropy, we use

$$w(x) = \frac{3}{2} + \frac{x}{10}.$$

This linear weight structure assumes that it is more costly for the buyer to learn about good states than about bad states. We will see that this induces the seller to offer the buyer more in good states, and hence makes the buyer's security more equity-like. The more general point is that almost any security design could be reverse-engineered as optimal given some weight matrix. This reinforces the need to consider what kinds of information costs are reasonable.

Our numerical results are shown in figures 6 and 7. The first of these shows the optimal security designs, the second the optimal doubly monotone security designs. Our numerical calculations recover the result of Yang (2020) for the case of Shannon's entropy. They also illustrate our point that, with upward-sloping weights, the result for weighted Shannon's entropy is equity-like. The "inverse hump-shape" of the optimal security with the generalized entropy index cost function is caused by the "hump-shape" of the prior.[33] The optimal securities for mutual information and weighted mutual information are monotone, and hence do not differ between the two graphs, whereas the optimal securities for the neighborhood based cost function and (imperceptibly) the generalized entropy index are non-monotone, and hence do differ. For weighted mutual information and the generalized entropy index, monotonicity or a lack thereof is not guaranteed, as the shape of the optimal

part of the prior.

[33] With a uniform prior, the optimal security with the generalized entropy index cost is also a debt.

security depends on the weights and prior, respectively.

Our results for the neighborhood cost function appear, regardless of parameters, to result in the same "debt-like," but non-monotone, optimal security. This security is non-monotone and rapidly changing in one area. Rapid changes in security values would cause rapid changes in buyer behavior with Shannon's entropy, and hence be sub-optimal, but this is not the case with neighborhood cost functions. As a result, it is possible for the optimal security to have rapid changes.[34] However, when we restrict the security to be monotone, the optimal security is a debt, suggesting that the result of Yang (2020) is robust to using neighborhood cost functions (but not the other two alternatives) under this additional restriction. This is consistent with our result in the previous appendix section, which shows the optimality of debt among monotone securities for the acceptance with certainty case on a continuous state space. We discuss this case with a discrete state space next.

Suppose the seller designs the security to induce the buyer to accept with probability one. In other words, the buyer's "consideration set" in his rational inattention problem consists only of L, instead of both L and R. As mentioned above, we have chosen the parameters of our numerical example to ensure that, for all of the cost functions, the seller is better off inducing information acquisition ($\pi_L < 1$) than avoiding information acquisition ($\pi_L = 1$). Note that the $\pi_L = 0$ case is equivalent to trading a "nothing" security at zero price, and hence assuming $\pi_L > 0$ is without loss of generality.

We will begin by restating the acceptance with certainty problem for the discrete state case (the problem for the continuous state case is described in the text). Consider the buyer's problem,

$$V(q;s,K) = \max_{\pi_L \in [0,1], q_L, q_R \in \mathscr{P}(X)} \pi_L q_L^T (s - K\iota) \\ - \theta \pi_L D_H(q_L || q) - \theta(1-\pi_L) D_H(q_R || q),$$

[34] Sharp-eyed readers might notice a second feature of the optimal security for neighborhood-based cost functions: the "flat" part isn't exactly flat. This feature arises from the "tri-diagonal" nature of the information cost matrix function $k(q)$, which leads to a difference equation describing the optimal security. As the number of states increases, the "flat" part of the security becomes increasingly flat. In the continuous state case, the difference equation becomes a differential equation, and we conjecture that the flat part is truly flat.

subject to the constraint that $\pi_L q_L + (1-\pi_L) q_R = q$. Rewrite the choice variables as $\hat{q}_L = \pi_L q_L$ and $\hat{q}_R = (1-\pi_L) q_R$, and use the homogeneity of the H function, so that the problem is

$$\begin{aligned} V(q;s,K) = & \max_{\hat{q}_L, \hat{q}_R \in \mathbb{R}_+^{|X|}} \hat{q}_L^T (s - K\iota) \\ & - \theta D_H(\hat{q}_L || q) - \theta D_H(\hat{q}_R || q), \end{aligned}$$

subject to $\hat{q}_L + \hat{q}_R = q$. Observe that the objective is concave and the constraints linear, so it suffices to consider local perturbations.

Suppose that it is optimal to set $\pi_L = 1$, implying $\hat{q}_L = q$. Consider a perturbation to $\hat{q}_L = q - \varepsilon q_R$, $\hat{q}_R = \varepsilon q_R$, for any arbitrary $q_R \in \mathscr{P}(X)$. For such a perturbation to reduce utility, we must have

$$-\varepsilon q_R^T (s - K\iota) - \theta D_H(q - \varepsilon q_R || q) - \theta \varepsilon D_H(q_R || q) \leq 0.$$

Taking the limit as $\varepsilon \to 0^+$, we must have, for all q_R, and hence for the minimizer,

$$\min_{q_R \in \mathscr{P}(X)} q_R^T (s - K\iota) + \theta D_H(q_R || q) \geq 0.$$

Note that this condition closely resembles the problem for the continuous state case above (Proposition 3).

If this condition is satisfied, it is at least weakly optimal for the buyer to choose $\pi_L = 1$ and gather no information. Consequently, the Lagrangian version of the optimal security design problem, subject to the constraint of inducing no information acquisition, is

$$\max_{\{s \in \mathbb{R}_+^{|X|}, K \geq 0\}} \min_{\{\lambda \geq 0, q_R \in \mathscr{P}(X), \omega \in \mathbb{R}_+^{|X|}\}} q^T (K\iota - \beta s) + \lambda (q_R^T (s - K\iota) + \theta D_H(q_R || q)) + \omega^T (v - s),$$

where λ is the multiplier on the no-information-gathering constraint, $v \in \mathbb{R}^{|X|}$ is a vector with $v_x = x$, and ω is the multiplier on the upper-bound of the limited liability requirement.

Defining $\tilde{q}_R = \lambda q_R$, the dual of this problem is

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}} \max_{s \in \mathbb{R}_+^{|X|}, K \geq 0} q^T(K\iota - \beta s) + \tilde{q}_R^T(s - K\iota) + \theta D_H(\tilde{q}_R || q) + \omega^T(v - s),$$

which can be understood as

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}} \theta D_H(\tilde{q}_R || q) + \omega^T v,$$

subject to

$$\tilde{q}_R - \beta q - \omega \leq 0,$$

$$1 - q_R^T \iota \leq 0.$$

The multipliers of this convex minimization problem are the optimal security design and price. After solving the problem for $\tilde{q}_R$ and ω, we can use the first-order condition to recover the security design:

$$s - K\iota = H_q(q) - H_q(\tilde{q}_R).$$

We use the convention that in the lowest state, the asset value is zero, and therefore $s_0 = 0$, and hence

$$s_x = (e_x - e_0)^T (H_q(q) - H_q(\tilde{q}_R)),$$

where e_x and e_0 are basis vectors associated with the states $x \in X$ and $0 \in X$.

To implement the problem with the additional requirement of monotonicity for the security design, write the monotonicity requirement as $Ms \gg 0$, where M is an $|X| - 1 \times |X|$ matrix. The dual problem is

$$\min_{\tilde{q}_R \in \mathbb{R}_+^{|X|}, \omega \in \mathbb{R}_+^{|X|}, \rho \in \mathbb{R}_+^{|X|}} \theta D_H(\tilde{q}_R || q) + \omega^T v,$$

subject to

$$\tilde{q}_R - \beta q - \omega + M^T \rho \leq 0,$$

$$1 - q_R^T \iota \leq 0.$$

As mentioned above, under our parameters it is not optimal for the seller to avoid information acquisition. We first present the optimal securities that induce information acquisition. We then present the optimal securities that avoid information acquisition below. Note the shapes of these securities are very similar in these two cases, although the level is often quite different.

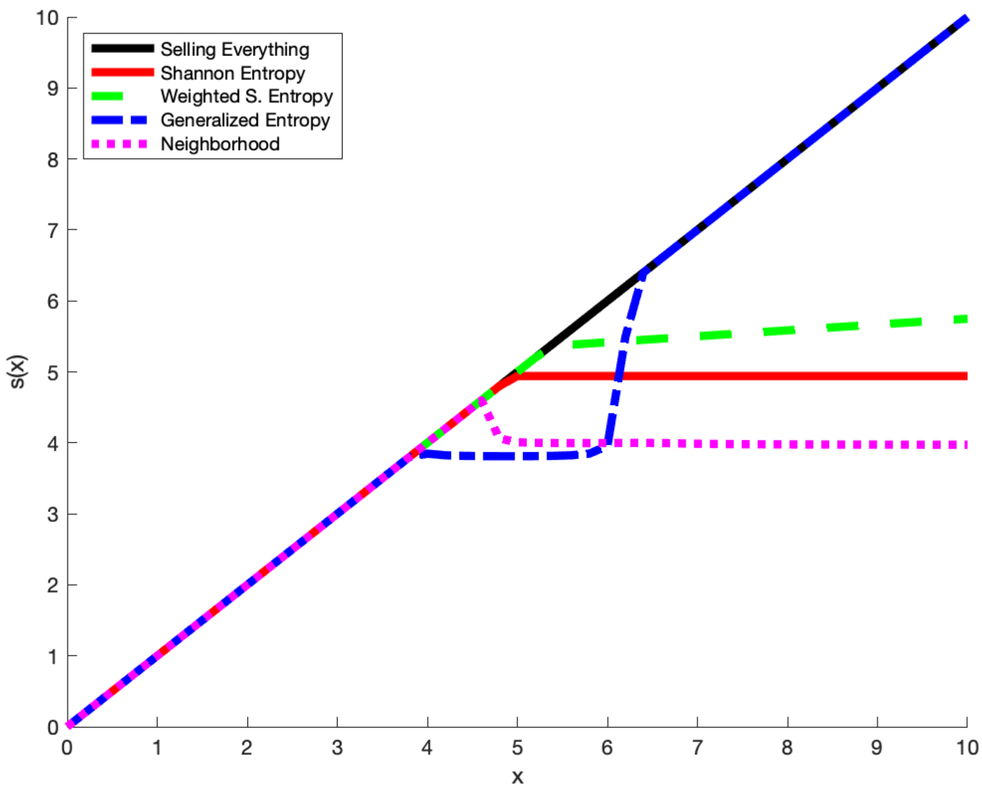

Figure 6: Optimal Security Designs

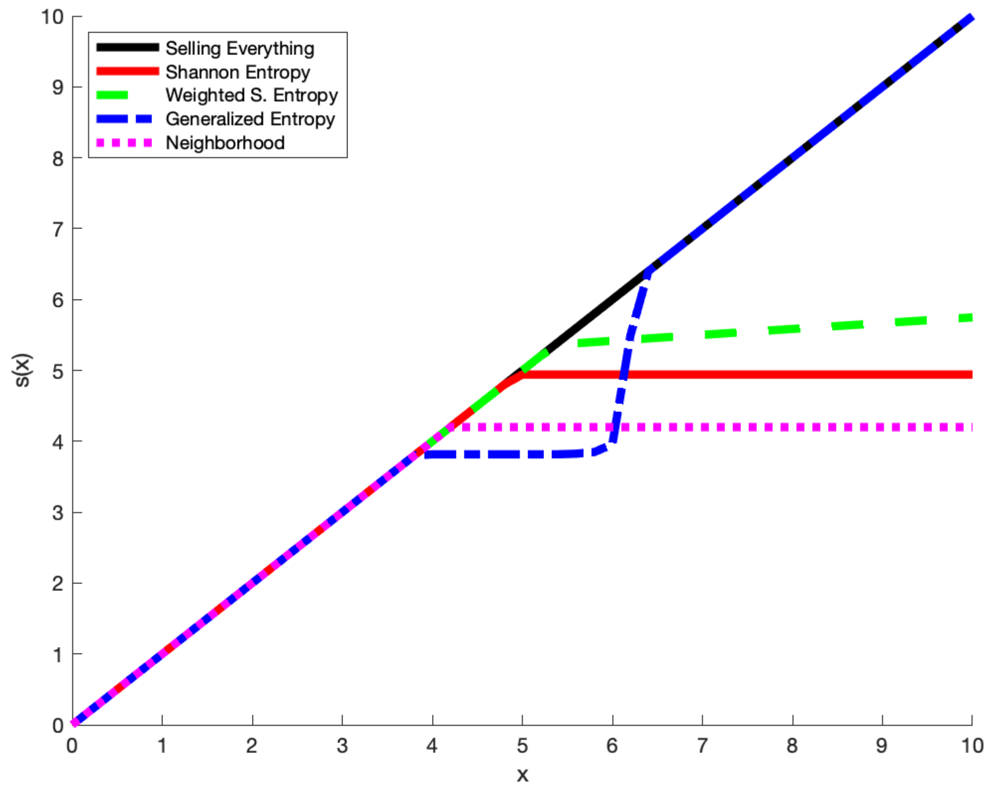

Figure 7: Optimal Monotone Security Designs

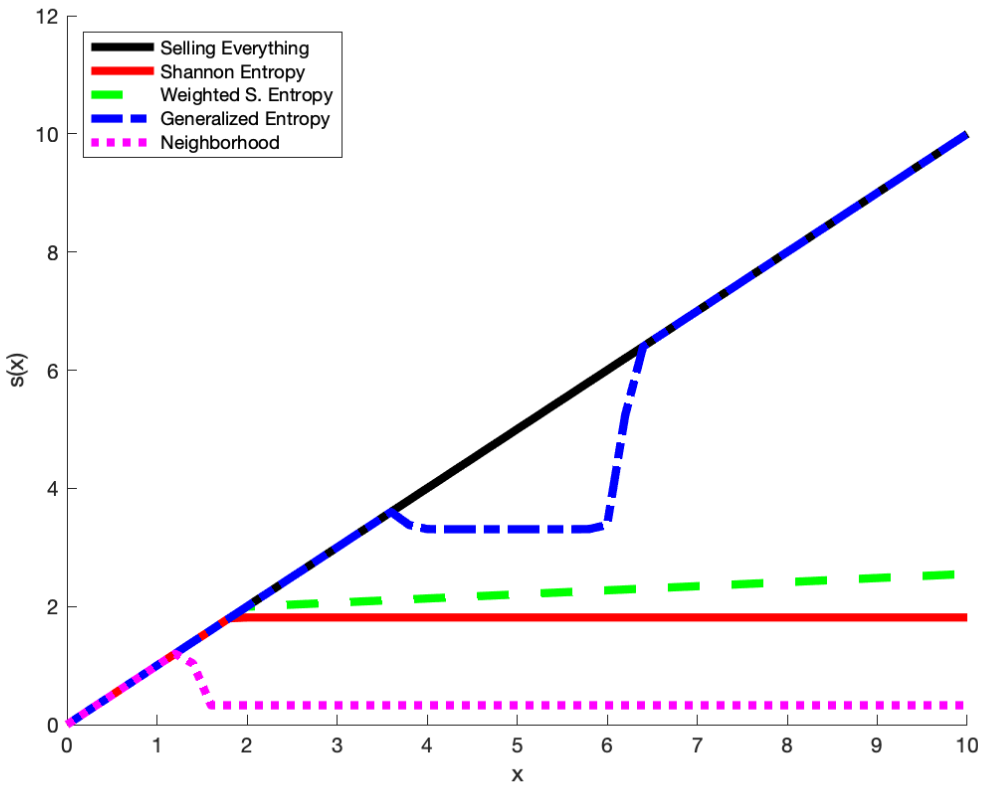

Figure 8: Optimal Security Designs that Avoid Info. Acquisition

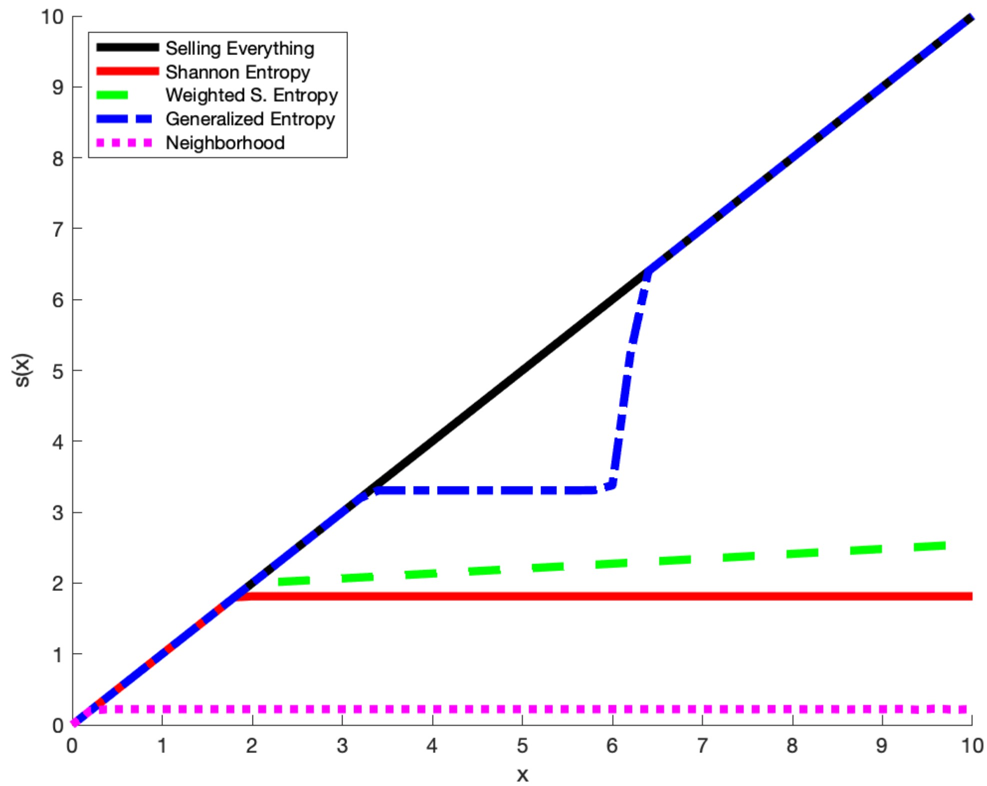

Figure 9: Optimal Monotone Security Designs that Avoid Info Acquisition

C Proofs

C.1 Proof of Proposition 1

We begin with the following lemma, which restates results in Hébert and Woodford (2019). For completeness, we include a proof of this lemma in the technical appendix.

Lemma 4. *Let $C(p,q_0;S,X)$ be any cost function satisfying Assumption 1 (i.e. any* *continuously twice-differentiable UPS cost function). Suppose that, for all $x \in X$,*

$$p_x = r + \varepsilon v_x$$

for some $\varepsilon > 0$, $r \in \mathscr{P}(S)$ with full support on S, and $v_x \in \mathbb{R}^{|S|}$, and that q_0 has full support on X. Then for the matrix-valued function

$$k(q) = Diag(q) \cdot H_{qq}(q;X) \cdot Diag(q),$$

where $Diag(q)$ is the diagonal matrix with q on its diagonal and $H_{qq}(q;X)$ is the Hessian of the H function associated with C,

$$C(p,q_0;S) = \frac{1}{2}\varepsilon^2 \sum_{x \in X, x' \in X} k_{x,x'}(q_0) v_x^T \cdot Diag(r)^{-1} \cdot v_{x'} + o(\varepsilon^2),$$

where $Diag(r)$ is a diagonal matrix with r on the diagonal and ι is a vector of ones.

Proof. See Hébert and Woodford (2019) or the Technical Appendix, section D.2.1. □

Consider now the signal structures p, p', and p'' defined in (6) and (7). Applying this lemma to those particular signal structures, with

$$r' = r + \varepsilon v,$$

we have

$$C(p,q_0;S,X) = \frac{1}{2}\varepsilon^2 k_{x,x}(q_0) v^T \cdot Diag(r)^{-1} \cdot v + o(\varepsilon^2),$$
$$C(p',q_0;S,X) = \frac{1}{2}\varepsilon^2 k_{x',x'}(q_0) v^T \cdot Diag(r)^{-1} \cdot v + o(\varepsilon^2),$$

and

$$
\begin{aligned}
C(p'', q_0; S, X) = &\frac{1}{2}\varepsilon^2 k_{x,x}(q_0)\nu^T \cdot Diag(r)^{-1} \cdot \nu + o(\varepsilon^2) \\
&+ \frac{1}{2}\varepsilon^2 k_{x',x'}(q_0)\nu^T \cdot Diag(r)^{-1} \cdot \nu + o(\varepsilon^2) \\
&+ \varepsilon^2 k_{x,x'}(q_0)\nu^T \cdot Diag(r)^{-1} \cdot \nu + o(\varepsilon^2).
\end{aligned}
$$

Consequently, by Assumption 2, if x and x' do not share a neighborhood in $\mathscr{N}$, $k_{x,x'}(q_0) = 0$, and if they do, $k_{x,x'}(q_0) < 0$. By definition, this property also applies to the Hessian matrix of H. That is, if x and x' share a neighborhood in $\mathscr{N}$, then

$$
\frac{\partial^2}{\partial q_x \partial q_{x'}} H(q; X, \mathscr{N}) < 0,
$$

and otherwise

$$
\frac{\partial^2}{\partial q_x \partial q_{x'}} H(q; X, \mathscr{N}) = 0.
$$

It follows that if x and x' do not share a neighborhood, then

$$
\frac{\partial}{\partial q_x} H(q; X, \mathscr{N}) = \frac{\partial}{\partial q_x} H(q'; X, \mathscr{N}) \tag{20}
$$

for all measures q, q' that differ only in the mass on x'.

Let us now suppose we are given some $\bar{X}$ such that the $\bar{X}_i = \bar{X} \cap N_i$ are disjoint. Observe that it is without loss of generality to suppose H is strictly positive on $q \in \mathscr{P}(X)$ (shifting H by a constant does not change the cost function C). By the strict positivity and homogeneity of degree one of H, at least one partial derivative must be positive, and note that this must continue to hold even if we assume instead that H is only weakly positive. Consequently, by the General Theorem on Functional Dependence (see Leontief (1947) and Gorman (1968)), separability holds:

$$
H(q; \bar{X}, \mathscr{N}) = f(\hat{H}^1(q_1(q), \bar{q}_1(q)), \hat{H}^2(q_2(q), \bar{q}_2(q)), \ldots),
$$

where the $\hat{H}^i$ are continuously differentiable functions only of the values of q_x within the neighborhood $\bar{X}_i$ (and hence of q_i and $\bar{q}_i$), and f is a continuously differentiable function.

By the condition

$$\frac{\partial^2}{\partial q_x \partial q_{x'}} H(q;\bar{X},\mathscr{N}) = 0$$

for x, x' that do not share a neighborhood, the function f must be linear in its arguments. The constant term in f is irrelevant for cost function under Assumption 1, and hence without loss of generality we assume it is zero. We have concluded that $f(x) = \alpha x$ for some constant α, and without loss of generality to rescale the $\hat{H}^i$ functions and assume $\alpha = 1$. Therefore, we can write

$$H(q;\bar{X},\mathscr{N}) = \sum_{i\in\mathscr{I}} \hat{H}^i(q_i(q),\bar{q}_i(q);\bar{X}_i).$$

Under Assumption 1, the level of the cost functions $\hat{H}^i(q_i,\bar{q}_i;\bar{X}_i)$ has no impact on the cost functions. We can therefore assume without loss of generality that $\hat{H}^i(q_i,0;X_i) = 0$, consistent with the assumption of homogeneity of degree one for $H(q;X,\mathscr{N})$. Considering distributions that place all support within a single neighborhood, it follows that the $\hat{H}^i$ are homogenous of degree one in $\bar{q}_i$ and twice-differentiable in q_i. We can therefore write

$$H(q;\bar{X},\mathscr{N}) = \sum_{i\in\mathscr{I}} \bar{q}_i(q)\hat{H}^i(q_i(q),1;\bar{X}_i).$$

Let us now applying Assumption 3 in the $k = 1$ case. In this case,

$$C(p,q;S,\bar{X}) = C(p',q';S,\bar{X}') = C(p,q;S,\bar{X}')$$

for any X', and hence the $\hat{H}^i$ function can depend on $\bar{X}_i$, holding fixed $|\bar{X}_i|$, only in ways that are irrelevant for information costs. It follows that it is without loss of generality to assume $\hat{H}^i$ depends only on the cardinality of $\bar{X}_i$.

As with standard UPS cost functions, any strictly increasing affine transformation of the $\hat{H}^i$ functions generates an equivalent cost function. It is therefore without loss of generality to assume they reach their minima at the uniform distribution, and also without loss of generality to extend $\hat{H}^i$ to the set of measures by assuming homogeneity of degree one.

This completes the proof for the case in which the X_i are disjoint. Let us now suppose we are given some X such that the X_i are not disjoint. Repeatedly applying Assumption 3, for any S, $p \in \mathscr{P}(S)^{|X|}$, and $q \in \mathscr{P}(X)$, there exists a $\bar{X}$, surjection $m : \bar{X} \to X$, $\bar{p} \in \mathscr{P}(S)^{|\bar{X}|}$, and $\bar{q} \in \mathscr{P}(\bar{X})$ such that the $\bar{X}_i$ are disjoint and

$$C(p,q;S,X) = C(\bar{p},\bar{q};S,\bar{X}),$$

where for all $\bar{x} \in \bar{X}$,

$$\begin{aligned}\bar{q}_{\bar{x}} &= q_{m(\bar{x})},\\ \bar{p}_{\bar{x}} &= p_{m(\bar{x}).}\end{aligned}$$

Moreover, $x \in X_i$ if and only if there is exactly one $\bar{x} \in \bar{X}$ such that $\bar{x} \in \bar{X}_i$ and $m(\bar{x}) = x$, and therefore $|X_i| = |\bar{X}_i|$.

Define

$$\hat{q}_s(p,q) = \pi(p,q)q_s(p,q)$$

and observe by Bayes' rule that

$$\hat{q}_{s,x}(p,q) = p_{s,x}q_x.$$

By Assumption 1 and homogeneity of degree one, and using the fact that the $\bar{X}_i$ are disjoint,

$$\begin{aligned}C(\bar{p},\bar{q};S,\bar{X}) &= -H(\bar{q};\bar{X},\mathscr{N}) + \sum_{s\in S} H(\hat{q}_s(\bar{p},\bar{q});\bar{X},\mathscr{N})\\ &= \sum_{i\in\mathscr{I}} \{-\bar{q}_i(\bar{q})\hat{H}^i(q_i(\bar{q}),1;|\bar{X}_i|) + \sum_{s\in S} \bar{q}_i(\hat{q}_s(\bar{p},\bar{q}))\hat{H}^i(q_i(\hat{q}_s(\bar{p},\bar{q})),1;|\bar{X}_i|)\}.\end{aligned}$$

By definition,

$$\bar{q}_i(\bar{q}) = \sum_{\bar{x}\in\bar{X}_i} \bar{q}_{\bar{x}} = \sum_{\bar{x}\in\bar{X}_i} q_{m(\bar{x})} = \sum_{x\in X_i} q_x = \bar{q}_i(q),$$

and (assuming $\bar{q}_i(q) > 0$)

$$q_{i,\bar{x}}(\bar{q}) = \frac{\bar{q}_{\bar{x}}}{\bar{q}_i(\bar{q})} = \frac{q_{m(\bar{x})}}{\bar{q}_i(q)} = q_{i,m(\bar{x})}(q).$$

Hence it follows that

$$\bar{q}_i(\bar{q})\hat{H}^i(q_i(\bar{q}),1;|\bar{X}_i|) = \bar{q}_i(q)\hat{H}^i(q_i(q),1;|X_i|).$$

By a similar argument,

$$\bar{q}_i(\hat{q}_s(\bar{p},\bar{q})) = \sum_{\bar{x}\in\bar{X}_i} \bar{p}_{s,x}\bar{q}_x = \sum_{\bar{x}\in\bar{X}_i} p_{s,m(\bar{x})}q_{m(\bar{x})} = \sum_{x\in X_i} p_{s,m(\bar{x})}q_{m(\bar{x})} = \bar{q}_i(\hat{q}_s(p,q))$$

and

$$q_{i,\bar{x}}(\hat{q}_s(\bar{p},\bar{q})) = \frac{\bar{p}_{s,\bar{x}}\bar{q}_{\bar{x}}}{\bar{q}_i(\bar{q})} = \frac{p_{s,m(\bar{x})}q_{m(\bar{x})}}{\bar{q}_i(q)} = q_{i,m(\bar{x})}(\hat{q}_s(p,q)),$$

and therefore

$$C(p,q;S,X) = \sum_{i\in\mathscr{I}}\{-\bar{q}_i(q)\hat{H}^i(q_i(q),1;|X_i|) + \sum_{s\in S}\bar{q}_i(\hat{q}_s(p,q))\hat{H}^i(q_i(\hat{q}_s(p,q)),1;|X_i|)\}.$$

Consequently, it is without loss of generality to suppose that

$$H(q;X,\mathscr{N}) = \sum_{i\in\mathscr{I}}\bar{q}_i(q)\hat{H}^i(q_i(q),1;|X_i|),$$

concluding the proof.

C.2 Proof of Proposition 2

As argued in the proof of section C.1, it is without loss of generality to suppose that the neighborhoods are disjoint. It follows immediately by Assumption 4 that the Hessian matrix of H^i is invariant to all embeddings in the sense of Chentsov (1982) (see also Amari and Nagaoka (2007) or Hébert and Woodford (2019) for a discussion of this invariance). Consequently, by Theorem 11.1 in Chentsov (1982), the Hessian matrix is proportional to the Fisher matrix. Let c_i denote the constant of proportionality, and note by the convexity of H^i that it is weakly positive. Integrating the Hessian of H^i, it follows that H^i must be proportional to the negative of Shannon's entropy.

C.3 Proof of Lemma 1

First, note that if $\rho \geq 2$ and q_s does not have full support, then p_x will not have full support for the state x such that $e_x^T q_s = 0$, and we will have $D_\rho(p_x || pE_i^T q_i) = \infty$ for any i with $x \in X_i$, as required. For $\rho < 2$, continuity holds, and therefore both boundary cases are satisfied, provided the result holds for interior q_s.

To prove this claim, it is sufficient to show that, if all q_s are interior,

$$\sum_{i \in \mathscr{I}} c_i |X_i|^{1-\rho} \bar{q}_i(q)^{\rho-1} \sum_{x \in X_i} (q_x)^{2-\rho} D_\rho(p_x || \pi_i) = -H_N(q) + \sum_{s \in S} \pi_s(p,q) H_N(q_s(p,q)).$$

By definition,

$$\sum_{s \in S} \pi_s H_N(q_s) = \sum_{s \in S: \pi_s > 0} \pi_s \sum_{i \in \mathscr{I}} c_i \bar{q}_i(q_s) \frac{1}{|X_i|} \frac{1}{(\rho-2)(\rho-1)} \sum_{x \in X_i} \{ (\frac{q_{s,x}}{\frac{1}{|X_i|} \bar{q}_i(q_s)})^{2-\rho} - 1 \}.$$

Using Bayes' rule, $\pi_s \bar{q}_i(q_s) = \sum_{x \in X_i} p_{s,x} q_x = \bar{q}_i(q) \pi_{i,s}$, and therefore

$$\begin{aligned} \sum_{s \in S} \pi_s H_N(q_s) = & \sum_{i \in \mathscr{I}} c_i |X_i|^{1-\rho} \bar{q}_i(q)^{\rho-1} \frac{1}{(\rho-2)(\rho-1)} \sum_{x \in X_i} (q_x)^{2-\rho} \sum_{s \in S: \pi_{i,s} > 0} \pi_{i,s} (\frac{p_{s,x}}{\pi_{i,s}})^{2-\rho} \\ & - \sum_{i \in \mathscr{I}} c_i \bar{q}_i(q) \frac{1}{(\rho-2)(\rho-1)}. \end{aligned}$$

Therefore,

$$-H_N(q) + \sum_{s \in S} \pi_s H_N(q_s) = \sum_{i \in \mathscr{I}} c_i |X_i|^{1-\rho} \bar{q}_i(q)^{\rho-1} \sum_{x \in X_i} (q_s)^{2-\rho} D_\rho(p_x || \pi_i),$$

as required. The proof is essentially identical in the $\rho = 1$ and $\rho = 2$ cases.

C.4 Proof of Proposition 3

It is convenient to work with the transformed variable

$$y = G(x) = \int_{x_L}^{x} \frac{dx}{q(x)},$$

which is well-defined by the compactness of X and the full-support property of $q(x)$.

Using this change-of-variable,

$$V_N(q) = \max\{\sup_{p_R \in C^1([0,\bar{y}],(0,1))} \int_0^{\bar{y}} g(y)p_R(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p_R'(y))^2}{p_R(y)(1-p_R(y))}dy,$$
$$\int_0^{\bar{y}} g(y)u_R(y)dy, 0\},$$

where $\bar{y} = G(x_H)$ and $g(y) = q(G^{-1}(y))^2$.

A necessary condition for always-L to be the optimal strategy is that

$$\int_0^{\bar{y}} g(y)\hat{p}_R(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(\hat{p}_R'(y))^2}{\hat{p}_R(y)(1-\hat{p}_R(y))}dy \leq 0$$

where

$$\hat{p}_R(y) = \varepsilon p_R(y)$$

for some $\varepsilon > 0$ and all $p_R(y) \in C^1([0,\bar{y}],(0,1))$. Considering the limit as $\varepsilon \to 0^+$, we must have

$$\int_0^{\bar{y}} g(y)p_R(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p_R'(y))^2}{p_R(y)}dy \leq 0.$$

Using a change of variable back to the x variable, this is

$$\int_X q(x)p_R(x)u_R(x)dx - \frac{\theta}{4}\int_X q(x)\frac{(p_R'(x))^2}{p_R(x)}dx \leq 0. \tag{21}$$

Now suppose this condition holds for all $p_R(x) \in C^1(X,(0,1))$. It must hold for p_R constant, and hence $\int_X q(x)u_R(x)dx \leq 0$, meaning that always-$L$ is preferred to always-R. Define the functional $J : C^1([0,\bar{y}],(0,1)) \to \mathbb{R}$ by

$$J[p] = \int_0^{\bar{y}} g(y)p(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p'(y))^2}{p(y)(1-p(y))}dy \tag{22}$$

The following lemma demonstrates that this functional is concave on its domain.

Lemma 5. *The functional $J : C^1([0,\bar{y}],(0,1)) \to \mathbb{R}$ defined in (22) is concave on $C^1([0,\bar{y}],(0,1))$.*

Proof. See the Technical Appendix, Section D.3.2. □

Consequently, for all $p \in C^1([0,\bar{y}],(0,1))$, and $\varepsilon > 0$,

$$J[p] \leq J[\varepsilon p] + \delta J[\varepsilon p, p - \varepsilon p],$$

where $\delta J[\varepsilon p, p-\varepsilon p]$ is the first variation from εp in the direction $p - \varepsilon p$,

$$\begin{aligned}\delta J[\varepsilon p, p-\varepsilon p] = &\int_0^{\bar{y}} g(y)(1-\varepsilon)p(y)u_R(y)dy \\ &+ \frac{\theta}{4}(1-\varepsilon)\int_0^{\bar{y}} \frac{(\varepsilon p'(y))^2}{(\varepsilon p(y)(1-\varepsilon p(y)))^2}(1-2\varepsilon p(y))p(y)dy \\ &- \frac{\theta}{2}(1-\varepsilon)\int_0^{\bar{y}} \frac{(\varepsilon p'(y))}{(\varepsilon p(y)(1-\varepsilon p(y)))}p'(y)dy.\end{aligned}$$

In the limit as $\varepsilon \to 0^+$,

$$J[p] \leq \int_0^{\bar{y}} g(y)p(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p'(y))^2}{p(y)}dy$$

and consequently $J[p] \leq 0$. It follows that if (21) holds for all $p_R(x) \in C^1(X,(0,1)$, the optimal policy is always-L. Consequently, the condition is both necessary and sufficient. Moreover, observe that this condition will hold if and only if it holds for all $p_R \in C^1(X,(0,\infty))$ such that $\int_X q(x)p(x)dx = 1$, by the homogeneity of degree of the functional J.

By symmetry, the analogous condition for always-R is, for all $p_L \in \{p \in C^1(X,(0,\infty)) : \int_X q(x)p(x)dx = 1\}$,

$$-\int_0^{\bar{x}} q(x)p_L(x)u_R(x)dx - \frac{\theta}{4}\int_0^{\bar{y}} q(x)\frac{(p_L'(x))^2}{p_L(x)}dx \leq 0$$

If neither of these conditions hold, then it must be the case that

$$\sup_{p_R \in C^1([0,\bar{y}],(0,1))} J[p_R] > \max\{\int_0^{\bar{y}} g(y)u_R(y)dy, 0\}.$$

The space $C^1([0,\bar{y}],(0,1))$ is not compact, so the existence of a maximizer does not follow immediately from concavity. However, the following lemma demonstrates that a maximizer does in fact exist.

Lemma 6. *If*

$$\sup_{p_R \in C^1([0,\bar{y}],(0,1))} J[p_R] > \max\{\int_0^{\bar{y}} g(y)u_R(y)dy, 0\},$$

then there exists an extremal $p_R^* \in C^1([0,\bar{y}],(0,1))$ that is a maximizer and is continuously twice-differentiable except at the discontinuities of $u_R(y)$.

Proof. See the Technical Appendix, Section D.3.1. □

Anywhere it is twice-differentiable, p_R^* must satisfy the Euler-Lagrange equation,

$$q(x)u_R(x) + \frac{\theta}{4}q(x)\frac{(p_R^{*\prime}(x))^2}{(p_R^*(x)(1-p_R^*(x)))^2}(1-2p_R^*(x)) = -\frac{\theta}{2}\frac{d}{dx}[q(x)\frac{p_R^{*\prime}(x)}{p_R^*(x)(1-p_R^*(x))}],$$

along with the natural boundary conditions

$$q(0)\frac{p_R^{*\prime}(0)}{p_R^*(0)(1-p_R(0))} = q(\bar{x})\frac{p_R^{*\prime}(\bar{x})}{p_R^*(\bar{x})(1-p_R^*(\bar{x}))} = 0.$$

The Euler-Lagrange equation can be rewritten as

$$q(x)u_R(x) - \frac{\theta}{4}q(x)\frac{(p_R^{*\prime}(x))^2}{(p_R^*(x)(1-p_R^*(x)))^2}(1-2p_R^*(x)) + \frac{\theta}{2}q'(x)\frac{p_R^{*\prime}(x)}{p_R^*(x)(1-p_R^*(x))} = -\frac{\theta}{2}q(x)\frac{p_R^{*\prime\prime}(x)}{p_R^*(x)(1-p_R^*(x))}$$

and further simplified to

$$\frac{p_R^*(x)(1-p_R^*(x))}{2\theta}u_R(x) - \frac{1}{2}\frac{(p_R^{*\prime}(x))^2}{p_R^*(x)(1-p_R^*(x))}(1-2p_R^*(x)) + \frac{q'(x)}{q(x)}p_R^{*\prime}(x) = -p_R^{*\prime\prime}(x).$$

By the concavity of J, any extremal satisfying these conditions is a maximizer.

C.5 Proof of Lemma 2

We first prove the "if": suppose a function

$$q(x)\psi(x) = \int_{x_L}^{x} q(x')[\frac{2}{\theta}u_R(x') - \frac{1}{2}\psi(x')^2]dx'$$

satisfying $\psi(x_H) = 0$ exists. Observe that this function is continuous. Defining the functional

$$J[p] = \int_{x_L}^{x_H} F(x, p(x), p'(x))dx, \tag{23}$$

$$F(x,p,v) = q(x)u_R(x)p + \frac{\theta}{4}q(x)\frac{v^2}{p},$$

we will prove that the existence of ψ implies that

$$\inf_{p_L \in \{p \in C^1(X,(0,\infty)) : \int_X q(x)p(x)dx = 1\}} J[p_L] = 0.$$

The integrated Euler-Lagrange equation associated with this functional is, for some constant c,

$$\frac{\theta}{2}q(x)\frac{p'(x)}{p(x)} = c + \int_{x_L}^{x} q(x')[u_R(x') - \frac{\theta}{4}(\frac{p'(x')}{p(x')})^2]dx'$$

and the natural boundary conditions are $\frac{\theta}{2}q(x_L)\frac{p'(x_L)}{p(x_L)} = \frac{\theta}{2}q(x_H)\frac{p'(x_H)}{p(x_H)} = 0$. Defining

$$\psi(x) = \frac{p'(x)}{p(x)}$$

demonstrates that if the function ψ exists, an extremal of the functional $J[p]$ on $p \in C^1(X,(0,\infty))$ exists,

$$p^*(x) = A\exp(\int_{x_L}^{x} \psi(x')dx')$$

for any constant $A > 0$.

We next invoke the following lemma to show that the functional $J[p]$ is convex on $C^1(X,(0,\infty))$.

Lemma 7. *The functional* $J : C^1(X,(0,\infty)) \to \mathbb{R}$ *defined in (23) is convex on* $C^1(X,(0,\infty))$.

Proof. See the Technical Appendix, Section D.3.2. □

Consequently, the $p^*(x)$ are minimizers, and must achieve the same value of the functional for all values of A, which by the homogeneity of degree one of $J[p]$ must be zero. Hence, for the particular value of A satisfying

$$A^{-1} = \int_{x_L}^{x_H} q(x)\exp(\int_{x_L}^{x} \psi(x')dx')dx,$$

the associated p^* must minimize $J[\cdot]$ on $\{p \in C^1(X,(0,\infty)) : \int_X q(x)p(x)dx = 1\}$ and have $J[p^*] = 0$.

We next prove the "only if", a proof that largely follows the arguments of the proof of Proposition 3. We will show that if

$$\inf_{p_L \in \{p \in C^1(X,(0,\infty)) : \int_X q(x)p(x)dx = 1\}} J[p_L] = 0, \tag{24}$$

then the function ψ must exist.

It is convenient to work with the transformed variable

$$y = G(x) = \int_{x_L}^{x} \frac{dx}{q(x)},$$

which is well-defined by the compactness of X and the full-support property of $q(x)$. Define $y_L = G(x_L)$ and $y_H = G(x_H)$ as the boundary points, and define $g(y) = q(G^{-1}(y))^2$.

Employing the change of variable

$$\phi(y) = \sqrt{p(y)},$$

we use the domain $[y_L - \varepsilon, y_H + \varepsilon]$ for some $\varepsilon > 0$, and define

$$\hat{J}[\phi] = \int_{y_L-\varepsilon}^{y_H+\varepsilon} F(y, \phi(y), \phi'(y))dy, \tag{25}$$

$$F(x, \phi, v) = \begin{cases} g(y)u_R(y)\phi^2 + \theta v^2 & y \in [y_L, y_H] \\ \theta v^2 & y \notin [y_L, y_H]. \end{cases}$$

If (24) holds, there must exist a sequence $\{\phi_n \in \{\phi \in C^1([y_L - \varepsilon, y_H + \varepsilon], (0, \infty)) : \int_{y_L}^{y_H} g(y)\phi(y)^2 dy = 1\}\}_{n=1}^{\infty}$ satisfying

$$\lim_{n\to\infty} \hat{J}[\phi_n] = 0.$$

The functions ϕ_n are elements of the Sobolev space $W^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$. By definition, for any $\delta > 0$, there exists an n_0 such that for all $n > n_0$, $|\hat{J}[\phi_n]| < \delta$. Consequently,

$$\theta \int_{y_L-\varepsilon}^{y_H+\varepsilon} \phi_n'(y)^2 dy < \delta,$$

and

$$B \int_{y_L-\varepsilon}^{y_H+\varepsilon} \phi_n(y) dy < \delta,$$

where $B = \max_{y\in[y_L, y_H]} |g(y)u_R(y)|$. The sequence $\{\phi_n\}_{n=n_0}^{\infty}$ is therefore bounded in the $W^{1,2}$ norm, and hence converges weakly to some $\phi^* \in W^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$, immediately implying that

$$\hat{J}[\phi^*] = 0$$

and

$$\int_{y_L}^{y_H} g(y)\phi^*(y)^2 dy = 1.$$

By the homogeneity of degree two of $\hat{J}$ and the observation that $F(y, \phi, v) = F(y, -\phi, v)$, ϕ^* must be a minimizer of $\hat{J}$ on $W^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$, and it is without loss of generality to assume $\phi^*(y) \geq 0$ for all $y \in [y_L - \varepsilon, y_H + \varepsilon]$.

We invoke the following regularity result, proven in the technical appendix, to show that ϕ^* is continuously differentiable, and continuous twice differentiable on any interval on which u_R is continuous.

Lemma 8. *If $\phi^* \in W^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$ is a minimizer of the functional $\hat{J}$ defined in (25), then $\phi^* \in C^1([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$, and ϕ^* is continuously twice-differentiable on any interval on which u_R is continuous.*

Proof. See the Technical Appendix, Section D.3.3. □

Let $y_1, \ldots y_{k-1}$ be the (possibly empty) set of points of discontinuity for u_R, and let $y_0 = y_L$ and $y_k = y_H$. This regularity result implies that the Euler-Lagrange equation,

$$2\theta\phi^{*\prime\prime}(y) = 2g(y)\phi^*(y)u_R(y)$$

must hold on all $y \in (y_{i-1}, y_i)$.

Suppose that for some $y' \in [y_L, y_H]$, $\phi^*(y') = 0$. By the fact that $\phi^*(y)$ is continuously differentiable and it is without loss of generality to assume $\phi^*(y) \geq 0$, it must be the case that $\phi^{*\prime}(y') = 0$. In this case, $\phi^*(y)$ constant on $y \in [y_L, y_H]$ satisfies the Euler-Lagrange equations. The system

$$\frac{d}{dy}\begin{bmatrix} \phi^{*\prime}(y) \\ \phi^*(y) \end{bmatrix} = \begin{bmatrix} \theta^{-1}g(y)u_R(y)\phi^*(y) \\ \phi^{*\prime}(y) \end{bmatrix}$$

is uniformly Lipschitz-continuous in $(\phi^*(y), \phi^{*\prime}(y))$ and continuous in y on all intervals (y_{i-1}, y_i), and hence by the Picard-Lindelof theorem, a unique solution to any initial value problem on any interval $[y_{i-1}, y_i]$ exists. Consequently, if $\phi^*(y') = 0$ for any $y' \in [y_L, y_H]$, $\phi^*(y) = 0$ for all $y \in [y_L, y_H]$.

But by the result that

$$\int_{y_L}^{y_H} q(y)\phi^*(y)^2 dy = 1,$$

$\phi^*(y)$ cannot be zero everywhere. It follows that $\phi^*(y) > 0$.

Defining $\hat{\psi} : [y_L, y_H] \to \mathbb{R}$ by

$$\hat{\psi}(y) = \frac{2\phi^{*\prime}(y)}{\phi^*(y)},$$

the Euler-Lagrange equation implies that everywhere u_R is continuous,

$$\hat{\psi}'(y) = 2\theta^{-1} g(y) u_R(y) - \frac{1}{2} g(y) \hat{\psi}(y)^2.$$

In the x variable, this is, for $\psi(x) = \hat{\psi}(G(x))$,

$$\frac{d}{dx}[\psi'(x)q(x)] = 2\theta^{-1} q(x) u_R(x) - \frac{1}{2} q(x) \psi(x)^2$$

Integrating and applying $\phi^{*\prime}(y_L) = \phi^{*\prime}(y_H) = 0$ proves the result.

C.6 Proof of Corollary 2

We first prove that $p_R^*(x)$ is strictly increasing on (x_L, x_H).

It is convenient to work with the transformed variable

$$y = G(x) = \int_{x_L}^{x} \frac{dx}{q(x)},$$

which is well-defined by the compactness of X and the full-support property of $q(x)$. We have

$$p_R^{*\prime}(y) = p_R^{*\prime}(x(y)) \frac{dx}{dy} = q(x(y)) p_R^{*\prime}(x(y)).$$

The Euler-Lagrange equation rewritten with this change of variable is

$$g(y) u_R(y) + \frac{\theta}{4} \frac{(p_R^{*\prime}(y))^2}{(p_R^*(y)(1 - p_R^*(y)))^2} (1 - 2p_R^*(y)) = -\frac{\theta}{2} \frac{d}{dy} [\frac{p_R^{*\prime}(y)}{p_R^*(y)(1 - p_R^*(y))}].$$

It also also convenient to work with the transformed function

$$\phi(y) = \cos^{-1}(\sqrt{p_R^*(y)}),$$

which satisfies

$$\phi'(y) = -\frac{1}{2}\frac{p_R^{*\prime}(y)}{\sqrt{p_R^*(y)(1-p_R^*(y))}}.$$

We assume (without loss of generality) that $\phi(y) \in [0, \frac{\pi}{2}]$. The corresponding Euler-Lagrange equation is

$$g(y)u_R(y) + \frac{\theta}{4}\frac{(\phi'(y))^2}{p_R^*(y)(1-p_R^*(y))}(1-2p_R^*(y)) = \frac{\theta}{2}\frac{d}{dy}[\frac{\phi'(y)}{\sqrt{p_R^*(y)(1-p_R^*(y))}}]$$

which further simplifies to

$$\sin(2\phi(y))g(y)u_R(y) = \frac{\theta}{2}\phi''(y).$$

By Proposition 3, this equation is satisfies everywhere $u_R(y)$ is continuous, and ϕ is continuously differentiable everywhere. The boundary conditions are $\phi'(0) = \phi'(\bar{y}) = 0$, where $\bar{y} = G(x_H)$.

Define $y^* = G(x^*)$. By the single-crossing property, $u_R(y) < 0$ for all $y < y^*$, and hence ϕ is strictly concave on $y \in (0, y^*)$. It follows by $\phi'(0) = 0$ that $\phi'(y) < 0$ for all $y \in (0, y^*)$. Similarly, $u_R(y) > 0$ for all $y > y^*$. It follows that ϕ strictly convex on $(y^*, \bar{y})$, and by $\phi'(\bar{y}) = 0$ we must have $\phi'(y) < 0$ for all $y \in (y^*, \bar{y})$. By the continuity of $\phi'(y)$, $\phi'(y) < 0$ for all $y \in (0, \bar{y})$. It follows immediately that $p_R^{*\prime}(x) > 0$ for all $x \in (x_L, x_H)$.

We next prove that for some $x_1 \in (x_L, x_H)$, $p_R^*(x)$ is strictly convex on $x \in [x_L, x_1)$. Define x_1' as the lesser of x^* and the smallest point of discontinuity for $u_R(x)$ on (x_L, x_H). On the interval (x_L, x_1'), $u_R(x)$ is continuous and strictly negative (by single crossing). The Euler-Lagrange equation from Proposition 3 can be written as

$$\frac{1}{2\theta}u_R(x) - \frac{1}{2}\frac{(p_R^{*\prime}(x))^2}{(p_R^*(x)(1-p_R^*(x)))^2}(1-2p_R^*(x)) + \frac{q'(x)}{q(x)}\frac{p_R^{*\prime}(x)}{p_R^*(x)(1-p_R^*(x))} = -\frac{p_R^{*\prime\prime}(x)}{p_R^*(x)(1-p_R^*(x))}.$$

By the continuity of $p_R^{*\prime}(x)$ and $u_R(x)$, the fact that $p_R^*(x)$ has a strictly positive minimum on X, the boundary condition $p_R^{*\prime}(x_L) = 0$, and $q \in \mathscr{P}_{LipG}(x)$ (implying $\frac{q'(x)}{q(x)}$ bounded), we must have

$$\lim_{x \to x_L} p_R^{*\prime\prime}(x) > 0,$$

and by the continuous twice-differentiability of p_R^* on (x_L, x_1'), this must hold on some interval (x_L, x_1) with $x_1 \in (x_L, x_1']$, implying that $p_R^*(x)$ is convex on $[x_L, x_1)$. The argument that for some $x_2 \in (x_L, x_H)$, $p_R^*(x)$ is strictly concave on $x \in (x_2, x_H]$ is symmetric, and $x_2 \geq x^* \geq x_1$ proves the result.

C.7 Proof of Proposition 4

Here we solve the multi-variate problem in the calculus of variations stated in Section 3.4,

$$\inf_{\{p_a(x)\}_{a \in \mathbb{A}} \in \mathscr{P}_{LipG}(A)} \int_X q(x) \int_A [p_a(x)(a - \gamma^T x)^2 + \frac{\theta}{4} \frac{|\nabla_x p_a(x)|^2}{p_a(x)}] da dx$$

where under the prior $q(x)$ $x \sim N(\mu_0, \Sigma_0)$, $X = \mathbb{R}^L$, and $A = \mathbb{R}$.

We can write this as

$$\int_X q(x) \int_A F(a, p_a(x), \nabla_x p_a(x); x)\, da\, dx,$$

where for each pair (x, a), the function

$$F(a, f, g; x) \;\equiv\; f \cdot (a - \gamma^T x)^2 \,+\, \frac{\theta}{4} \frac{|g|^2}{f}$$

is a convex function of the arguments (f, g) everywhere on its domain (the half-plane on which $f > 0$). To prove convexity, observe that

$$\begin{bmatrix} F_{gg} & F_{fg} \\ F_{gf} & F_{ff} \end{bmatrix} = \frac{\theta}{4} \begin{bmatrix} \frac{1}{f} I & -\frac{g}{f^2} \\ -\frac{g^T}{f^2} & 2\frac{g^T g}{f^3} \end{bmatrix}.$$

The upper left block is positive definite, and the determinant of the matrix is strictly positive, and consequently the matrix is strictly positive-definite.

Given the convexity of the objective, the first-order conditions are both necessary and sufficient for an optimum. The relevant first-order conditions are furthermore the same as those for minimization of the Lagrangian

$$\int_X q(x) \int_A \mathscr{L}(a, p_a(x), \nabla p_a(x); x)\, da\, dx,$$

where

$$\mathscr{L}(a, f, g; x) \;=\; F(a, f, g; x) + \varphi(x) f + \psi_a(x) f. \tag{26}$$

Here $\varphi(x)$ is the Lagrange multiplier associated with the constraint

$$\int_A p_a(x) da \;=\; 1 \tag{27}$$

for each $x \in X$, as is required in order for $p_a(x)$ to be a probability density function, and $\psi_a(x)$ is the multiplier on the constraint that $p_a(x)$ be weakly positive.

For given Lagrange multipliers, the problem of minimizing the Lagrangian can further be expressed as a separate minimization problem for each possible action a. Then if we can find a function $\varphi(x)$ and a function $p_a(x)$ for each $a \in A$, with $p_a(x) > 0$ for all x, such that (i) for each $a \in A$, the function $p_a(x)$ minimizes

$$\int_X q(x) \mathscr{L}(a, p_a(x), \nabla_x p_a(x); x)\, dx, \tag{28}$$

and (ii) condition (27) holds for all $x \in X$, then we will have derived an optimal information structure.

For the problem of choosing a function $p_a(x)$ to minimize (28), the first-order conditions are given by the Euler-Lagrange equations

$$q(x) \frac{\partial \mathscr{L}}{\partial f}(a, p_a(x), \nabla_x p_a(x); x) \;= \sum_{k=1}^{L} \frac{d}{dx^k} \left[q(x) \frac{\partial L}{\partial g^k}(a, p_a(x), \nabla_x p_a(x); x) \right],$$

or equivalently,

$$\frac{\partial \mathscr{L}}{\partial f}(a, p_a(x), \nabla_x p_a(x); x) \;=\; \nabla_g \mathscr{L}(a, p_a(x), \nabla_x p_a(x); x) \cdot \nabla_x[\log q(x)] \\ + \nabla_x \cdot \left[\nabla_g \mathscr{L}(a, p_a(x), p_a'(x); x)\right].$$

In the case of the objective function (26), we have

$$\frac{\partial \mathscr{L}}{\partial f} \;=\; (a-\gamma^T x)^2 - \frac{\theta}{4}|\nabla_x v_a(x)|^2 + \varphi(x) + \psi_a(x),$$

$$\nabla_g \mathscr{L} \;=\; \frac{\theta}{2}\nabla_x v_a(x),$$

where $v_a(x) \equiv \log p_a(x)$. Under our assumption of a Gaussian prior, we also have

$$\nabla_x[\log q(x)] \;=\; \Sigma_0^{-1}(\mu_0 - x).$$

Substituting these expressions, the Euler-Lagrange equations take the form

$$(a-\gamma^T x)^2 + \varphi(x) + \psi_a(x) - \frac{\theta}{4}|\nabla_x v_a(x)|^2 = \frac{\theta}{2}(\mu_0 - x)^T \Sigma_0^{-1} \nabla_x v_a(x) + \frac{\theta}{2}\nabla_x \cdot \nabla_x v_a(x)$$

for all x and a.

In the case that $4|\Sigma_0 \gamma|^2 > \theta$, we conjecture and verify that these equations have a solution given by

$$\psi_a(x) = 0,$$
$$\nabla_x v_a(x) \;= \lambda\,[a - \gamma^T \mu_0 - \sigma^{-2}\lambda^T(x-\mu_0)], \tag{29}$$

for some values of $\sigma \in \mathbb{R}, \lambda \in \mathbb{R}^L$ and some $\phi(x)$. Note that this conjecture can be integrated, with

$$\exp(v_a(x)) = p_a(x) = -\frac{\sigma}{\sqrt{2\pi}}\exp(-\frac{\sigma^2}{2}(a - \gamma^T \mu - \sigma^{-2}\lambda^T(x-\mu))^2).$$

Plugging in this conjecture,

$$\begin{aligned}\varphi(x) = -(a-\gamma^T x)^2 + \frac{\theta}{4}\lambda^T\lambda(a-\gamma^T x + (\gamma-\sigma^{-2}\lambda)^T(x-\mu_0))^2 \\ + \frac{\theta}{2}(\mu_0 - x)^T\Sigma_0^{-1}\lambda(a-\gamma^T x) \\ + \frac{\theta}{2}(\mu_0 - x)^T\Sigma_0^{-1}\lambda(\gamma-\sigma^{-2}\lambda)^T(x-\mu_0) \\ + \frac{\theta}{2}\sigma^{-2}\lambda^T\lambda.\end{aligned}$$

By variation of parameters in a, we must have (as in the proposition)

$$\lambda^T\lambda = \frac{4}{\theta}$$

and, for all x,

$$(x-\mu_0)^T\Sigma_0^{-1}\lambda = \lambda^T\lambda(x-\mu_0)^T(\gamma-\sigma^{-2}\lambda).$$

Hence we require

$$\frac{\theta}{4}\Sigma_0^{-1}\lambda = \gamma - \sigma^{-2}\lambda,$$

which implies (as stated in the text) that

$$\lambda = (\frac{\theta}{4}\Sigma_0^{-1} + \sigma^{-2}I)^{-1}\gamma, \tag{30}$$

and

$$|(\frac{\theta}{4}\Sigma_0^{-1} + \sigma^{-2}I)^{-1}\gamma|^2 = \frac{4}{\theta}, \tag{31}$$

which is feasible for $\sigma > 0$ under the assumption that $|\Sigma_0\gamma|^2 > \frac{\theta}{4}$. Note that this formula is a rescaled version of the one stated in the proposition.

Observe that we can rewrite this equations as

$$\Sigma_0^{-1}\lambda = \frac{4}{\theta}\gamma - \sigma^{-2}\lambda\lambda^T\lambda,$$

and hence that

$$\lambda = \frac{4}{\theta}(\Sigma_0^{-1} + \sigma^{-2}\lambda\lambda^T)\gamma. \tag{32}$$

Now suppose the DM receives a Gaussian signal $s = \lambda^T x + \varepsilon$, where the "observation error" ε is normally distributed, with mean zero and a variance σ^2, and independent of the value of x. Here, σ and λ are the solutions to (30) and (31) above.

With such a signal, and given the Gaussian prior beliefs, the DM's posterior beliefs are Gaussian. The posterior precision of the DM's belief about $\lambda^T x$ is

$$(\lambda^T \Sigma_0 \lambda)^{-1} + \sigma^{-2},$$

and the posterior mean is

$$((\lambda^T \Sigma_0 \lambda)^{-1} + \sigma^{-2})^{-1}((\lambda^T \Sigma_0 \lambda)^{-1} \lambda^T \mu_0 + \sigma^{-2} s),$$

while the posterior mean and precision about any $z^T x$ with $z^T \Sigma_0 \lambda = 0$ is unchanged. An orthogonal basis of these z vectors and λ form an orthogonal basis, and let

$$\gamma = b_0 \lambda + b_1 z_1 + \dots,$$

observing that

$$b_0 = \frac{\gamma^T \Sigma_0 \lambda}{\lambda^T \Sigma_0 \lambda}.$$

The posterior variance-covariance matrix is

$$\Sigma_s = \Sigma_0 + \frac{\Sigma_0 \lambda \lambda^T \Sigma_0}{(\lambda^T \Sigma_0 \lambda)^2}\left(\frac{1}{(\lambda^T \Sigma_0 \lambda)^{-1} + \sigma^{-2}} - \lambda^T \Sigma_0 \lambda\right),$$

which simplifies to

$$\begin{aligned}\Sigma_s &= \Sigma_0 + \frac{\Sigma_0 \lambda \lambda^T \Sigma_0}{(\lambda^T \Sigma_0 \lambda)}\left(\frac{1}{1 + \sigma^{-2} \lambda^T \Sigma_0 \lambda} - 1\right) \\ &= \Sigma_0 - \Sigma_0 \lambda \lambda^T \Sigma_0 \frac{\sigma^{-2}}{1 + \sigma^{-2} \lambda^T \Sigma_0 \lambda},\end{aligned}$$

and therefore by the Sherman-Morrison lemma,

$$\Sigma_s^{-1} = \Sigma_0^{-1} + \sigma^{-2} \lambda \lambda^T.$$

The posterior mean of $\gamma^T x$ (and hence optimal action $a(s)$) is

$$E[\gamma^T x|s] = \frac{\gamma^T \Sigma_0 \lambda}{\lambda^T \Sigma_0 \lambda}[((\lambda^T \Sigma_0 \lambda)^{-1} + \sigma^{-2})^{-1}((\lambda^T \Sigma_0 \lambda)^{-1} \lambda^T \mu_0 + \sigma^{-2} s) - \lambda^T \mu_0]$$
$$+ \gamma^T \mu_0,$$

which simplifies to (as given in the text)

$$E[\gamma^T x|s] = \gamma^T \mu_0 + \frac{\gamma^T \Sigma_0 \lambda}{\lambda^T \Sigma_0 \lambda} \frac{\sigma^{-2}}{(\lambda^T \Sigma_0 \lambda)^{-1} + \sigma^{-2}} (s - \lambda^T \mu).$$

Observe by the definitions of λ and σ that

$$1 = \lambda^T \Sigma_0 \gamma - \sigma^{-2} \lambda^T \Sigma_0 \lambda$$

and therefore (as stated in the text)

$$E[\gamma^T x|s] = \gamma^T \mu_0 + \sigma^{-2}(s - \lambda^T \mu_0).$$

Consequently, a is normally distributed conditional on x, with conditional mean

$$E[a(s)|x] = \gamma^T \mu_0 + \sigma^{-2} \lambda^T (x - \mu_0)$$

and conditional variance

$$Var[a(s)|x] = \sigma^{-2}.$$

That is,

$$p_a(x) = \frac{\sigma}{\sqrt{2\pi}} \exp(-\frac{\sigma^2}{2}(a - \gamma^T \mu_0 - \sigma^{-2} \lambda^T (x - \mu_0))^2,$$

and

$$\nabla_x \ln(p_a(x)) = \lambda (a - \gamma^T \mu_0 - \sigma^{-2} \lambda^T (x - \mu_0)),$$

which is the conjectured and verified functional form in (29).

Now consider the problem

$$z^* \in \arg\min_{z:|z|^2=1} z^T(\Sigma_0^{-1} + \sigma^{-2}\lambda\lambda^T)^{-1}\gamma.$$

The first-order condition is

$$\Sigma_s\gamma - \psi z^* = 0,$$

where ψ is the multiplier on $z^T z = 1$, and therefore by (32)

$$z^* \propto \lambda,$$

concluding the proof.

C.8 Proof of Corollary 3

In this corollary we rewrite the problem in terms of a choice of a normally distributed signal $s \in \mathbb{R}^L$ with conditional mean μ_x and positive-semidefinite variance matrix Ω. Given such a signal, the posterior is normally distributed with mean μ_s and posterior variance

$$\Sigma_s = (\Sigma_0^{-1} + \Omega^{-1})^{-1}.$$

Observe by Proposition 4 that the optimal signal structure falls into this class.

Now consider the original problem in posterior form (as in the multi-dimensional generalization of equation (11)). Because the posteriors of this problem are normally distributed, we have

$$\int_{\mathbb{R}^k} \frac{|\nabla_x q_s(x)|^2}{q_s(x)} dx = E[|\Sigma_s^{-1}(x-\mu_s)|^2 | s]$$

and therefore

$$\begin{aligned}\int_{R^k} \pi(s) \int_X \frac{|\nabla_x q_s(x)|^2}{q_s(x)} dxds &= E[tr[\Sigma_s^{-1}(x-\mu_s)(x-\mu_s)^T\Sigma_s^{-1}]] \\ &= tr[\Sigma_s^{-1}].\end{aligned}$$

By the same argument, for the prior q,

$$\int_X \frac{|\nabla_x q(x)|^2}{q(x)} dx = tr[\Sigma_0^{-1}].$$

Given such a signal structure, the optimal action is

$$a^*(s) = \gamma^T \mu_s,$$

and therefore

$$\begin{aligned}\int_X q(x) \int_{\mathbb{R}^k} p_s(x)(a^*(s) - \gamma^T x)^2 ds dx &= E[Var[\gamma^T x|s]] \\ &= \gamma^T \Sigma_s \gamma.\end{aligned}$$

Let $\mathscr{M}_k$ be the set of $k \times k$ real symmetric positive-definite matrices. We can write the posterior-based problem as

$$\inf_{\Sigma_s \in \mathscr{M}_k} \gamma^T \Sigma_s \gamma - \frac{\theta}{4} tr[\Sigma_s^{-1}] + \frac{\theta}{4} tr[\Sigma_0^{-1}]$$

subject to the constraint

$$\Sigma_s^{-1} \succeq \Sigma_0^{-1},$$

which equivalent to $\Sigma_s \preceq \Sigma_0$. By Proposition 4, the optimal solution to this problem is

$$\Sigma_s^* = (\Sigma_0^{-1} + \sigma^{-2} \lambda \lambda^T)^{-1}.$$

C.9 Proof of Corollary 4

In the case that $\theta \geq 4|\Sigma_0 \gamma|^2$, instead, there is no solution to the Euler-Lagrange equations from the proof of Proposition 4, and we can show that there is no interior solution to the optimization problem. Instead it is optimal to choose a completely uninformative information structure, and to choose the estimate $a = \mu$ at all times. This is because in this case, one can show that any information structure and estimation rule implies that

$$V \equiv \mathrm{E}[(a-\gamma^T x)^2] + \frac{\theta}{4}\mathrm{E}[I(x)] \geq \mathrm{E}[(\gamma^T(x-\mu))^T] = \gamma^T \Sigma_0 \gamma,$$

where $I(x)$ is the Fisher information, with the lower bound achieved only in the case that $a = \mu$ with probability 1.

Consider some hypothetical policy $p_a(x)$. We begin by observing that the Cramér-Rao bound for a biased estimator[35] implies that

$$\mathrm{E}^p[(a-\gamma^T x)^2|x] \geq (\nabla_x \bar{a}(x))^T \cdot I(x;p)^{-1} \cdot \nabla_x \bar{a}(x) + (\bar{a}(x) - \gamma^T x)^2.$$

where $\bar{a}(x) \equiv \mathrm{E}^p[a|x]$ under the measure $p_a(x)$, and $I(x;p)$ is the Fisher information of x under $p_a(x)$.

Thus,

$$\begin{aligned}
&\mathrm{E}^p[(a-\gamma^T x)^2|x] + \frac{\theta}{4} tr[I(x)] \\
&\geq (\nabla_x \bar{a}(x))^T \cdot I(x;p)^{-1} \cdot \nabla_x \bar{a}(x) + \frac{\theta}{4} tr[I(x;p)] + (\bar{a}(x) - \gamma^T x)^2 \\
&\geq \inf_I \{\nabla_x \bar{a}(x))^T \cdot I^{-1} \cdot \nabla_x \bar{a}(x) + \frac{\theta}{4} tr[I]\} + (\bar{a}(x) - \gamma^T x)^2
\end{aligned}$$

where the minimization is taken over the set of positive-definite matrices.

In the technical appendix, we prove the following lemma:

Lemma 9. *Let Λ_0 be a $k \times k$ real symmetric positive-semidefinite matrix, let $\mathscr{M}_k$ be the set of $k \times k$ real symmetric positive-definite matrices, and let $v \in \mathbb{R}^k$ be a vector. Then*

$$2|v| = \inf_{\Lambda \in \mathscr{M}_k} v^T \Lambda^{-1} v + tr[\Lambda]$$

Proof. See the Technical Appendix, D.2.2. □

[35] See Cover and Thomas (2006), p. 396.

By this lemma,

$$\inf_I \{ \frac{4}{\theta} \nabla_x \bar{a}(x))^T \cdot I^{-1} \cdot \nabla_x \bar{a}(x) + tr[I] \} = 4\theta^{-\frac{1}{2}} |\nabla_x \bar{a}(x)|.$$

Therefore,

$$\begin{aligned} \mathrm{E}^p[(a-\gamma^T x)^2 | x] + \frac{\theta}{4} tr[I(x)] &\geq \theta^{1/2} |\nabla_x \bar{a}(x)| + (\bar{a}(x) - \gamma^T x)^2 \\ &\geq 2|\Sigma_0 \gamma| \, |\nabla_x \bar{a}(x)| + (\bar{a}(x) - \gamma^T x)^2 \\ &\geq 2\gamma^T \Sigma_0 \nabla_x \bar{a}(x) + (\bar{a}(x) - \gamma^T x)^2, \end{aligned}$$

where the next-to-last inequality follows from the assumption that $\theta \geq 4|\Sigma_0 \gamma|^2$ and the last from the Cauchy-Schwarz inequality. Taking the expected value under the prior $q(x)$, it then follows that

$$V \geq \int_X q(x) \, [2\gamma^T \Sigma_0 \nabla_x \bar{a}(x) + (\bar{a}(x) - \gamma^T x)^2] \, dx. \tag{33}$$

We wish to obtain a lower bound for the integral on the right-hand side of (33). To do this, we solve for the function $\bar{a}(x)$ that minimizes this integral, using the calculus of variations. Once again, we note that the integrand is a convex function of $\bar{a}$ and $\nabla_x \bar{a}$, so that the first-order conditions are both necessary and sufficient for a minimum. The first-order conditions are given by the Euler-Lagrange equations

$$\begin{aligned} 2q(x)(\bar{a}(x) - \gamma^T x) &= 2\gamma^T \Sigma_0 \nabla_x q(x) \\ &= 2q(x) \gamma^T (x - \mu_0) \end{aligned}$$

which have a unique solution $\bar{a}(x) = \gamma^T \mu_0$ for all x.

Substituting this solution into the integral (33), we obtain the tighter lower bound

$$V \geq \int_X q(x) (\gamma^T (x - \mu_0))^2 \, dx = \gamma^T \Sigma_0 \gamma. \tag{34}$$

But this lower bound is achievable by choosing $a = \gamma^T \mu_0$ with probability 1, regardless of the value of x (the optimal estimate in the case of a perfectly uninformative information structure). Hence a perfectly uninformative information structure is optimal for all $\theta \geq 4|\Sigma_0\gamma|^2$.

This solution is not only *one* way of achieving the lower bound, it is the only way. It follows from the reasoning used to derive the lower bound for V that the lower bound can be achieved only if each of the weak inequalities holds as an equality. But the bound in (34) is equal to the bound in (33) only if $\bar{a}(x) = \gamma^T\mu_0$ almost surely; thus optimality requires this. And the restriction that $\mathrm{E}[a|x] = \gamma^T\mu_0$ for a set of x with full measure implies that we must have

$$\mathrm{E}[(a-\gamma^T x)^2|x] = (\gamma^T(x-\mu_0))^2 + Var[a|x].$$

This in turn implies that

$$\mathrm{E}[(a-\gamma^T x)^2] \;=\; \mathrm{E}[(\gamma^T(x-\mu_0))^2] + \mathrm{E}[Var[a|x]] \;=\gamma^T\Sigma_0\gamma + \mathrm{E}[Var[a|x]].$$

Hence the lower bound can be achieved only if $\mathrm{E}[Var[a|x]] = 0$.

Given that the variance is necessarily non-negative, this requires that $Var[a|x] = 0$ almost surely. This together with the requirement that $\mathrm{E}[a|x] = \gamma^T\mu_0$ almost surely implies that $a = \gamma^T\mu_0$ almost surely. Hence optimality requires that $a = \gamma^T\mu_0$ with probability 1, whenever $\theta \geq 4|\Sigma\gamma|^2$.

C.10 Proof of Lemma 3

Applying Lemma 1, for the $\rho = 1$ case and any $p^1 \in \mathscr{P}(S)^{|X|}$,

$$C_{NG}(p^1, q_0; S, X; \rho = 1) = \sum_{i\in\mathscr{I}} c_i \sum_{x\in X_i} q_x \sum_{s\in S:\pi_{i,s}>0} p^1_{x,s}\ln(\frac{p^1_{x,s}}{\pi^1_{i,s}})$$

where

$$\pi^1_i = \sum_{x\in X_i} p^1_x q_{i,x}(q).$$

Therefore, defining $p^{12} = p^1 \otimes p^2$ as in Definition 3,

$$C_{NG}(p^{12}, q_0; S, X; \rho = 1) = \sum_{i\in\mathscr{I}} c_i \sum_{x\in X_i} q_x \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} p^1_{x,s_1} p^2_{x,s_2} \ln(\frac{p^1_{x,s_1} p^2_{x,s_2}}{\pi^{12}_{i,s_1,s_2}}),$$

where

$$\pi^1_{i,s_1,s_2} = \sum_{x\in X_i} p^1_{x,s_1} p^2_{x,s_2} q_{i,x}(q).$$

It follows that

$$C_{NG}(p^{12},q_0;S,X;\rho=1) - C_{NG}(p^1,q_0;S,X;\rho=1) - C_{NG}(p^2,q_0;S,X;\rho=1) = \\ \sum_{i\in\mathscr{I}} c_i \sum_{x\in X_i} q_x \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} p^1_{x,s_1} p^2_{x,s_2} \ln(\frac{\pi^1_{i,s_1}\pi^1_{i,s_2}}{\pi^{12}_{i,s_1,s_2}}) = \\ -\sum_{i\in\mathscr{I}} c_i \bar{q}_i(q) \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} \pi^{12}_{i,s_1,s_2} \ln(\frac{\pi^{12}_{i,s_1,s_2}}{\pi^1_{i,s_1}\pi^1_{i,s_2}}).$$

This quantity is the negative of the conditional (on being in $i \in \mathscr{I}$) mutual information between s_1 and s_2, and hence is negative, strictly so if the signals are not independent. Therefore, the $\rho = 1$ case exhibits decreasing marginal costs.

Next consider the $\rho = 2$ case:

$$C_{NG}(p^1,q_0;S,X;\rho=2) = \sum_{i\in\mathscr{I}} c_i |X_i|^{-1} \bar{q}_i(q) \sum_{x\in X_i} \sum_{s\in S:\pi_{i,s}>0} \pi^1_{i,s} \ln(\frac{\pi^1_{i,s}}{p^1_{x,s}}),$$

and therefore

$$C_{NG}(p^{12},q_0;S,X;\rho=2) = \sum_{i\in\mathscr{I}} c_i |X_i|^{-1} \bar{q}_i(q) \sum_{x\in X_i} \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} \pi^{12}_{i,s_1,s_2} \ln(\frac{\pi^{12}_{i,s_1,s_2}}{p^1_{x,s_1} p^2_{x,s_2}}).$$

It follows that

$$C_{NG}(p^{12},q_0;S,X;\rho=2) - C_{NG}(p^1,q_0;S,X;\rho=2) - C_{NG}(p^2,q_0;S,X;\rho=2) = \\ \sum_{i\in\mathscr{I}} c_i \bar{q}_i(q) \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} \pi^{12}_{i,s_1,s_2} \ln(\pi^{12}_{i,s_1,s_2}) - \\ \sum_{i\in\mathscr{I}} c_i \bar{q}_i(q) \sum_{(s_1,s_2)\in S\times S:\pi^{12}_{i,s_1,s_2}>0} \pi^1_{i,s_1}\pi^1_{i,s_2} \ln(\pi^1_{i,s_1}\pi^1_{i,s_2}).$$

This quantity is also the conditional (on being in $i \in \mathscr{I}$) mutual information between s_1 and s_2, and hence is positive, strictly so if the signals are not independent. Therefore, the $\rho = 2$ case exhibits increasing marginal costs.

C.11 Proof of Proposition 5

By Definition 1,

$$H^G(q_i;1,|X_i|)+H^G(q_i;2,|X_i|) = (\sum_{j=1}^{|X_i|} q_{i,j}\ln(q_{i,j})) - \frac{1}{|X_i|}\sum_{j=1}^{|X_i|}\ln(q_{i,j})$$

A little algebra shows that

$$\begin{aligned} H^G(q_i;1,|X_i|)+H^G(q_i;2,|X_i|) &= \frac{1}{|X_i|}\sum_{j=1}^{|X_i|}\sum_{k=1}^{|X_i|} q_{i,j}\ln(\frac{q_{i,j}}{q_{i,k}}) \\ &= H^{CM}(q_i(\bar{q});|X_i|). \end{aligned}$$

By the results of Bloedel and Zhong (2020), if the neighborhood-based cost function $H(q;X,\mathcal{N})$ has constant marginal costs, it must satisfy the functional form in (19).

By the arguments used to prove Proposition 1 (invoking Assumption 3), it is without loss of generality to suppose that a set $\bar{X}$, measure $\bar{q} \in \mathbb{R}_+^{|\bar{X}|}$, and surjection $m : \bar{X} \to X$ exists such that the $\bar{X}_i = \bar{X} \cap N_i$ are disjoint and

$$H(q;X,\mathcal{N}) = H(\bar{q};\bar{X},\mathcal{N}),$$

where $\bar{q}_{\bar{x}} = q_{m(x)}$ for all $x \in \bar{X}$. By Assumption 2, we must have $\gamma_{x,x'} = 0$ for any x,x' that do not share a neighborhood. By Proposition 1, within each disjoint neighborhood i we must (due to the symmetry of the local information cost) have $\gamma_{x,x'} = \gamma_{x',x} = \gamma_{x,x''}$ for all $x,x',x'' \in \bar{X}_i$. Consequently, defining $c_i = |X_i|\gamma_{x,x'}$ for some (any) $x, x' \in \bar{X}_i$,

$$H(\bar{q};\bar{X},\mathscr{N}) = \sum_{i\in\mathscr{I}} \frac{c_i}{|X_i|} \sum_{x\in\bar{X}_i} \sum_{x'\in\bar{X}_i\setminus\{x\}} \bar{q}_x \ln(\frac{\bar{q}_x}{\bar{q}_{x'}})$$
$$= \sum_{i\in\mathscr{I}} \frac{c_i}{|X_i|}\bar{q}_i(\bar{q}) \sum_{x\in\bar{X}_i} \sum_{x'\in\bar{X}_i\setminus\{x\}} q_{i,x}(\bar{q}) \ln(\frac{q_{i,x}(\bar{q})}{q_{i,x'}(\bar{q})})$$
$$= \sum_{i\in\mathscr{I}} c_i\bar{q}_i(\bar{q}) H^{CM}(q_i(\bar{q});|X_i|).$$

It follows immediately that for all X, regardless of whether the X_i are disjoint,

$$H(q;X,\mathscr{N}) = \sum_{i\in\mathscr{I}} c_i\bar{q}_i(q)H^{CM}(q_i(q);|X_i|).$$

The representation for V_{CM} with the KL divergence follows immediately (or see Bloedel and Zhong (2020)).

C.12 Proof of Proposition 6

It is convenient to work with the transformed variable

$$y = G(x) = \int_{x_L}^{x} \frac{dx}{q(x)},$$

which is well-defined by the compactness of X and the full-support property of $q(x)$. Define $\bar{y} = G(\bar{x})$ and $g(y) = q(G^{-1}(y))^2$.

The associated problem is

$$\max_{s\in S,K\in\mathbb{R}} \int_0^{\bar{y}} g(y)(K-\beta s(y))dy$$

subject to

$$\inf_{p_L\in\{p\in C^1([0,\bar{y}],(0,\infty)):\int_X g(y)p(y)dy=1\}} \int_0^{\bar{y}} g(y)p_L(y)(s(y)-K)dy + \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p_L'(y))^2}{p_L(y)}dy \geq 0.$$

Note that fixing any K, maximizer on S exists (provided the problem is feasible), by the Lipschitz property of S (which ensures S is compact).

It is without loss of generality to restrict K to the set $[0,\bar{x}]$, because $K<0$ will always be dominated by $K=0$ and $K>\bar{x}$ will never satisfy the constraint (due to the limited liability of $s\in S$). Therefore, a maximizing (s,K) exists.

We proceed by taking $K\geq 0$ as given, and determining the optimal security s, and then consider the optimal choice of K. If $K=0$, the optimal security is $s(y)=0$ for all $y\in[0,\bar{y}]$.

Suppose $K>0$. As argued in the text, it is without loss of generality to assume the constraint binds, and hence that the results of Lemma 2 apply. This can only occur if $s(\bar{y})>K$, as otherwise rejection must be a strictly dominating action.

Defining $q(y)=\bar{y}^{-1}$ and $u_R(y)=g(y)(s(y)-K)$, there exists a function $\psi:[0,\bar{y}]\to\mathbb{R}$ satisfying $\psi(0)=\psi(\bar{y})=0$ and

$$\frac{1}{\bar{y}}\psi(y)=\int_0^{\bar{y}}\frac{1}{\bar{y}}[\frac{2}{\theta}g(y')(s(y')-K)-\frac{1}{2}\psi(y')^2]dy.$$

We begin by proving that $\psi(y)<0$. Define the function

$$\phi(y)=\exp(\frac{1}{2}\int_0^{y}\psi(y')dy),$$

and observe that

$$\frac{2\phi'(y)}{\phi(y)}=\psi(y).$$

Plugging $\phi(y)$ into the Euler-Lagrange equation for ψ,

$$\frac{d}{dy}[\frac{2\phi'(y)}{\phi(y)}]=2\theta^{-1}(s(y)-K)g(y)-\frac{1}{2}(\frac{2\phi'(y)}{\phi(y)})^2,$$

which simplifies to

$$\phi''(y)=2\theta^{-1}(s(y)-K)g(y)\phi(y).$$

By definition, $\phi'(0)=\phi'(\bar{y})=0$. By the single-crossing property of $g(y)(s(y)-K)\phi(y)$, $\phi(y)$ must be strictly concave wherever $s(y)<K$ and convex wherever $s(y)>K$. A single crossing must exist, by the continuity of s.

Consequently, if $K>0$, we must have $\phi'(y)<0$ and hence that $\psi(y)<0$. In the case of $K = 0$, $s(y) = 0$, Lemma 2 holds with $\psi(y) = 0$ for all $y \in [0, \bar{y}]$.

The Hamiltonian can be written as

$$H(s, \psi, v, \lambda_1, \lambda_2, y; K) = g(y)(K - \beta s) + \lambda_1 v + \lambda_2 (\frac{2}{\theta}(s - K) g(y) - \frac{1}{2}\psi(y)^2).$$

The constraints on v are $v \geq 0$ and $v \leq \sqrt{g(y)}$, ensuring that $s'(y) y'(x) \in [0, 1]$.

The associated necessary conditions are

$$\lambda_1(y) + \rho_0(y) - \rho_1(y) = 0,$$

where $\rho_0(y)$ and $\rho_1(y)$ are the multipliers on the constraints $v \geq 0$ and $v \leq \sqrt{g(y)}$, respectively, and

$$\begin{aligned} -\lambda_1'(y) &= g(y)(\frac{2}{\theta}\lambda_2(y) - \beta), \\ -\lambda_2'(y) &= -\lambda_2(y)\psi(y). \end{aligned} \tag{35}$$

The associated boundary conditions are $\psi(0) = \psi(\bar{y}) = 0$, $s(0) = 0$, and $\lambda_1(\bar{y}) = 0$.

If $\lambda_2(0) \leq 0$, by (35) we will have $\lambda_2(y) \leq 0$ for all $y \in [0, \bar{y}]$, implying $\lambda_1'(y) > 0$ for all $y \in [0, \bar{y}]$. By the boundary condition $\lambda_1(\bar{y}) = 0$, this requires $\lambda_1(y) < 0$ for all $y \in [0, \bar{y})$, and hence $s'(y) = 0$ for all such y. It follows in this case that $s(y) = 0$ for all $y \in [0, \bar{y}]$. This can occur only if $K = 0$.

If $K > 0$, we must have $\lambda_2(0) > 0$, and hence that $\lambda_2(y) > 0$ for all y. In this case, $\lambda_2(y)$ must be strictly decreasing (by $\psi(y) < 0$), and hence crosses β at most once. It follows that for some $\hat{y} \in [0, \bar{y}]$, $\lambda_1'(y) < 0$ on some interval $[0, \hat{y})$, if such an interval exists, and $\lambda_1'(y) > 0$ on the interval $[\hat{y}, \bar{y}]$, if such an interval exists. By the boundary condition $\lambda_1(\bar{y}) = 0$, three outcomes are possible: (1) $\lambda_1(y) < 0$ on $y \in [0, \bar{y}]$, or (2) $\lambda_1(y) > 0$ on $[0, \bar{y}]$, or (3) $\lambda_1(y) > 0$ for all $y \in [0, y^*)$ and $\lambda_1(y) < 0$ for all $y \in [y^*, \bar{y}]$. The first of these, however, is ruled out by $s(\bar{y}) > K > 0$, as above.

The optimal securities in cases (2) and (3) are described by $s(y) = \int_0^{\min\{y, y^*\}} g(y)^{\frac{1}{2}} dy$, for some $y^* \in (0, \bar{y}]$. This integral can be written as

$$
\begin{aligned}
s(y) &= \int_0^{\min\{G^{-1}(y), G^{-1}(y^*)\}} dx \\
&= \min\{G^{-1}(y), G^{-1}(y^*)\},
\end{aligned}
$$

and therefore

$$s(x) = \min\{x, x^*\}$$

for $x^* = G^{-1}(y^*) \in (0, \bar{x}]$. By $s(\bar{x}) > K$, we must have $x^* > K > 0$.

Hence, the optimal security design is either a debt with a strictly positive price, or the zero contract with zero price.

Suppose $K = 0$ was optimal. We would require $\lambda_1^*(y) \leq 0$ for all $y \in [0, \bar{y}]$, and therefore

$$\int_0^{\bar{y}} g(y)(\frac{2}{\theta}\lambda_2(y) - \beta)dy \leq 0.$$

The first order condition for K requires that

$$\frac{\partial}{\partial K}\int_0^{\bar{y}} H(s^*(y), \psi^*(y), v^*(y), \lambda_1^*(y), \lambda_2^*(y), y; K)dy < 0,$$

and therefore

$$\int_0^{\bar{y}} g(y)(1 - \lambda_2^*(y)\frac{2}{\theta})dy = 0.$$

By $\beta < 1$, this is a contradiction, and therefore $K > 0$ and $s(\bar{x}) > K$.

D Technical Appendix

D.1 Convergence to the Continuous State Model

For each of a sequence of values for the integer M, we assume a neighborhood structure of the kind discussed in section 2.3 with $M+1$ states. The set of states is ordered, $X^M = \{0, 1, \ldots, M\}$, and each pair of adjacent states forms a neighborhood, $X_i = \{i, i+1\}$, for all $i \in \{0, 1, \ldots, M-1\}$. We will also assume that there is an $M+1$st neighborhood containing all of the states. Note that M indexes both the number of states and the number of neighborhoods. We consider the limit as $M \to \infty$.

To study this limit, we need to define how the prior beliefs, q_M, and the magnitude of the information costs vary with M. For the initial beliefs, we shall assume that there is a differentiable probability density function $q : [0,1] \to \mathbb{R}^+$, with full support on the unit interval and with a derivative that is Lipschitz continuous.

For this section and its proofs, we will use the notation e_i to indicate a basis vector equal to one for the i-th element of X^M and zero otherwise. Using the q function, we define, for any $i \in X^M$, the prior $q_M \in \mathscr{P}(X^M)$ by

$$e_i^T q_M = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)dx.$$

That is, for each value of M, the prior q_M is assumed to be a discrete approximation to the p.d.f. $q(x)$, which becomes increasingly accurate as $M \to \infty$.

For our neighborhood structures, we assume that that the constants associated with the cost of each neighborhood, c_j, are equal to M^2 for all $j < M$, and M^{-1} for $j = M$. In this particular example, the scaling ensures that the DM is neither able to determine the state with certainty, nor prevented from gathering any useful information, even as M is made arbitrarily large; moreover, the scaling ensures that the neighborhood containing all states plays no role in the limiting behavior, so that in the limit all information costs are local. We also scale the entire cost function by a constant, $\theta > 0$.

We also need to define the set of actions, and the utility from those actions. We will assume the set of actions, A, remains fixed as N grows, and define the utility from a particular action, in a particular state, as

$$e_i^T u_{a,M} = \frac{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)u_a(x)dx}{e_i^T q_M}.$$

Here, the utility $u_a : [0,1] \to \mathbb{R}$ is a bounded measurable function for each action $a \in A$.[36] In other words, as M grows large, the prior converges to $q(x)$ and the utilities converge to the functions $u_a(x)$.

We consider only the case of generalized entropy index neighborhood cost functions with $\rho = 1$ (see Definition 2). Under these assumptions, the static model of section §1 can be written as

$$V_N(q_M;M) = \max_{\pi_M \in \mathscr{P}(A), \{q_{a,M} \in \mathscr{P}(X^M)\}_{a\in A}} \sum_{a\in A} \pi_M(a)(u_{a,M}^T \cdot q_{a,M}) - \theta \sum_{a\in A} \pi_M(a) D_{NG}(q_{a,M}||q_M; \rho = 1, X^M, \mathscr{N}^M), \tag{36}$$

subject to the constraint that

$$\sum_{a\in A} \pi_N(a) q_{a,M} = q_M.$$

The following theorem shows that the solution to this problem, both in terms of the value function and the optimal policies, converges to the solution of a static rational inattention problem with a continuous state space.

Proposition 7. *Consider the sequence of finite-state-space static rational inattention problems (36), with progressively larger state spaces indexed by the natural numbers M. There exists a sub-sequence of integers $n \in \mathbb{N}$ for which the solutions to the sub-sequence of problems converge, in the sense that, for some $\pi^* \in \mathscr{P}(A)$ and $\{q_a^* \in \mathscr{P}([0,1])\}_{a\in A}$,*

[36] Note that we do not require the payoff resulting from an action to be a continuous function of x at all points, though it will be continuous almost everywhere. This allows for the possibility that a DM's payoffs change discontinuously when the state x crosses some threshold, as in some of our applications.

i) $\lim_{n\to\infty} V_N(q_n;n) = V_N(q)$*;*

ii) $\lim_{n\to\infty} \pi_n^* = \pi^*$*; and*

iii) for all $a \in A$ *and all* $x \in [0,1]$*,* $\lim_{n\to\infty} \sum_{i=0}^{\lfloor xn \rfloor} e_i^T q_{a,n}^* = \int_0^x q_a^*(y)dy$*.*

Moreover, the limiting value function $V_N(q)$ *is the value function for the following continuous-state-space static rational inattention problem:*

$$V_N(q) = \sup_{\pi \in \mathscr{P}(A), \{q_a \in \mathscr{P}_{LipG}([0,1])\}_{a\in A}} \sum_{a\in A} \pi(a) \int_{supp(q)} u_a(x) q_a(x) dx$$
$$-\frac{\theta}{4} \sum_{a\in A} \{\pi(a) \int_0^1 \frac{(q_a'(x))^2}{q_a(x)} dx\} + \frac{\theta}{4} \int_0^1 \frac{(q'(x))^2}{q(x)} dx,$$

subject to the constraint that, for all $x \in [0,1]$*,*

$$\sum_{a\in A} \pi(a) q_a(x) = q(x), \tag{37}$$

and where $\mathscr{P}_{LipG}([0,1])$ *denotes the set of differentiable probability density functions with full support on* $[0,1]$*, whose derivatives are Lipschitz-continuous. Furthermore, the limiting action probabilities* $\pi^*(a)$ *and posteriors* q_a^* *are the optimal policies for this continuous-state-space problem.*

Proof. See the technical appendix, section D.4.1. □

This theorem demonstrates that the value function, choice probabilities, and posterior beliefs of the discrete state problem converge to the value function, choice probabilities, and posterior beliefs associated with a continuous state problem. The continuous state problem uses a particular cost function, the expected value of the Fisher information $I^{Fisher}(x;p)$, defined locally for each element of the continuum of possible states x, with the expectation taken with respect to the prior over possible states. The posterior beliefs in the continuous state problem, $q_a(x)$, are required to be differentiable, with a Lipschitz-continuous derivative, on their support. This is a result; the limiting posterior beliefs of the discrete state problem will have these properties. This restriction also ensures that the Fisher information is finite, so that the optimization associated with the continuous state problem is well-behaved.

The static rational inattention problem for the limiting case of a continuous state space can be given an alternative, equivalent formulation, in which the objects of choice are the conditional probabilities of taking different actions in the different possible states, rather than the posteriors associated with different actions. This is essentially the continuous state analog of Lemma 1.

Lemma 10. *Consider the alternative continuous-state-space static rational inattention problem:*

$$\bar{V}_N(q) \;=\; \sup_{p \in \mathscr{P}_{LipG}(A)} \int_0^1 q(x) \sum_{a \in A} p_a(x) u_a(x) dx - \frac{\theta}{4} \int_0^1 q(x)\, I^{Fisher}(x;p) dx,$$

where $\mathscr{P}_{LipG}(A)$ is the set of mappings $\{p_a : [0,1] \to [0,1]\}_{a \in A}$ such that for each action a, the function $p_a(x)$[37] *is either everywhere zero or a strictly positive differentiable function of x with a Lipschitz-continuous derivative, and for any information structure $p \in \mathscr{P}_{LipG}(A)$, the* Fisher information *at state $x \in X$ is defined as*

$$I^{Fisher}(x;p) \;\equiv \sum_{a \in A : p_a(x) > 0} \frac{(p'_a(x))^2}{p_a(x)}.$$

This problem is equivalent to the one defined in Theorem 7, in the sense that the information structure p^* that is the limiting optimal policy of this problem defines action probabilities and posteriors

$$\pi^*(a) \;=\; \int_0^1 q(x) p^*_a(x), \qquad q^*_a(x) \;=\; \frac{q(x) p^*_a(x)}{\pi^*(a)} \tag{38}$$

that solve the problem in Theorem 7, and conversely, the action probabilities and posteriors $\{\pi^*(a), q^*_a\}$ that solve the problem stated in the theorem define state-contingent action probabilities

$$p^*_a(x) \;=\; \frac{\pi^*(a) q^*_a(x)}{q(x)} \tag{39}$$

that are the limiting optimal policies in the problem stated here. Moreover, the

[37] Here for any $x \in [0,1]$, we use the notation $p_a(x)$ to indicate the probability of action a implied by the probability distribution $p(x) \in \mathscr{P}(A)$.

maximum achievable value is the same for both problems: $\bar{V}_N(q) = V_N(q)$.

Proof. See the appendix, section D.4.2. $\square$

D.2 Additional Technical Lemmas for Proposition 1 and Corollary 4

D.2.1 Proof of Lemma 4

We first state the lemma.

Lemma. *Let* $C(p, q_0; S, X)$ *be any cost function satisfying Assumption 1 (i.e. any continuously twice-differentiable UPS cost function). Suppose that, for all* $x \in X$,

$$p_x = r + \varepsilon v_x$$

for some $\varepsilon > 0$, $r \in \mathscr{P}(S)$ *with full support on S, and* $v_x \in \mathbb{R}^{|S|}$, *and that* q_0 *has full support on* X. *Then for the matrix-valued function*

$$k(q) = Diag(q) \cdot H_{qq}(q; X, \mathscr{N}) \cdot Diag(q),$$

where $Diag(q)$ *is the diagonal matrix with q on its diagonal and* $H_{qq}(q; X, \mathscr{N})$ *is the Hessian of the H function associated with C,*

$$C(p, q_0; S) = \frac{1}{2}\varepsilon^2 \sum_{x \in X, x' \in X} k_{x,x'}(q_0) \nu_x^T \cdot Diag(r)^{-1} \cdot \nu_{x'} + o(\varepsilon^2),$$

where $Diag(r)$ *is a diagonal matrix with r on the diagonal and* ι *is a vector of ones.*

Under the stated assumptions,

$$p_{s,x} = r_s + \varepsilon \nu_{s,x} + o(\varepsilon).$$

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

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

where

$$\pi_s(p,q) = r_s + \varepsilon \sum_{x' \in X} v_{s,x'} q_{x'}.$$

It follows immediately that

$$\lim_{\varepsilon \to 0^+} q_{s,x}(p,q) = \frac{q_x r_s}{r_s} = q_x.$$

We also have

$$\varepsilon^{-1}(q_{s,x}(p,q) - q_x) = \frac{q_x(v_{s,x} - \sum_{x' \in X} v_{s,x'} q_{x'})}{\pi_s(p,q)},$$

and therefore for any s,

$$\lim_{\varepsilon \to 0^+} \varepsilon^{-1}(q_{s,x}(p,q) - q_x) = \frac{q_x(v_{s,x} - \sum_{x' \in X} v_{s,x'} q_{x'})}{r_s}$$

and hence

$$q_{s,x}(p,q) - q_x = q_x \frac{v_{s,x} - \sum_{x' \in X} v_{s,x'} q_{x'}}{r_s} + o(\varepsilon).$$

In matrix form, where $v_s \in \mathbb{R}^{|X|}$ is the vector of $\{v_{s,x}\}_{x \in X}$,

$$(q_s(p,q) - q) = \frac{1}{r_s} Diag(q) \cdot (v_s - \iota q^T v_s) + o(\varepsilon).$$

By Assumption 1,

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

Taylor-expanding up to second-order,

$$C(p,q_0;S) = \frac{1}{2} \sum_{s \in S} \frac{r_s}{r_s^2} (v_s - \iota q^T v_s)^T \cdot Diag(q_0) \cdot H_{qq}(q_0) \cdot Diag(q_0) \cdot (v_s - \iota q^T v_s) + o(\varepsilon^2).$$

Recalling that H is homogenous of degree one, we must have

$$\iota^T Diag(q_0) H_{qq}(q_0) = q_0^T H_{qq}(q_0) = \vec{0},$$

and consequently this expression is

$$C(p, q_0; S) = \frac{1}{2} \sum_{s \in S} \sum_{x \in X, x' \in X} k_{x.x'}(q_0) \frac{1}{r_s} v_{s,x} v_{s,x'} + o(\varepsilon^2),$$

which is the result.

D.2.2 Proof of Lemma 9

We first state the lemma.

Lemma. *Let Λ_0 be a $k \times k$ real symmetric positive-semidefinite matrix, let $\mathscr{M}_k$ be the set of $k \times k$ real symmetric positive-definite matrices, and let $v \in \mathbb{R}^k$ be a vector. Then*

$$2|v| = \inf_{\Lambda \in \mathscr{M}_k} v^T \Lambda^{-1} v + tr[\Lambda]$$

Proof. Let $\frac{v}{|v|} = z_1, z_2, \ldots, z_k$ be an orthonormal basis, and let V be the associated orthonormal matrix ($V^T V = I$) whose columns are the basis vectors. Suppose there is a minimizer, Λ^*, with

$$\Lambda^* = VMV^T$$

for some positive-definite, real symmetric M.

Consider a perturbation

$$\Lambda(\varepsilon) = \Lambda^* + \varepsilon VMzz^T MV^T$$

for some arbitrary vector z. Such a perturbation is always feasible for $\varepsilon > 0$, and is feasible for $\varepsilon < 0$ if

$$z^T MV^T \Lambda^* VMz > 0.$$

We have

$$\frac{d}{d\varepsilon}(\Lambda(\varepsilon))^{-1}|_{\varepsilon=0} = -(\Lambda^*)^{-1}VMzz^TMV^T(\Lambda^*)^{-1}.$$

Observing that

$$(\Lambda^*)^{-1} = VM^{-1}V^T$$

and using the orthonormality of V,

$$\frac{d}{d\varepsilon}(\Lambda(\varepsilon))^{-1}|_{\varepsilon=0} = -Vzz^TV^T.$$

It follows that optimality requires

$$-v^TVzz^TV^Tv + tr[VMzz^TMV^T] \geq 0,$$

with equality if the perturbation is feasible in both directions.

Because v is a basis vector of the orthonormal basis that defines V,

$$v^TV = \frac{v^Tv}{|v|}e_1^T,$$

where e_1 is a basis vector with one in index 1 and zero otherwise. Again using orthonormality to insert $V^TV = I$, we must have

$$-|v|^2e_1^Tzz^Te_1 + tr[VMV^TVzz^TV^TVMV^T] \geq 0,$$

which simplifies to

$$|v|^2e_1^Tzz^Te_1 \leq tr[\Lambda^*Vzz^TV^T\Lambda^*],$$

which is

$$z^T(V^T\Lambda^*\Lambda^*V - |v|^2e_1e_1^T)z \geq 0.$$

It follows that for all z with $e_1^Tz = 0$, we must have

$$z^TV^T\Lambda^*\Lambda^*\Lambda^*Vz = 0,$$

which requires

$$z_j^T \Lambda^* \Lambda^* \Lambda^* z_j = 0$$

for all $j \neq 1$. It follows immediately that the nullity of Λ^* is at least $k-1$, and hence that the rank is at most one. Conjecture therefore that

$$\Lambda^* = xx^T$$

for some vector x. The objective is

$$\lim_{\varepsilon \to 0^+} v^T (\varepsilon I + xx^T) v + x^T x,$$

which by the Sherman-Morrison lemma is

$$\lim_{\varepsilon \to 0^+} \varepsilon^{-1} v^T v - \frac{\varepsilon^{-2} v^T x x^T v}{1 + \varepsilon^{-1} x^T x} + x^T x.$$

By Cauchy-Schwarz,

$$\varepsilon^{-1} v^T v - \frac{\varepsilon^{-2} v^T x x^T v}{1 + \varepsilon^{-1} x^T x} \geq \frac{\varepsilon^{-1} v^T v}{1 + \varepsilon^{-1} x^T x},$$

and therefore holding fixed $|x|$ is optimal to set

$$\frac{x}{|x|} = v,$$

and the problem solves

$$\inf_{|x|^2 \geq 0} \frac{|v|^2}{|x|^2} + |x|^2,$$

and hence

$$|x|^2 = |v|.$$

It follows that

$$\inf_{\Lambda \in \mathscr{M}_k} v^T \Lambda^{-1} v + tr[\Lambda] = 2|v|.$$

$\square$

D.3 Additional Technical Lemmas for Binary Choice

D.3.1 Proof of Lemma 6

We first repeat the lemma to be proven (from the proof of Proposition 3).

Lemma. *If*

$$\sup_{p_R\in C^1([0,\bar{y}],(0,1))} J[p_R] > \max\{\int_0^{\bar{y}} g(y)u_R(y)dy, 0\},$$

then there exists an extremal $p_R^* \in C^1([0,\bar{y}],(0,1))$ that is a maximizer and is continuously twice-differentiable except at the discontinuities of $u_R(y)$.

We begin by proving the existence of a maximizer of $J[p_R]$ on $C^1([0,\bar{y}],(0,1))$. Observe first by the concavity of J (Lemma 5) that any extremal must a maximizer.

Let us define a transformed domain for the problem, $\phi \in C^1([0,\bar{y}],\mathbb{R})$, by

$$\phi(y) = \cos^{-1}(\sqrt{p_R(y)}),$$

which satisfies

$$\phi'(y) = -\frac{p_R'(y)}{2\sqrt{p_R(y)(1-p_R(y))}}.$$

The corresponding functional is

$$\hat{J}[\phi] = \int_0^{\bar{y}} g(y)\cos(\phi(y))^2 u_R(y)dy - \theta\int_0^{\bar{y}} \phi'(y)^2 dy. \tag{40}$$

Consider the relaxed problem, for some $y_H > \bar{y} > 0 > y_L$,

$$\inf_{\phi\in C^1([0,\bar{y}+\varepsilon],\mathbb{R})} \int_{y_L}^{y_H} F(y,\phi(y),\phi'(y))dy$$

where

$$F(y,\phi,v) = \begin{cases} \theta v^2 - g(y)\cos(\phi)^2 u_R(y) & y\in[0,\bar{y}], \\ \theta v^2 & y\notin[0,\bar{y}]. \end{cases}$$

Note that this problem does not restrict the range of ϕ, but it is without loss of generality to assume that $\phi(y)\in[0,\frac{1}{2}\pi]$ for all $y\in[y_L,y_H]$. The problem is relaxed by the possibility that $\phi(y) = 0$ or $\phi(y) = \frac{1}{2}\pi$ (which corresponds to $p(y) = 0$ or $p(y) = 1$) and extended to the domain $[y_L, y_H]$.

Because it is without loss of generality to assume bounded $\phi(y)$, and always optimal to satisfy

$$\int_{y_L}^{y_H} \phi'(y)^2 dy < \infty,$$

it is without loss of generality to assume $\phi \in W^{1,2}([y_L, y_H], \mathbb{R})$ (the Sobolev space with square-integrable weak first derivatives).

Observing that $F(y, \phi, v)$ is convex in v and satisfies, for $B = \max_{y \in [0,\bar{y}]} |g(y)u_R(y)|$,

$$F(y, \phi, v) \geq \theta v^2 - B.$$

By theorem 4.1 of Dacorogna (2007), for any given values ϕ_L and ϕ_H, the problem

$$\inf_{\phi \in \{W^{1,2}([y_L, y_H], \mathbb{R}): \phi(y_L) = \phi_L, \phi(y_H) = \phi_H\}} \int_{y_L}^{y_H} F(y, \phi(y), \phi'(y)) dy$$

has a minimizer (where ϕ' is understood as a weak derivative). Minimizing over the compact set $(\phi_L, \phi_H) \in [0, \frac{1}{2}\pi]^2$ demonstrates that a minimizer exists for $W^{1,2}([y_L, y_H], \mathbb{R})$.

We next invoke the following lemma to show that the minimizer ϕ^* is in fact continuously differentiable, and continuously twice-differentiable everywhere $u_R(y)$ is continuous.

Lemma 11. *If $\phi^* \in W^{1,2}([-\varepsilon, \bar{y} + \varepsilon], \mathbb{R})$ is a minimizer of the functional $\hat{J}$ defined above, then $\phi^* \in C^1([-\varepsilon, \bar{y} + \varepsilon], \mathbb{R})$, and ϕ^* is continuously twice-differentiable on any interval on which u_R is continuous.*

Proof. See the Technical Appendix, Section D.3.3, defining (in the context of that proof) $u(y) = g(y)u_R(y)$. □

Let $y_1, \ldots y_{k-1}$ be the (possibly empty) set of points of discontinuity for u_R, and let $y_0 = y_L$ and $y_k = y_H$. This regularity result implies that the Euler-Lagrange equation,

$$\phi^{*\prime\prime}(y) = g(y)\sin(2\phi^*(y))u_R(y)$$

must hold on all $y \in (y_{i-1}, y_i)$.

Suppose that for some $y \in [0, \bar{y}]$, $\phi^*(y) \in \{0, \frac{\pi}{2}\}$. By the fact that $\phi^*(y)$ is continuously differentiable and it is without loss of generality to assume $\phi^*(y) \in [0, \frac{\pi}{2}]$, it must be the case that $\phi^{*\prime}(y) = 0$ if $\phi^*(y) \in \{0, \frac{\pi}{2}\}$. In this case, $\phi^*(y)$ constant on $y \in [y_L, y_H]$ satisfies the Euler-Lagrange equations. The system

$$\frac{d}{dy}\begin{bmatrix} \phi^{*\prime}(y) \\ \phi^*(y) \end{bmatrix} = \begin{bmatrix} g(y)\sin(2\phi^*(y))u_R(y) \\ \phi^{*\prime}(y) \end{bmatrix}$$

is uniformly Lipschitz-continuous in $(\phi^*, \phi^{*\prime})$ and continuous in y on all intervals (y_{i-1}, y_i), and hence by the Picard-Lindelof theorem, a unique solution to the initial value problem on any interval $[y_{i-1}, y_i]$ exists. Consequently, if $\phi^*(y) \in \{0, \frac{\pi}{2}\}$ for any $y \in [0, \bar{y}]$, $\phi^*(y) \in \{0, \frac{\pi}{2}\}$ for all $y \in [0, \bar{y}]$.

But by the assumption that

$$\sup_{p_R \in C^1([0,\bar{y}],(0,1))} J[p_R] > \max\{\int_0^{\bar{y}} g(y)u_R(y)dy, 0\},$$

a constant solution cannot be a optimal. Therefore, $\phi^*(y) \in (0, \frac{\pi}{2})$ for all $y \in [y_L, y_H]$. Consequently, the function $p^* \in C^1([0, \bar{y}], (0,1))$ defined by

$$p^*(y) = \cos(\phi^*(y))^2$$

for $y \in [0, \bar{y}]$ is a maximizer of $J[\cdot]$ and is continuously twice-differentiable everywhere $u_R(y)$ is continuous.

D.3.2 Proof of Lemmas 5 and 7

We prove that the functionals $J_0 : C^1([0, \bar{y}], (0,1)) \to \mathbb{R}$ and $J_1 : C^1([x_L, x_H], (0, \infty)) \to \mathbb{R}$, defined by

$$\begin{aligned} J_0[p] &= \int_0^{\bar{y}} g(y)p(y)u_R(y)dy - \frac{\theta}{4}\int_0^{\bar{y}} \frac{(p'(y))^2}{p(y)(1-p(y))}dy, \\ J_1[p] &= -\int_{x_L}^{x_H} q(x)p(x)u_R(x)dx - \frac{\theta}{4}\int_{x_L}^{x_H} q(x)\frac{(p'(x))^2}{p(x)}dx, \end{aligned}$$

are concave.

Proof. Per chapter 4, section 2.2 of Giaquinta and Hildebrandt (1996), a sufficient condition for the concavity of the functional

$$J[p] = \int_0^{\bar{y}} F(y, p(y), p'(y))dy$$

is that the Hessian

$$\begin{bmatrix} F_{22} & F_{23} \\ F_{32} & F_{33} \end{bmatrix}$$

be negative semi-definite for all p on the relevant domain. In this context,

$$F_0(y, p, v) = g(y)u_R(y)p - \frac{\theta}{4}\frac{v^2}{p(1-p)},$$

and therefore

$$\begin{bmatrix} F_{0,22}(y,p,v) & F_{0,23}(y,p,v) \\ F_{0,32}(y,p,v) & F_{0,33}(y,p,v) \end{bmatrix} = -\frac{\theta}{4}\begin{bmatrix} \frac{2}{p(1-p)} & -\frac{2v(1-2p)}{(p(1-p))^2} \\ -\frac{2v(1-2p)}{(p(1-p))^2} & \frac{2v^2(1-2p)^2}{(p(1-p))^3} + \frac{2v^2}{(p(1-p))^2} \end{bmatrix}.$$

The trace (and hence sum of the eigenvalues) is strictly negative for all $p \in (0,1)$, and the determinant (and hence product of the eigenvalues) is positive (strictly so if $v^2 > 0$), implying that all eigenvalues are weakly negative and hence that the matrix is negative semi-definite.

Similarly,

$$\begin{bmatrix} F_{1,22}(x,p,v) & F_{1,23}(x,p,v) \\ F_{1,32}(x,p,v) & F_{1,33}(x,p,v) \end{bmatrix} = -\frac{\theta}{2}q(x)\begin{bmatrix} \frac{v^2}{p^3} & -\frac{v}{p^2} \\ -\frac{v}{p^2} & \frac{1}{p} \end{bmatrix}.$$

On $p > 0$, the determinant is zero and trace negative, and hence one eigenvalue is negative and the other is zero, implying this matrix is negative semi-definite. $\square$

D.3.3 Proof of Lemmas 11 and 8

Let $u : [y_L, y_H] \to \mathbb{R}$ be a bounded function with finitely many discontinuities. If for some $\varepsilon > 0$, $\theta > 0$, $\phi^* \in W^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$ is a minimizer of

$$J[p] = \int_{y_L-\varepsilon}^{y_H+\varepsilon} F(y, p(y), p'(y))dy,$$

where either $F = F_1$ or $F = F_0$,

$$F_0(y, \phi, v) = \begin{cases} \theta v^2 - u(y)\cos(\phi)^2 & y \in [y_L, y_H] \\ \theta v^2 & y \notin [y_L, y_H] \end{cases}$$

$$F_1(y, \phi, v) = \begin{cases} \theta v^2 + u(y)\phi^2 & y \in [y_L, y_H] \\ \theta v^2 & y \notin [y_L, y_H] \end{cases}$$

then $\phi^* \in C^1([y_L - \varepsilon, y_L + \varepsilon], \mathbb{R})$, and ϕ^* is continuously twice-differentiable on any interval on which u_R is continuous.

Proof. The functional $F(y, \phi, v)$ satisfies the growth conditions of theorem 4.12 of Dacorogna (2007). Define, for any $R > 0$,

$$\alpha_1(y) = 2\max\{R^2, 1\}|u(y)|,$$

For all $|\phi| \leq R$,

$$\begin{aligned} |F(y, \phi, v)| &\leq \alpha_1(y) + 2\theta v^2, \\ |F_\phi(y, \phi, v)| &\leq \alpha_1(y) + 2\theta v^2, \\ |F_v(y, \phi, v)|| &\leq 2\theta |v|. \end{aligned}$$

Consequently, by theorem 4.12 of Dacorogna (2007), for all $\omega \in W_0^{1,2}([y_L - \varepsilon, y_H + \varepsilon], \mathbb{R})$ (the set of $W^{1,2}$ functions with $\omega(y_L - \varepsilon) = \omega(y_H + \varepsilon) = 0$), the integrated

Euler-Lagrange equation holds:

$$\int_{y_L-\varepsilon}^{y_H+\varepsilon}[F_\phi(y,\phi^*(y),\phi^{*\prime}(y))\omega(y)+2\theta\phi^{*\prime}(y)\omega'(y)]dy=0.$$

Consider the particular test function defined by some $y_L \le y < y' \le y_H$,

$$\omega'(x)=\begin{cases}0 & y\in[y_L-\varepsilon,y),\\ 1 & y\in[y,y'),\\ 0 & y\in[y',y_H],\\ -\frac{y'-y}{\varepsilon} & y\in(y_H,y_H+\varepsilon].\end{cases}$$

It is immediate from the definition of $F(y,\phi,v)$ that if ϕ^* is a minimizer it must satisfy $\phi^{*\prime}(y)=0$ for all $y\notin[y_L,y_H]$. Consequently, for this test function,

$$\begin{aligned}-\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))\omega(y)dy&=2\theta\int_x^{x'}\phi^{*\prime}(x)dx\\&=2\theta(\phi^*(x')-\phi^*(x)).\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}|\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))\omega(y)dy|&\le(\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))^2dy)^{\frac{1}{2}}(\int_{y_L}^{y_H}\omega(y)^2dy)^{\frac{1}{2}}\\&\le(\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))^2dy)^{\frac{1}{2}}(y'-y).\end{aligned}$$

Define $B=\max_{y\in[y_L,y_H]}|u(y)|$. For $F=F_0$, $|F_\phi(y,\phi,v)|=|u(y)\sin(2\phi)|\le B$, and consequently

$$(\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))^2dy)^{\frac{1}{2}}\le B(y_H-y_L).$$

For $F=F_1$, $|F_\phi(y,\phi,v)|=|2u(y)\phi|$, and

$$\int_{y_L}^{y_H}F_\phi(y,\phi^*(y),\phi^{*\prime}(y))^2dy\le 4B^2\int_{y_L}^{y_H}\phi^*(y)^2dy,$$

and consequently

$$(\int_{y_L}^{y_H} F_\phi(y, \phi^*(y), \phi^{*\prime}(y))^2 dy)^{\frac{1}{2}} < K$$

for some constant $K > 0$ by the square integrability of ϕ^*. We conclude that ϕ^* is Lipschitz-continuous.

Let $y_1, \ldots y_{k-1}$ be the (possibly empty) set of points of discontinuity for u, and let $y_0 = y_L$ and $y_k = y_H$. On the intervals (y_{i-1}, y_i) for $i \in \{1, \ldots, k\}$, the derivative F_ϕ is continuous. Following the arguments of propositions 1-3 in section 3.1, chapter 1 of Giaquinta and Hildebrandt (1996) proves that ϕ^* is continuously twice-differentiable on (y_{i-1}, y_i) for all $i \in \{1, \ldots, k\}$.[38] Moreover, the Euler-Lagrange equation

$$2\theta\phi^{*\prime\prime}(y) = F_\phi(y, \phi^*(y), \phi^{*\prime}(y))$$

must hold on all $y \in (y_{i-1}, y_i)$.

By the Weierstrauss-Erdmann corner conditions (or see also proposition 1 in section 3.1, chapter 1 of Giaquinta and Hildebrandt (1996)), at a hypothetical corner at y_i, we would have

$$F_v(y_i, \phi^*(y_i), v_i^-) = F_v(y_i, \phi^*(y_i), v_i^+),$$

where $v_i^- = \lim_{y \uparrow y_i} \phi^{*\prime}(y)$ and $v_i^+ = \lim_{y \downarrow y_i} \phi^{*\prime}(y)$. It follows immediately no corners exist, and hence that $\phi^{*\prime}(y)$ is continuous. □

D.4 Additional Definition and Lemmas for Convergence

Definition 4. Let X^M be a sequence of state spaces, as described in section 3.3. A sequence of policies $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$ satisfies the "convergence condition" if:

i) The sequence satisfies, for some constants $c_H > c_L > 0$, all M, and all $i \in X^M$,

$$\frac{c_H}{M+1} \geq e_i^T p_M \geq \frac{c_L}{M+1}.$$

[38] In the aforementioned section of Giaquinta and Hildebrandt (1996), it is assumed that $F(y, \phi, v)$ is continuously differentiable. However, the proofs given in that section require only that F_ϕ and F_v be continuous, and not that $F(y, \phi, v)$ be differentiable in y.

ii) The sequence satisfies, for some constant $K_1 > 0$, all M, and all $i \in X^M \setminus \{0, M\}$,

$$M^3 \left|\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M\right| \leq K_1,$$

and

$$M^2 \left|\frac{1}{2}(e_M^T - e_{M-1}^T)p_M\right| \leq K_1$$

and

$$M^2 \left|\frac{1}{2}(e_1^T - e_0^T)p_M\right| \leq K_1.$$

Definition 5. Let $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$ be a sequence of probability distributions over the state spaces associated with Theorem 7. The interpolating functions $\{\hat{p}_M \in \mathscr{P}([0,1])\}_{M\in\mathbb{N}}$ are, for $x \in [\frac{1}{2(M+1)}, 1 - \frac{1}{2(M+1)})$,

$$\hat{p}_M(x) = (M+1)((M+1)x + \frac{1}{2} - \lfloor (M+1)x + \frac{1}{2} \rfloor) e^T_{\lfloor (M+1)x + \frac{1}{2} \rfloor} p_M + \\ + (M+1)(\frac{1}{2} - (M+1)x + \lfloor (M+1)x + \frac{1}{2} \rfloor) e^T_{\lfloor (M+1)x + \frac{1}{2} \rfloor - 1} p_M,$$

and, for $x \in [0, \frac{1}{2(M+1)})$,

$$\hat{p}_M(x) = (M+1)e_0^T q_M,$$

and. for $x \in [1 - \frac{1}{2(M+1)}, 1]$,

$$\hat{p}_M(x) = (M+1)e_M^T q_M.$$

Lemma 12. *Given a function $p \in \mathscr{P}([0,1])$, define the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M\in\mathbb{N}}$,*

$$e_i^T p_M = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} p(x)dx,$$

where X^M is the state space described in section 3.3. If the function p is strictly greater than zero for all $x \in [0,1]$, differentiable, and its derivative is Lipschitz continuous, then the sequence $\{p_M \in \mathscr{P}(X^M)\}_{N\in\mathbb{N}}$ satisfies the convergence condition, *and satisfies, for some constant $K > 0$, all M, and all $i \in X^N \setminus \{0, M\}$,*

$$M^2 |\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_M) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_M) - 2\ln(e_i^T q_M)| \leq K,$$

and

$$M|\ln(\frac{1}{2}(e_1^T + e_0^T)q_M) - \ln(e_0^T q_M))| < K$$

and

$$M|\ln(\frac{1}{2}(e_M^T + e_{M-1}^T)q_M) - \ln(e_M^T q_M))| < K.$$

Proof. See the technical appendix, D.4.3. □

Lemma 13. *Let $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$ be a sequence of probability distributions over the state spaces associated with Theorem 7. If the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$ satisfies the convergence condition (Definition 4), then there exists a sub-sequence, whose elements we denote by n, such that:*

i) The interpolating functions (5) $\hat{p}_n(x)$ converge point-wise to a differentiable function $p(x) \in \mathscr{P}([0,1])$, whose derivative is Lipschitz-continuous, with $p(x) > 0$ for all $x \in [0,1]$,

ii) the following sum converges:

$$\lim_{n\to\infty} n^2 \sum_{i \in X^n \setminus \{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_n)\} = \frac{1}{4}\int_0^1 \frac{(p'(x))^2}{p(x)} dx,$$

where $g(x) = x\ln(x)$,

iii) for all $a \in A$, $\lim_{n\to\infty} u_{a,n}^T p_n = \int_0^1 u_a(x)p(x)dx$,

iv) and, if the sequence $\{p_M \in \mathscr{P}(X^M)\}_{M \in \mathbb{N}}$ is constructed from some function $\tilde{p}(x)$, as in Lemma 12, then $p(x) = \tilde{p}(x)$ for all $x \in [0,1]$.

Proof. See the technical appendix, section D.4.4. □

Lemma 14. *Let $\pi_M(a) \in \mathscr{P}(A)$ and $\{q_{a,M} \in \mathscr{P}(X^M)\}_{a \in A}$ denote optimal policies in the discrete state setting described in section 3.3. For each $a \in A$, the sequence $\{q_{a,N}\}$ satisfies the convergence condition (Definition 4).*

Proof. See the technical appendix, section D.4.5. □

D.4.1 Proof of Theorem 7

By the boundedness of $\mathscr{P}(A)$, there exists a convergent sub-sequence of the optimal policy $\pi_n(a)$, which we also denote by n. Define

$$\pi(a) = \lim_{n\to\infty} \pi_n(a).$$

By Lemma 14, for all $a \in A$, each sequence of optimal policies $\{q_{a,n}\}$ satisfies the convergence condition (Definition 4). Therefore, by Lemma 13, each sequence of interpolating functions (5), $\{\hat{q}_{a,n}(x)\}$, has a convergent sub-sequence that converges to a differentiable function $q_a(x)$, whose derivative is Lipschitz continuous. We can construct a sub-sequence in which $\pi_n(a)$ and all $\{\hat{q}_{a,n}(x)\}$ converge by iteratively applying this argument. Pass to this subsequence.

We can write the discrete value function, defining $g(x) = x\ln x$, as

$$\begin{aligned} V_N(q_n;n) = &\max_{\{p_{x,n}\in\mathscr{P}(A)\}_{i\in X}} \sum_{a\in A} e_a^T p_n Diag(q) u_n e_a \\ &- \theta n^2 \sum_{a\in A} (e_a^T p_n q_n) \sum_{i=0}^{n-1} [g(\frac{e_i^T q_{a,n}}{\bar{q}_{i,a,n}}) + g(\frac{e_{i+1}^T q_{a,n}}{\bar{q}_{i,a,n}})] \\ &+ \theta n^2 \sum_{i=0}^{n-1} [g(\frac{e_i^T q_N}{\bar{q}_{i,a,N}}) + g(\frac{e_{i+1}^T q_N}{\bar{q}_{i,a,N}})] \\ &- \theta n^{-1} \sum_{i=0}^{n-1} (e_i^T q_n) D_{KL}(p_n e_i || p_n q_n). \end{aligned}$$

We can re-arrange this to

$$\begin{aligned} V_N(q_n;n) = &\max_{\{p_{x,n}\in\mathscr{P}(A)\}_{i\in X}} \sum_{a\in A} e_a^T p_n Diag(q) u_n e_a \\ &- \theta n^2 \sum_{a\in A} (e_a^T pq) \sum_{i=0}^{n-1} [g(e_i^T q_{a,n}) + g(e_{i+1}^T q_{a,n}) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) q_{a,n})] \\ &+ \theta n^2 \sum_{i=0}^{N-1} [g(e_i^T q_n) + g(e_{i+1}^T q_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) q_n)] \\ &- \theta n^{-1} \sum_{i=0}^{N-1} (e_i^T q_N) D_{KL}(p_{i,n} || p_n q_n). \end{aligned}$$

By Lemma 13 and the boundedness of the KL divergence,

$$\lim_{n\to\infty} V_N(q_n;n) = \sum_{a\in A}\pi(a)\int_0^1 u_a(x)q_a(x)dx$$
$$-\frac{\theta}{4}\sum_{a\in A}\{\pi(a)\int_0^1 \frac{(q_a'(x))^2}{q_a(x)}dx\} + \frac{\theta}{4}\int_0^1 \frac{(q'(x))^2}{q(x)}dx.$$

Suppose that $\pi(a)$ and the $q_a(x)$ functions do not maximize this expression (subject to the constraints stated in Theorem 7). Let $\pi^*(a)$ and $q_a^*(x)$ be some superior policy. Define, for all n,

$$\tilde{\pi}_n(a) = \pi^*(a),$$
$$e_i^T \tilde{q}_{a,n} = \int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} q_a^*(x)dx.$$

Note that, by construction, $\tilde{q}_{a,n} \in \mathscr{P}(X^n)$ and $\sum_{a\in A}\tilde{\pi}_N(a)\tilde{q}_{a,n} = q_n$. That is, the constraints of the discrete-state problem are satisfied for all n. Denote the value function under these policies as $\tilde{V}_N(q_n;n)$.

Because of the constraints stated in Theorem 7, each q_a^* satisfies the conditions of Lemma 12, and therefore the sequence $\tilde{q}_{a,n}$ satisfies the convergence condition for all $a \in A$. It follows by Lemma 13 that this sequence of policies delivers, in the limit, the value function $V_N(q)$. If this function is strictly larger than $\lim_{n\to\infty} V_N(q_n;n)$, there must exist some $\bar{n}$ such that

$$\tilde{V}_N(q_{\bar{n}};\bar{n}) > V_N(q_{\bar{n}};\bar{n}),$$

contradicting optimality. Therefore, the functions $q_a(x)$ and $\pi(a)$ are maximizers.

It remains to show that

$$\lim_{n\to\infty}\sum_{i=0}^{\lfloor xn\rfloor} e_i^T q_{a,n} = \int_0^x q_a(y)dy.$$

Note that

$$e_i^T q_{a,n} = (n+1)\int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} \hat{q}_{a,n}(\frac{2i+1}{2(n+1)})dy,$$

where $\hat{q}_{a,n}$ is the function defined in Lemma 13. Therefore, the sum is equal to

$$\sum_{i=0}^{\lfloor xn \rfloor} e_i^T q_{a,n} = \int_0^{\frac{\lfloor xn \rfloor+1}{n+1}} \hat{q}_{a,n}(\frac{\lfloor (n+1)y + \frac{1}{2} \rfloor + \frac{1}{2}}{(n+1)})dy.$$

By the boundedness of $\hat{q}_{a,n}$ (which follows from the convergence condition) and the dominated convergence theorem,

$$\lim_{n\to\infty} \int_0^{\frac{\lfloor xn \rfloor+1}{n+1}} \hat{q}(\frac{\lfloor (n+1)y + \frac{1}{2} \rfloor + \frac{1}{2}}{(n+1)})dy = \int_0^x q_a(y)dy,$$

as required.

D.4.2 Proof of Lemma 10

We begin by observing that any information structure $p \in \mathscr{P}_{LipG}(A)$ defines unconditional action frequencies $\pi \in \mathscr{P}(A)$ and posteriors $q_a \in \mathscr{P}_{LipG}([0,1])$ satisfying (37), using definitions (38). And conversely, any unconditional action frequencies and posteriors satisfying (37) define an information structure, using definitions (39). Hence the set of candidate structures is the same in both problems, and the problems are equivalent if the two objective functions are equivalent as well. It is also easily seen that in each problem, the first term of the objective function is the expected value of the DM's reward $u(x,a)$, integrating over the joint distribution for (x,a). Hence it remains only to establish that the remaining terms of the objective function are equivalent as well.

Consider any information structure $p \in \mathscr{P}_{LipG}(A)$ and the corresponding unconditional action frequencies and posteriors, and let x be any point at which $q(x) > 0$, and at which $p_a(x)$ is twice differentiable for all a (and as a consequence, $q_a(x)$ is twice differentiable for all a as well). (We note that, given the Lipschitz continuity of the first derivatives, the set of x for which this is true must be of full measure.) Then the fact that $\sum_{a\in A} p_a(x) = 1$ for all x implies that

$$\sum_{a\in A} p''_a(x) = 0, \tag{41}$$

and similarly, constraint (37) implies that

$$\sum_{a \in A} \pi(a) q_a''(x) = q''(x). \tag{42}$$

At any such point, the definition of the Fisher information implies that

$$\begin{aligned}
I^{Fisher}(x) &\equiv \sum_{a \in A} \frac{(p_a'(x))^2}{p_a(x)} \\
&= \sum_a p_a''(x) - \sum_{a \in A} p_a(x) \frac{\partial^2 \log p_a(x)}{\partial x^2} \\
&= -\frac{\pi(a) q_a(x)}{q(x)} \frac{\partial^2}{\partial x^2} [\log \pi(a) + \log q_a(x) - \log q(x)] \\
&= \frac{1}{q(x)} \left[\sum_{a \in A} \pi(a) \frac{(q_a'(x))^2}{q_a(x)} - \sum_{a \in A} \pi(a) q_a''(x) - \frac{(q'(x))^2}{q(x)} + q''(x) \right] \\
&= \frac{1}{q(x)} \left[\sum_{a \in A} \pi(a) \frac{(q_a'(x))^2}{q_a(x)} - \frac{(q'(x))^2}{q(x)} \right].
\end{aligned}$$

Here the first line is the definition of the Fisher information (given in the lemma), and the second line follows from twice differentiating the function $\log p_a(x)$ with respect to x. In the third line, the first term from the second line vanishes because of (41); the remaining term from the second line is rewritten using (39). The fourth line follows from the third line by twice differentiating each of the terms inside the square brackets with respect to x. The fifth line then follows from (42).

Since this result holds for a set of x of full measure, we obtain expression

$$\int_0^1 q(x) I^{Fisher}(x) dx = \sum_{a \in A} \pi(a) \int_0^1 \frac{(q_a'(x))^2}{q_a(x)} dx - \int_0^1 \frac{(q'(x))^2}{q(x)} dx$$

for the mean Fisher information. This shows that the information-cost terms in both objective functions are equivalent, and hence the two problems are equivalent, and have equivalent solutions.

D.4.3 Proof of Lemma 12

Proof. The function p is strictly greater than zero, and continuous, and therefore attains a maximum and minimum on $[0,1]$, which we denote with c_H and c_L, respectively. By construction,

$$e_i^T p_M \geq \frac{c_L}{M+1}$$

and likewise for c_H, satisfying the bounds.

For all $i \in X^M \setminus \{M\}$,

$$\begin{aligned}(e_{i+1}^T - e_i^T)p_M &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} (p(x + \frac{1}{M+1}) - p(x))dx \\ &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \int_0^{\frac{1}{M+1}} p'(x+y)dydx\end{aligned}$$

and therefore, letting K_2 be the maximum of the absolute value of p' on $[0,1]$ (which exists by the continuity of p'), we have

$$|(e_{i+1}^T - e_i^T)p_M| \leq \frac{1}{(M+1)^2}K_2, \tag{43}$$

satisfying the convergence condition for the endpoints.

For all $i \in X^M \setminus \{0, M\}$,

$$\begin{aligned}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} (p(x + \frac{1}{M+1}) + p(x - \frac{1}{M+1}) - 2p(x))dx \\ &= \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \int_0^{\frac{1}{M+1}} (p'(x+y) - p'(x-y))dydx.\end{aligned}$$

Let K_3 denote the Lipschitz constant associated with p'. It follows that

$$|(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M| \leq \frac{2K_3}{(M+1)^3}.$$

Therefore, the convergence condition is satisfied for $K_1 = \max(\frac{1}{2}K_2, K_3)$.

By the concavity of the log function, and the inequality $\ln(x) \leq x-1$,

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \leq 2\ln(\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1} + 2e_i^T)p_M}{e_i^T p_M})$$
$$\leq \frac{\frac{1}{2}(e_{i+1}^T + e_{i-1} - 2e_i^T)p_M}{e_i^T p_M}.$$

Therefore, by the convergence condition we have established,

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \leq \frac{(M+1)K_1}{M^3 c_L} \leq \frac{2K_1}{M^2 c_L}.$$

By the inequality $-\ln(\frac{1}{x}) \leq x-1$,

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \geq \frac{\frac{1}{2}(e_{i+1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}.$$

We can rewrite this as

$$\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}) \geq$$
$$(\frac{\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M} - 1)).$$

By the bounds above,

$$\frac{\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M} \geq -\frac{2K_1}{M^2 c_L}$$

and, using equation (43),

$$\frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}\left(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M} - 1\right) = \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)p_M}{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}\left(\frac{\frac{1}{2}(e_{i+1}^T - e_{i-1}^T)p_M}{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}\right)$$
$$\geq -\frac{M^2}{c_L^2}\frac{1}{(M+1)^4}(K_2)^2$$
$$\geq -\left(\frac{K_2}{2Mc_L}\right)^2.$$

Therefore,

$$M^2\left|\ln\left(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)p_M}{e_i^T p_M}\right) + \ln\left(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)p_M}{e_i^T p_M}\right)\right| \leq \frac{2K_1}{c_L} + \left(\frac{K_2}{2c_L}\right)^2.$$

For the end-points,

$$\frac{\frac{1}{2}(e_1^T - e_0^T)q_M}{\frac{1}{2}(e_1^T + e_0^T)q_M} \leq \ln\left(\frac{\frac{1}{2}(e_1^T + e_0^T)q_M}{e_0^T q_M}\right) \leq \frac{\frac{1}{2}(e_1^T - e_0^T)q_M}{e_0^T q_M}$$

and therefore

$$\left|\ln\left(\frac{\frac{1}{2}(e_1^T + e_0^T)q_M}{e_0^T q_M}\right)\right| \leq \frac{K_2}{Mc_L}.$$

A similar property holds for the other endpoint, and therefore the claim holds for $K = \max(\frac{K_2}{c_L}, \frac{2K_1}{c_L} + (\frac{K_2}{2c_L})^2)$. □

D.4.4 Proof of Lemma 13

Proof. We begin by noting that the functions $\hat{p}_M(x)$ are absolutely continuous. Almost everywhere in $[\frac{1}{2(M+1)}, 1 - \frac{1}{2(M+1)}]$,

$$\hat{p}_M'(x) = (M+1)^2(e^T_{\lfloor (M+1)x + \frac{1}{2}\rfloor} - e^T_{\lfloor (M+1)x + \frac{1}{2}\rfloor - 1})p_M,$$

and outside this region, $\hat{p}_M'(x) = 0$. Let $\tilde{p}_M'(x)$ denote the right-continuous Lebesgue-integrable function on $[0,1]$ such that

$$\hat{p}_M(x) = \hat{p}_M(0) + \int_0^x \tilde{p}_M'(y)dy,$$

which is equal to $\hat{p}'_M(x)$ anywhere the latter exists.

The total variation of $\tilde{p}'_M(x)$ is equal to

$$TV(\tilde{p}'_M) = \sum_{i=1}^{M-1}(M+1)^2|(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M)|+ \\ + (M+1)^2|(e_M^T - e_{M-1}^T)p_M| + (M+1)^2|(e_1^T - e_0^T)p_M|.$$

By the convergence condition,

$$TV(\tilde{p}'_M) \leq \frac{(M+1)^3}{M^3}2K_1,$$

and therefore the sequence of functions $\tilde{p}'_M(x)$ has uniformly bounded variation.

For any $1-\frac{1}{2(M+1)} > x > y \geq \frac{1}{2(M+1)}$, the quantity

$$|\tilde{p}'_M(x) - \tilde{p}'_M(y)| = (M+1)^2|\sum_{i=\lfloor (M+1)y+\frac{1}{2}\rfloor}^{\lfloor (M+1)x+\frac{1}{2}\rfloor}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)p_M| \\ \leq \frac{(M+1)^2((M+1)(x-y)+2)}{M^3}2K_1.$$

At the end points, for all $x \in [0, \frac{1}{2(M+1)})$,

$$|\tilde{p}'_M(\frac{1}{2(M+1)}) - \tilde{p}'_M(x)| \leq \frac{2K_1}{M+1},$$

and for all $x \in [1-\frac{1}{2(M+1)}, 1]$,

$$|\tilde{p}'_M(x) - \lim_{y\uparrow 1-\frac{1}{2(M+1)}}\tilde{p}'_M(y)| \leq \frac{2K_1}{M+1}.$$

By $\tilde{p}'_M(0) = 0$, we have, for all $x \in [0,1]$,

$$|\tilde{p}'_M(x)| \leq (\frac{(M+1)^2((M+1)(1-\frac{1}{2(M+1)})+2)}{M^3} + \frac{1}{M+1})2K_1,$$

proving that $\tilde{p}'_M(x)$ is bounded uniformly in M for all $x \in [0,1]$.

Therefore Helly's selection theorem applies. That is, there exists a sub-sequence, which we denote by n, such that $\tilde{p}_n'(x)$ converges point-wise to some $p'(x)$. Moreover, by the point-wise convergence of $\tilde{p}_M'$ to p', for all $x > y$,

$$|p'(x) - p'(y)| \leq 2K_1(x-y),$$

meaning that p' is Lipschitz-continuous. By the fact that $p'(0) = 0$, this implies that $|p'(x)| \leq 2K_1$ for all $x \in [0,1]$.

By the convergence condition, $c_L \leq \hat{p}_N(0) \leq c_H$. Therefore, there exists a convergent sub-sequence. We now use n to denote the sub-sequence for which $\lim_{n\to\infty} \hat{p}_n(0) = p(0)$ and for which $\tilde{p}_n'(x)$ converges point-wise to $p'(x)$. By the dominated convergence theorem, for all $x \in [0,1]$,

$$\lim_{n\to\infty} \hat{p}_n(x) = \lim_{n\to\infty} \{\hat{p}_n(0) + \int_0^x \tilde{p}_n'(y)dy\} = p(0) + \int_0^x p'(y)dy.$$

Define the function $p(x) = p(0) + \int_0^x p'(y)dy$ for all $x \in [0,1]$. By the convergence conditions, this function is bounded, $0 < c_L \leq p(x) \leq c_H$, by construction it is differentiable, and its derivative is Lipschitz continuous. Moreover,

$$\int_0^1 p(x)dx = 1,$$

and therefore $p \in \mathscr{P}([0,1])$.

Next, consider the limiting cost function. We have, using the function $g(x) = x\ln x$ and Taylor-expanding,

$$g(y) = g(x) + g'(x)(y-x) + \frac{1}{2}g''(cy+(1-c)x)(y-x)^2$$

for some $c \in (0,1)$. Therefore,

$$
\begin{aligned}
g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_M) = \\
\frac{1}{8}g''(c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T)p_M)((e_{i+1}^T - e_i^T)p_M)^2 \\
+ \frac{1}{8}g''(c_2 e_i^T p_M + (1-c_2)\frac{1}{2}(e_i^T + e_{i+1}^T)p_M)((e_{i+1}^T - e_i^T)p_M)^2
\end{aligned}
$$

for constants $c_1, c_2 \in (0,1)$. Note that, by the boundedness $\hat{p}_M(x)$ from below, $e_i^T p_M \geq (M+1)^{-1} c_L$ for all $i \in X^M$. It follows that

$$
g''(c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T)p_M) = \frac{1}{c_1 e_i^T p_M + (1-c_1)\frac{1}{2}(e_i^T + e_{i+1}^T)p_M} \leq (M+1)c_L^{-1}.
$$

Therefore,

$$
0 \leq g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_M) \leq \frac{(M+1)c_L^{-1}}{4}((e_{i+1}^T - e_i^T)p_M)^2.
$$

By construction,

$$
e_i^T p_M = \frac{1}{(M+1)}\hat{p}_M(\frac{2i+1}{2(M+1)}).
$$

Therefore,

$$
\begin{aligned}
(M+1)(g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_M)) = \\
g(\hat{p}_M(\frac{2i+1}{2(M+1)})) + g(\hat{p}_M(\frac{2i+3}{2(M+1)})) - 2g(\hat{p}_M(\frac{2i+2}{2(M+1)})).
\end{aligned}
$$

and

$$
g(e_i^T p_M) + g(e_{i+1}^T p_M) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T)p_M) \leq \frac{c_L^{-1}}{4(M+1)}(\hat{p}(\frac{2i+3}{2(M+1)}) - \hat{p}(\frac{2i+1}{2(M+1)}))^2.
$$

By the boundedness of $\tilde{p}'_M(x)$,

$$
g(\hat{p}(\frac{2i+1}{2(M+1)})) + g(\hat{p}(\frac{2i+3}{2(M+1)})) - 2g(\hat{p}(\frac{2i+2}{2(M+1)})) \leq \frac{B}{(M+1)^2}
$$

for some finite bound B.

Writing the limiting cost as an integral, and switching to the sub-sequence n defined above,

$$n^2 \sum_{i \in X^n \setminus \{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_n)\} =$$

$$\frac{n^3}{n+1} \int_0^1 \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx.$$

By the bound above,

$$\frac{n^3}{n+1} \int_0^1 \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx \leq$$

$$\frac{n^3}{(n+1)^3} \int_0^1 B dx.$$

Applying the dominated convergence theorem,

$$\lim_{n \to \infty} n^2 \sum_{i \in X^n \setminus \{n\}} \{g(e_i^T p_n) + g(e_{i+1}^T p_n) - 2g(\frac{1}{2}(e_i^T + e_{i+1}^T) p_n)\} =$$

$$\int_0^1 \lim_{n \to \infty} \frac{n^3}{n+1} \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} dx.$$

By the Taylor expansion above,

$$\lim_{n \to \infty} \frac{n^3}{n+1} \{g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) + g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)})) - 2g(\hat{p}_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)}))\} =$$

$$\lim_{n \to \infty} \frac{1}{8} \frac{n^3}{n+1} \{g''(\cdot) + g''(\cdot)\} (\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)}) - \hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)}))^2.$$

By definition,

$$(n+1)(\hat{p}_n(\frac{2\lfloor nx \rfloor + 3}{2(n+1)}) - \hat{p}_n(\frac{2\lfloor nx \rfloor + 1}{2(n+1)})) = \tilde{p}'_n(\frac{2\lfloor nx \rfloor + 2}{2(n+1)})$$

and

$$\lim_{n\to\infty} g''(\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)})+c_n(\hat{p}_n(\frac{2\lfloor nx\rfloor+3}{2(n+1)})-\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)}))) = \frac{1}{p(x)},$$

with $c_n \in (0,1)$ for all n, and therefore

$$\lim_{n\to\infty} \frac{n^3}{n+1}\{g(\hat{p}_n(\frac{2\lfloor nx\rfloor+1}{2(n+1)}))+g(\hat{p}_n(\frac{2\lfloor nx\rfloor+3}{2(n+1)}))-2g(\hat{p}_n(\frac{2\lfloor nx\rfloor+2}{2(n+1)}))\} = \lim_{n\to\infty} \frac{1}{4}\frac{(p'(x))^2}{p(x)},$$

proving the second claim.

Turning to the third claim, recall that, by definition,

$$e_i^T u_{a,M} = \frac{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} u_a(x)q(x)dx}{\int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} q(x)dx.}$$

We define the function, for $x \in [0,1)$, as

$$u_{a,M}(x) = e^T_{\lfloor (M+1)x\rfloor} u_{a,M},$$

and let $u_{a,M}(1) = e_M^T u_{a,M}$. We also define the function

$$\tilde{x}_M(x) = \frac{2\lfloor (M+1)x\rfloor+1}{2(M+1)}.$$

By construction, $\hat{p}_M(\tilde{x}_M(x)) = (M+1)e^T_{\lfloor (M+1)x\rfloor} p_{a,M}$ for all $x \in [0,1)$, and equals $e_M^T p_{a,M}$ for $x = 1$. Therefore,

$$\begin{aligned} u_{a,M}^T p_M &= \sum_{i\in X^M} (e_i^T u_{a,M})(e_i^T p_M) \\ &= \int_0^1 \hat{p}_M(\tilde{x}_M(x)) u_{a,M}(x)dx. \end{aligned}$$

By the measurability of $u_a(x)$,

$$\lim_{M\to\infty} u_{a,M}(x) = u_a(x).$$

Therefore, by the boundedness of utilities and the dominated convergence theorem,

$$\lim_{n\to\infty} u_{a,n}^T p_n = \int_0^1 p(x)u_a(x)dx.$$

Finally, suppose that, for all M

$$e_i^T p_{a,M} = \int_{\frac{i}{M+1}}^{\frac{i+1}{M+1}} \tilde{p}(x)dx.$$

It follows that $\lim_{n\to\infty} \hat{p}_{a,n}(x) = \tilde{p}(x)$ for all $x \in [0,1]$, and therefore $\tilde{p}(x) = p(x)$. $\square$

D.4.5 Proof of Lemma 14

Proof. We begin by noting that the conditions given for the function $q(x)$ satisfy the conditions of Lemma 12, and therefore the sequence q_M satisfies the convergence condition. We will use the constants c_H and c_L to refer to its bounds,

$$\frac{c_H}{M+1} \geq e_i^T q_M \geq \frac{c_L}{M+1},$$

and the constants K_1 and K to refer to the constants described by convergence condition and Lemma 12 for the sequence q_M. By the convention that $q_{a,M} = q_M$ if $\pi_M(a) = 0$, $q_{a,M}$ also satisfies the convergence condition whenever $\pi_M(a) = 0$.

The problem of size M is

$$\begin{aligned} V_N(q_M;M) = &\max_{\pi_M \in \mathscr{P}(A), \{q_{a,M} \in \mathscr{P}(X^M)\}_{a\in A}} \sum_{a\in A} \pi_M(a)(u_{a,M}^T \cdot q_{a,M}) \\ &- \theta \sum_{a\in A} \pi_M(a) D_N(q_{a,M}||q_M;M) \end{aligned}$$

subject to

$$\sum_{a\in A}\pi_M(a)q_{a,M}=q_M,$$

where

$$\begin{aligned}D_N(q_{a,M}||q_M;\rho,M)=M^2(H_N(q_{a,M};1,M)-H_N(q_M;1,M))\\+M^{-1}(H^S(q_{a,M};M)-H^S(q_M;M)\end{aligned}$$

and

$$H_N(q;1,M)=-\sum_{i=0}^{M-1}\bar{q}_iH^S(q_i).$$

Let u_M denote that $|X^M|\times|A|$ matrix whose columns are $u_{a,M}$. Using Lemma 1, we can rewrite the problem as

$$\begin{aligned}V_N(q_M;M)=&\max_{\{p_{i,M}\in\mathscr{P}(A)\}_{i\in X^M}}\sum_{a\in A}e_a^Tp_MDiag(q)u_Me_a\\&-\theta M^2\sum_{i=0}^{M-1}(e_i^Tq_M)D_{KL}(p_{i,M}||\frac{p_{i,M}(e_i^Tq_M)+p_{i+1,M}(e_{i+1}^Tq_M)}{(e_i^T+e_{i+1}^T)q_M})\\&-\theta M^2\sum_{i=1}^{M}(e_i^Tq_M)D_{KL}(p_{i,M}||\frac{p_{i,M}(e_i^Tq_N)+p_{i-1,M}(e_{i-1}^Tq_M)}{(e_i^T+e_{i-1}^T)q_M})\\&-\theta M^{-1}\sum_{i=0}^{M-1}(e_i^Tq_M)D_{KL}(p_{i,M}||p_Mq_M).\end{aligned}$$

The FOC for this problem is, for all $i\in[1,M-1]$ and $a\in A$ such that $e_a^Tp_{i,M}>0$,

$$\begin{aligned}e_i^Tu_{a,M}-\theta M^2\ln(\frac{e_a^Tp_{i,M}(e_i^T+e_{i+1}^T)q_M}{e_a^T(p_{i,M}(e_i^Tq_M)+p_{i+1,M}(e_{i+1}^Tq_M))})\\-\theta M^2\ln(\frac{e_a^Tp_{i,M}(e_i^T+e_{i-1}^T)q_M}{e_a^T(p_{i,M}(e_i^Tq_M)+p_{i-1,N}(e_{i-1}^Tq_M))})-\theta M^{-1}\ln(\frac{e_a^Tp_{i,M}}{e_a^Tp_Mq_M})-e_i^T\kappa_M=0,\end{aligned}$$

where $\kappa_M\in\mathbb{R}^{M+1}$ are the multipliers (scaled by $e_i^Tq_M$) on the constraints that $\sum_{a\in A}e_a^Tp_{i,M}=1$ for all $i\in X$. Defining $e_{i-1}^Tq_M=e_{M+1}^Tq_M=0$, and defining $p_{-1,M}$ and $p_{M+1,M}$ in arbitrary fashion, we can recover this FOC for all $i\in X$.

Rewriting the FOC in terms of the posteriors, and again defining $e_{i-1}^Tq_{a,M}=$

$e_{M+1}^T q_{a,M} = 0$, for any a such that $\pi_M(a) > 0$,

$$
\begin{aligned}
e_i^T(u_{a,M} - \kappa_M) &= \theta M^2 \ln\left(\frac{(e_i^T q_{a,M})(1+\frac{e_{i+1}^T q_M}{e_i^T q_M})}{(e_{i+1}+e_i)^T q_{a,M}}\right) + \theta M^2 \ln\left(\frac{(e_i^T q_{a,N})(1+\frac{e_{i-1}^T q_N}{e_i^T q_N})}{(e_{i-1}+e_i)^T q_{a,N}}\right) \\
&\quad + \theta M^{-1} \ln\left(\frac{e_a^T p_{i,M}}{e_a^T p_M q_M}\right) \\
&= -\theta M^2 \ln\left(1+\frac{e_{i+1}^T q_{a,M}}{e_i^T q_{a,M}}\right) + \theta M^2 \ln\left(1+\frac{e_{i+1}^T q_M}{e_i^T q_M}\right) \\
&\quad - \theta M^2 \ln\left(1+\frac{e_{i-1}^T q_{a,M}}{e_i^T q_{a,M}}\right) \\
&\quad + \theta M^2 \ln\left(1+\frac{e_{i-1}^T q_M}{e_i^T q_M}\right) + \theta M^{-1} \ln\left(\frac{e_i^T q_{a,M}}{e_i^T q_M}\right),
\end{aligned}
$$

which can be rewritten as

$$
\begin{aligned}
& e_i^T(u_{a,M} - \kappa_M) = \\
& -\theta M^2\left(\ln\left(\frac{1}{2}(e_{i+1}^T + e_i^T) q_{a,M}\right) + \ln\left(\frac{1}{2}(e_{i-1}^T + e_i^T) q_{a,M}\right) - (2+M^{-3})\ln(e_i^T q_{a,M})\right) \\
& +\theta M^2\left(\ln\left(\frac{1}{2}(e_{i+1}^T + e_i^T) q_M\right) + \ln\left(\frac{1}{2}(e_{i-1}^T + e_i^T) q_M\right) - (2+M^{-3})\ln(e_i^T q_M)\right). \quad (44)
\end{aligned}
$$

Our analysis proceeds by analyzing this first-order condition.

We next describe a series of lemmas that use this first-order condition to establish various bounds, which will ultimately be used to establish the bounds required by the convergence condition. As part of the proof, we find it useful to consider the interpolating functions $\hat{q}_{a,M}(x)$ (5) constructed from $q_{a,M}$. We define from these interpolating functions the function

$$l_{a,N}(x) = (M+1)\left(\ln(\hat{q}_{a,M}(x)) - \ln\left(\hat{q}_{a,M}\left(x - \frac{1}{2(M+1)}\right)\right)\right)$$

on $x \in [\frac{1}{2(M+1)}, 1]$, observing that, for any $i \in X^M \setminus \{0\}$,

$$l_{a,M}\left(\frac{2i+1}{2(M+1)}\right) = (M+1)\ln\left(\frac{(M+1)e_i^T q_{a,M}}{\frac{1}{2}(M+1)(e_i^T + e_{i-1}^T) q_{a,M}}\right),$$

and for any $i \in X^M \setminus \{M\}$,

$$l_{a,M}(\frac{2i+2}{2(M+1)}) = (M+1)\ln(\frac{\frac{1}{2}(M+1)(e_i^T + e_{i+1}^T)q_{a,M}}{(M+1)e_i^T q_{a,M}}).$$

$\square$

Lemma 15. *For all $M \in \mathbb{N}$ and $i \in X^M \setminus \{0, M\}$, $e_i^T \kappa_M \leq B_\kappa$ for some positive constant B_κ.*

Proof. See the technical appendix, section D.4.6. $\square$

Lemma 16. *For all $M \in \mathbb{N}$ and $i \in \{0, M\}$, $|e_i^T \kappa_M| \leq B_0$ for some positive constant B_0, and*

$$\ln(\frac{\frac{1}{2}(e_0^T + e_1^T)q_{a,M}}{e_0^T q_{a,M}}) \leq M^{-1}B_1$$

and

$$\ln(\frac{e_M^T q_{a,M}}{\frac{1}{2}(e_M^T + e_{M-1}^T)q_{a,M}}) \geq -M^{-1}B_1$$

for some positive constant B_1.

Proof. See the technical appendix, section D.4.7. $\square$

Lemma 17. *For all $M \in \mathbb{N}$ and $j \in \{2, 3, \ldots, 2M+1\}$, and some positive constant B_l,*

$$|l_{a,N}(\frac{j}{2(M+1)})| \leq B_l.$$

Proof. See the technical appendix, section D.4.8. The proof uses the previous two lemmas. $\square$

Armed with these lemmas, we prove that the convergence condition (Definition 4) is satisfied.

Proof that $\frac{c_H}{M+1} \geq e_i^T q_{a,M} \geq \frac{c_L}{M+1}$ We next apply the above lemmas to prove that the first part of the convergence condition is satisfied. Begin by observing that there must exist some $\tilde{i}_{a,M} \in X^M$ such that $e_{\tilde{i}_{a,M}}^T q_{a,M} \geq \frac{1}{N+1}$, implying that

$$\ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) \geq 0.$$

By the definition of $l_{a,M}$, for any $i \in X^M \setminus \{0\}$,

$$l_{a,M}(\frac{2i+1}{2(M+1)}) + l_{a,M}(\frac{2i}{2(M+1)}) = (M+1)\ln(\frac{(M+1)e_i^T q_{a,M}}{(M+1)e_{i-1}^T q_{a,M}}).$$

For any $i > \tilde{i}_{a,M}$, using Lemma 17,

$$\begin{aligned}
\ln((M+1)e_i^T q_{a,M}) &= \ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) + \sum_{j=\tilde{i}_{a,M}+1}^{i} \ln(\frac{(M+1)e_j^T q_{a,M}}{(M+1)e_{j-1}^T q_{a,M}}) \\
&= \ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) + \frac{1}{M+1} \sum_{j=\tilde{i}_{a,M}+1}^{i} l_{a,M}(\frac{2j+1}{2(M+1)}) \\
&+ l_{a,N}(\frac{2j}{2(M+1)}) \\
&\geq -\frac{1}{M+1} \sum_{j=\tilde{i}_{a,M}+1}^{i} 2B_l \\
&\geq -2B_l.
\end{aligned}$$

Similarly, for any $i < \tilde{i}_{a,M}$,

$$\ln((M+1)e_{\tilde{i}_{a,M}}^T q_{a,M}) = \ln((M+1)e_i^T q_{a,M}) + \sum_{j=i+1}^{\tilde{i}_{a,M}} \ln(\frac{(N+1)e_j^T q_{a,N}}{(N+1)e_{j-1}^T q_{a,N}}).$$

Therefore, for any $i < \tilde{i}_{a,M}$,

$$\ln((M+1)e_i^T q_{a,M}) \geq -\sum_{j=i+1}^{\tilde{i}_{a,M}} \ln(\frac{(M+1)e_j^T q_{a,M}}{(M+1)e_{j-1}^T q_{a,M}}),$$

and thus, using Lemma 17, for all $i \in X^M$,

$$\ln((M+1)e_i^T q_{a,M}) \geq -2B_l.$$

Repeating this argument, there must be some $\hat{i}_{a,M}$ such that $e^T_{\hat{i}_{a,M}} q_{a,M} \leq M^{-1}$, and using the bounds on $l_{a,M}$ in similar fashion yields

$$\ln((M+1)e_i^T q_{a,M}) \leq 2B_l.$$

It follows that, for all M, $a \in A$ such that $\pi_M(a) > 0$, and $i \in X^M$,

$$\frac{exp(2B_l)}{(M+1)} \geq e_i^T q_{a,M} \geq \frac{\exp(-2B_l)}{M+1}, \tag{45}$$

demonstrating that $q_{a,N}$ satisfies the first part of the convergence condition.

Proof that $M^3|\frac{1}{2}(e^T_{i+1} + e^T_{i-1} - 2e^T_i)q_{a,M}| \leq K_1$ We start by proving a bound on $(M+1)^2|\frac{1}{2}(e^T_{i+1} - e^T_i)q_{a,M}|$.

Using Lemma 17, and a Taylor expansion of $\ln(1+x)$, for some $c \in (0,1)$, for any $i \in X^M \setminus \{M\}$,

$$\begin{aligned}
|l_{a,M}(\frac{2i+2}{2(M+1)})| &= |(M+1)\ln(\frac{\frac{1}{2}(M+1)(e_i^T + e^T_{i+1})q_{a,M}}{(M+1)e_i^T q_{a,M}})| \\
&= \frac{(M+1)|\frac{1}{2}(e^T_{i+1} - e^T_i)q_{a,M}|}{e_i^T q_{a,M} + \frac{c}{2}(e^T_{i+1} - e^T_i)q_{a,M}} \\
&\leq B_l,
\end{aligned}$$

and therefore, by the bound on $e_i^T q_{a,M}$, for any $i \in X^M \setminus \{M\}$,

$$(M+1)^2|\frac{1}{2}(e^T_{i+1} - e^T_i)q_{a,M}| \leq B_l \exp(-2B_l). \tag{46}$$

Returning to the first-order condition, for $i \in X^N \setminus \{0,N\}$, and using the bounds on utility and on the terms involving q_M,

$$\begin{aligned}
e_i^T \kappa_M \geq& -\bar{u} - \theta K + \theta M^{-1} \ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \\
&+ \theta M^2(\ln(\frac{1}{2}(e^T_{i+1} + e^T_i)q_{a,M}) + \ln(\frac{1}{2}(e^T_{i-1} + e^T_i)q_{a,M}) - 2\ln(e_i^T q_{a,M})).
\end{aligned}$$

We have
$$M^{-1}\ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \geq M^{-1}\ln(\frac{c_L}{\exp(2B_l)}),$$
and therefore
$$\begin{aligned} e_i^T \kappa_M \geq -\bar{u} - \theta K + M^{-1}\ln(\frac{c_L}{\exp(2B_l)}) \\ + \theta M^2(\ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}}{e_i^T q_{a,M}}) + \ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}}{e_i^T q_{a,M}})). \end{aligned}$$
Using the mean-value theorem, for some $c_1 \in (0,1)$,
$$\begin{aligned} \ln(\frac{\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}}{e_i^T q_{a,M}}) &= \ln(1 + \frac{\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}}{e_i^T q_{a,M}}) \\ &= \frac{e_i^T q_{a,M}}{e_i^T q_{a,M} + c_1\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}} \frac{\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}}{e_i^T q_{a,M}}, \end{aligned}$$
and likewise
$$\ln(\frac{\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}}{e_i^T q_{a,M}}) = \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M} + \frac{1}{2}c_1 e_{i-1}^T q_{a,M}}$$
for some $c_2 \in (0,1)$. Therefore,
$$\begin{aligned} e_i^T \kappa_M \geq -\bar{u} - \theta K + M^{-1}\ln(\frac{c_L}{\exp(2B_l)}) \\ + \theta M^2(\frac{\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M}}{(1-\frac{1}{2}c_1)e_i^T q_{a,M} + \frac{1}{2}c_1 e_{i+1}^T q_{a,M}} + \frac{\frac{1}{2}(e_{i-1}^T - e_i^T)q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M} + \frac{1}{2}c_2 e_{i-1}^T q_{a,M}}). \end{aligned}$$

Multiplying through,

$$
\begin{aligned}
&[(1-\frac{1}{2}c_1)e_i^T q_{a,M} + \frac{1}{2}c_1 e_{i+1}^T q_{a,M}](e_i^T \kappa_M + \bar{u} + \theta K - M^{-1}\ln(\frac{c_L}{\exp(2B_l)}))\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T - e_i^T)q_{a,M} + \frac{1}{2}(e_{i-1}^T - e_i^T)q_{a,M}\frac{(1-\frac{1}{2}c_1)e_i^T q_{a,M} + \frac{1}{2}c_1 e_{i+1}^T q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M} + \frac{1}{2}c_2 e_{i-1}^T q_{a,M}}).\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M}\\
&+\frac{1}{2}(e_{i-1}^T - e_i^T)q_{a,M}(\frac{\frac{1}{2}c_1(e_{i+1}^T - e_i^T)q_{a,M} - \frac{1}{2}c_2(e_i^T - e_{i-1}^T)q_{a,M}}{(1-\frac{1}{2}c_2)e_i^T q_{a,M} + \frac{1}{2}c_2 e_{i-1}^T q_{a,M}})).
\end{aligned}
$$

Using equations (45) and (46),

$$
\begin{aligned}
&[(1-\frac{1}{2}c_1)e_i^T q_{a,M} + \frac{1}{2}c_1 e_{i+1}^T q_{a,M}](e_i^T \kappa_M + \bar{u} + \theta K - M^{-1}\ln(\frac{c_L}{\exp(2B_l)}))\\
&\geq \theta M^2(\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M} - \frac{B_l\exp(2B_l)}{(M+1)^2}(\frac{\frac{2B_l\exp(2B_l)}{(M+1)^2}}{\frac{\exp(-2B_l)}{M+1}}))\\
&\geq \theta M^2\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M} - \theta\frac{2B_l^2 M^2\exp(6B_l)}{(M+1)^3}.
\end{aligned}
$$

Summing over a, weighted by $\pi_N(a)$, and applying Lemma 12,

$$
\begin{aligned}
(e_i^T \kappa_M + \bar{u} + \theta K - M^{-1}\ln(\frac{c_L}{\exp(2B_l)})) &\geq -\theta\frac{\frac{K_1}{M} + \frac{2B_l^2M^2\exp(6B_l)}{(M+1)^3}}{\frac{c_L}{(M+1)}}\\
&\geq -\theta c_L^{-1}(2K_1 + 2B_l^2\exp(6B_l)).
\end{aligned}
$$

Therefore, $|e_i^T \kappa_N|$ is bounded below by some $B_\kappa^+ > 0$ for all $i \in X^N$ (recalling that this was shown for $i \in \{0,N\}$ in Lemma 16 and in the other direction in Lemma 15).

It also follows, using equation (45), that

$$
\begin{aligned}
\theta M^2(M+1)\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M} \leq exp(2B_l)(B_\kappa^+ + \bar{u} + \theta K - M^{-1}\ln(\frac{c_L}{\exp(2B_l)})\\
+\theta\frac{2B_l^2M^2\exp(6B_l)}{(M+1)^2},
\end{aligned}
$$

which establishes one side of the bound on $|\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M}|$.

Rewriting the FOC (equation (44)) and using Lemma 12 and the boundedness of the utility and the bound on $|e_i^T \kappa_N|$,

$$
\begin{aligned}
&-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1}\ln(\frac{e_i^T q_M}{e_i^T q_{a,M}})\\
\leq&\theta M^2(\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}) - 2\ln(e_i^T q_{a,M})).
\end{aligned}
$$

By equation (45),

$$M^{-1}\ln(\frac{e_i^T q_M}{e_i^T q_{a,M}}) \leq M^{-1}\ln(\frac{c_H}{\exp(-2B_l)}),$$

and therefore, by the concavity of the log function,

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1}\ln(\frac{c_H}{\exp(-2B_l)}) \leq 2\theta M^2 \ln(\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M}}{e_i^T q_{a,M}}).$$

By the inequality $\ln(x) \leq x - 1$,

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1}\ln(\frac{c_H}{\exp(-2B_l)}) \leq 2\theta M^2 (\frac{\frac{1}{4}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M}}{e_i^T q_{a,M}}),$$

and therefore, using the lower bound on $e_i^T q_{a,M}$ (equation (45)),

$$-B_\kappa^+ - \bar{u} - \theta K - \theta M^{-1}\ln(\frac{c_H}{\exp(-2B_l)}) \leq \theta M^2 (M+1)\frac{1}{2}(e_{i+1}^T + e_{i-1}^T - 2e_i^T)q_{a,M},$$

which proves the other side of the bound.

Proof that $M^2|\frac{1}{2}(e_1^T - e_0^T)q_{a,M}| \leq K_1$ By Lemma 17,

$$-B_l \leq (M+1)\ln(\frac{\frac{1}{2}(e_0^T + e_1^T)q_{a,M}}{e_0^T q_{a,M}}) \leq B_l.$$

Using the mean-value theorem, for some $c \in (0,1)$,

$$\ln(\frac{\frac{1}{2}(e_0^T + e_1^T)q_{a,M}}{e_0^T q_{a,M}}) = \frac{\frac{1}{2}(e_1^T - e_0^T)q_{a,M}}{(1-\frac{1}{2}c)e_0^T q_{a,M} + \frac{1}{2}c e_i^T q_{a,M}}.$$

Therefore, by equation (45),

$$\frac{\exp(2B_l)}{(M+1)^2}B_l \geq \frac{1}{2}(e_1^T - e_0^T)q_{a,M} \geq -\frac{\exp(2B_l)}{(M+1)^2}B_l,$$

proving the bound. The proof for the other endpoint is identical.

D.4.6 Proof of Lemma 15

First, using Lemma 12, for all $i \in X^M \setminus \{0,M\}$, observe that

$$M^2|\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_M) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_M) - 2\ln(e_i^T q_M)| \leq K.$$

Rewriting the FOC (equation (44)) and using this bound,

$$\begin{aligned} e_i^T \kappa_M \leq\ & e_i^T u_{a,M} + \theta K + \theta M^{-1}\ln(e_i^T q_M) \\ & + \theta M^2(\ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}) - (2+M^{-3})\ln(e_i^T q_{a,M})). \end{aligned}$$

By the boundedness of the utility function, this can be rewritten as

$$\begin{aligned} e_i^T \kappa_M \leq\ & \bar{u} + \theta K - \theta M^2(\ln(\frac{e_i^T q_{a,M}}{\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}}) \\ & + \ln(\frac{e_i^T q_{a,M}}{\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}})) - \theta M^{-1}\ln(\frac{e_i^T q_{a,M}}{e_i^T q_M}). \end{aligned}$$

By the concavity of the log function,

$$\begin{aligned} & \ln(\frac{1}{2}(e_{i+1}^T + e_i^T)q_{a,M}) + \ln(\frac{1}{2}(e_{i-1}^T + e_i^T)q_{a,M}) + M^{-3}\ln(e_i^T q_M) \leq \\ & \qquad (2+M^{-3})\ln(\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M} + \frac{M^{-3}}{2+M^{-3}}e_i^T q_M), \end{aligned}$$

It follows that

$$e_i^T \kappa_N \leq \bar{u} + \theta K + (2+M^{-3})\theta M^2 \ln\left(\frac{\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M} + \frac{N^{-3}}{2+N^{-3}} e_i^T q_M}{e_i^T q_{a,M}}\right).$$

Exponentiating,

$$(e_i^T q_{a,M}) \exp\left(-\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M)\right) \leq \frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_{a,M} + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M.$$

Summing over a, weighted by $\pi_N(a)$,

$$(e_i^T q_M) \exp\left(-\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M)\right) \leq \frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_M + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M.$$

Taking logs,

$$\begin{aligned}
&-\frac{1}{2+M^{-3}} \theta^{-1} M^{-2} (\bar{u} + \bar{\theta} K - e_i^T \kappa_M) \\
&\leq \ln\left(\frac{\frac{1}{2(2+M^{-3})}(e_{i+1}^T + e_{i-1}^T + 2e_i^T)q_M + \frac{M^{-3}}{2+M^{-3}} e_i^T q_M}{(e_i^T q_M)}\right) \\
&\leq \ln\left(1 + \frac{M^{-3}}{2+M^{-3}} + \frac{1}{2+M^{-3}} \frac{K_1 M^{-3}}{c_L M^{-1}}\right),
\end{aligned}$$

where the last step follows by Lemma 12, recalling that c_L is the lower bound on $q(x)$. We have

$$\begin{aligned}
e_i^T \kappa_N &\leq 3\theta M^2 \ln\left(1 + \frac{M^{-3}}{2+M^{-3}} + \frac{1}{2+M^{-3}} \frac{K_1}{c_L} M^{-2}\right) + \bar{u} + \bar{\theta} K \\
&\leq \bar{u} + \theta K + \frac{3\theta M^{-1}}{2+M^{-3}} + \frac{3\theta}{2+M^{-3}} \frac{K_1}{c_L} \\
&\leq \bar{u} + \theta K + \frac{3\theta}{2} + \frac{3\theta}{2} \frac{K_1}{c_L}.
\end{aligned}$$

where the second step follows by the inequality $\ln(1+x) < x$ for $x > 0$.

D.4.7 Proof of Lemma 16

For the lower end point, the FOC (equation (44)) can be simplified to

$$\begin{aligned}e_0^T(u_{a,M}-\kappa_M) = &-\theta M^2(\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M})+\ln(\frac{1}{2})-(1+M^{-3})\ln(e_0^T q_{a,M}))\\ &+\theta M^2(\ln(\frac{1}{2}(e_1^T+e_0^T)q_M)+\ln(\frac{1}{2})-(1+M^{-3})\ln(e_0^T q_M)).\end{aligned}$$

Rearranging this,

$$\begin{gathered}\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)+\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M}) =\\ (1+M^{-3})\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M})+\ln(\frac{1}{2}(e_1^T+e_0^T)q_M).\end{gathered}$$

Exponentiating,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)) = (\frac{e_0^T q_{a,M}}{e_0^T q_M})^{1+M^{-3}}\frac{1}{2}(e_1^T+e_0^T)q_M.$$

By the boundedness of the utility function,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(\bar{u}-e_0^T\kappa_M)) \geq (\frac{e_0^T q_{a,M}}{e_0^T q_M})^{1+M^{-3}}\frac{1}{2}(e_1^T+e_0^T)q_M.$$

Taking a sum over a, weighted by $\pi(a)$, and applying Jensen's inequality,

$$\frac{1}{2}(e_1^T+e_0^T)q_M\exp(\theta^{-1}M^{-2}(\bar{u}-e_0^T\kappa_M)) \geq \frac{1}{2}(e_1^T+e_0^T)q_M,$$

and therefore

$$e_0^T\kappa_M \leq \bar{u}.$$

Observing that

$$M^{-1}\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M}) \leq M^{-1}\ln(\frac{M}{c_L}) \leq M^{-1}(\frac{M}{c_L}-1) \leq c_L^{-1}, \tag{47}$$

we have

$$\theta^{-1}M^{-2}e_0^T(u_{a,M}-\kappa_M)+\ln(\frac{1}{2}(e_1^T+e_0^T)q_{a,M})\leq$$
$$M^{-2}c_L^{-1}+\ln(\frac{e_0^T q_{a,M}}{e_0^T q_M})+\ln(\frac{1}{2}(e_1^T+e_0^T)q_M).$$

Exponentiating,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(-\theta c_L^{-1}+e_0^T(u_{a,M}-\kappa_M)))\leq(\frac{e_0^T q_{a,M}}{e_0^T q_M})\frac{1}{2}(e_1^T+e_0^T)q_M$$

Using the boundedness of the utility function, then taking a sum over a, weighted by $\pi(a)$,

$$\frac{1}{2}(e_1^T+e_0^T)q_{a,M}\exp(\theta^{-1}M^{-2}(-\theta c_L^{-1}-\bar{u}-e_0^T\kappa_M))\leq\frac{1}{2}(e_1^T+e_0^T)q_M.$$

Therefore,

$$e_0^T\kappa_M\geq-\bar{u}-\theta c_L^{-1},$$

and thus

$$|e_0^T\kappa_M|\leq B_0$$

for $B_0=\bar{u}+\theta c_L^{-1}$. A similar argument applies to the other end-point ($e_M^T\kappa_M$).

Using the bound on utility and equation (47), the FOC requires that

$$\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_{a,M}}{e_0^T q_{a,M}})\leq\theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_M}{e_0^T q_M}).$$

By Lemma 12, it follows that

$$\ln(\frac{\frac{1}{2}(e_1^T+e_0^T)q_{a,M}}{e_0^T q_{a,M}})\leq\theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+M^{-1}K,$$

and therefore the constraint with $B_1=K+\theta^{-1}(\bar{u}+B_0+\theta c_L^{-1})$ is satisfied.

Similarly, the FOC for the highest state is

$$\theta^{-1}M^{-2}e_M^T(u_{a,M}-\kappa_M)+\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}{e_M^Tq_{a,M}}) =$$
$$(1+M^{-3})\ln(\frac{e_M^Tq_{a,M}}{e_M^Tq_M})+\ln(\frac{1}{2}(e_M^T+e_{M-1}^T)q_M),$$

and therefore

$$\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}{e_M^Tq_{a,M}}) \leq \theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_M}{e_M^Tq_M}),$$

implying that

$$\ln(\frac{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}{e_M^Tq_{a,M}}) \leq \theta^{-1}M^{-2}(\bar{u}+B_0+\theta c_L^{-1})+M^{-1}K,$$

and therefore

$$\ln(\frac{e_M^Tq_{a,M}}{\frac{1}{2}(e_M^T+e_{M-1}^T)q_{a,M}}) \geq -M^{-1}B_1.$$

D.4.8 Proof of Lemma 17

The first-order condition is, for any $i \in X^M \setminus \{0,M\}$ can be re-written using the function $l_{a,M}$ (and the function l_M, defined from $\hat{q}_M$ along the same lines) as

$$e_i^T(\kappa_M-u_{a,M})+\theta M^{-1}\ln(\frac{e_i^Tq_{a,M}}{e_i^Tq_M}) = \theta\frac{M^2}{(M+1)}(l_{a,M}(\frac{2i+2}{2(M+1)})-l_{a,M}(\frac{2i+1}{2(M+1)}))$$
$$-\theta\frac{M^2}{(M+1)}(l_M(\frac{2i+2}{2(M+1)})-l_M(\frac{2i+1}{2(M+1)})).$$

Note that

$$\theta M^{-1}\ln(\frac{e_i^Tq_{a,M}}{e_i^Tq_M}) \leq \theta M^{-1}\ln(\frac{1}{c_LM^{-1}}) \leq \theta M^{-1}(\frac{M}{c_L}-1) \leq \theta c_L^{-1}.$$

By Lemma 12 and Lemma 15 and the bound on utility,

$$\theta \frac{M^2}{(M+1)}(l_{a,M}(\frac{2i+2}{2(M+1)}) - l_{a,M}(\frac{2i+1}{2(M+1)}) \leq B_\kappa + \bar{u} + \theta K + \theta c_L^{-1}.$$

We also have, for all $i \in X^M \setminus \{M\}$

$$\begin{aligned} &\frac{M^2}{M+1}(l_{a,M}(\frac{2i+3}{2(M+1)}) - l_{a,M}(\frac{2i+2}{2(M+1)})) \\ = &M^2(\ln(\frac{(M+1)e_{i+1}^T q_{a,M}}{\frac{1}{2}(M+1)(e_{i+1}^T + e_i^T)q_{a,M}}) - \ln(\frac{\frac{1}{2}(M+1)(e_i^T + e_{i+1}^T)q_{a,M}}{(M+1)e_i^T q_{a,M}})) \\ &\leq 0, \end{aligned}$$

by the concavity of the log function. Observe also that, by Lemma 16,

$$l_{a,M}(\frac{2}{2(M+1)}) = (M+1)\ln(\frac{\frac{1}{2}(e_0^T + e_1^T)q_{a,M}}{e_0^T q_{a,M}}) \leq \frac{M+1}{M}B_1.$$

It follows that, for all $j \in \{2, 3, \ldots, 2M+1\}$,

$$\begin{aligned} l_{a,M}(\frac{j}{2(M+1)}) &= l_{a,M}(\frac{2}{2(M+1)}) + \sum_{k=2}^{j-1}(l_{a,M}(\frac{k+1}{2(N+1)}) - l_{a,M}(\frac{k}{2(M+1)})) \\ &\leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(j-2) + \frac{M+1}{M}B_1. \end{aligned}$$

Similarly, for all $j \in \{2, 3, \ldots, 2M+1\}$,

$$l_{a,M}(\frac{2M+1}{2(M+1)}) = l_{a,M}(\frac{j}{2(M+1)}) + \sum_{k=j}^{2M}(l_{a,M}(\frac{k+1}{2(M+1)}) - l_{a,M}(\frac{k}{2(M+1)})).$$

Observing that

$$-l_{a,M}(\frac{2M+1}{2(M+1)}) = -\ln(\frac{(M+1)e_M^T q_{a,M}}{\frac{1}{2}(M+1)(e_M^T + e_{M-1}^T)q_{a,M}}) \leq \frac{M+1}{M}B_1,$$

using Lemma 16,

$$-l_{a,M}(\frac{j}{2(M+1)}) \leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(2M - j + 1) + \frac{M+1}{M}B_1.$$

It follows that, for all $j \in \{2, 3, \ldots, 2M+1\}$,

$$\begin{aligned}|l_{a,N}(\frac{j}{2(N+1)})| &\leq \theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1})\frac{M+1}{M^2}(2M-1) + \frac{M+1}{M}B_1 \\ &\leq 4\theta^{-1}(B_\kappa + \bar{u} + \theta K + \theta c_L^{-1}) + 2B_1.\end{aligned}$$

References

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

Bloedel, Alexander, and Weijie Zhong. 2020. "The Cost of Optimally-Acquired Information." *Available at https://web.stanford.edu/~abloedel/AWB_JMP.pdf. Accessed 28 April 2021.*

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

Dacorogna, Bernard. 2007. *Direct methods in the calculus of variations.* Vol. 78, Springer Science & Business Media.

Dean, Mark, and Nathaniel Neligh. 2019. "Experimental tests of rational inattention." *Available at http://www.columbia.edu/~md3405/Working_Paper_21.pdf. Accessed 28 April 2021.*

Giaquinta, M, and S Hildebrandt. 1996. "Calculus of Variations, Vol. I. Number 310 in A Series of Comprehensive Studies in Mathematics."

Gorman, WM. 1968. "Conditions for additive separability." *Econometrica: Journal of the Econometric Society*, 605–609.

Hébert, Benjamin. 2018. "Moral Hazard and the Optimality of Debt." *The Review of Economic Studies*, 85(4): 2214–2252.

Hébert, Benjamin M, and Michael Woodford. 2019. "Rational inattention when decisions take time." National Bureau of Economic Research Working Paper no. w26415.

Leontief, Wassily. 1947. "A note on the interrelation of subsets of independent variables of a continuous function with continuous first derivatives." *Bulletin of the American mathematical Society*, 53(4): 343–350.

Myatt, David P, and Chris Wallace. 2011. "Endogenous information acquisition in coordination games." *The Review of Economic Studies*, 79(1): 340–374.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2020. "The cost of information." *arXiv preprint*, 1812.04211.

Woodford, Michael. 2008. "Inattention as a source of randomized discrete adjustment." *Unpublished*.

Yang, Ming. 2020. "Optimality of debt under flexible information acquisition." *The Review of Economic Studies*, 87(1): 487–536.