Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy

Andrew Caplin*, Mark Dean†, and John Leahy‡

Online Appendix

There are in total five appendices. The first establishes theorem 2, the recoverability theorem, which is a stand-alone result. The second establishes a series of lemmas on optimal strategies for PS cost functions and additional lemmas that link strategies and data. These are employed in all subsequent proofs. The third appendix proves theorem 3 which characterizes PS cost functions. The fourth establishes theorem 4 which characterizes UPS cost functions. The final appendix, which is significantly the longest, establishes theorem 1. While it is presented first in the paper, theorem 1 is proved last since it builds on all earlier results.

Appendix 1: Theorem 2

A1.1: NIAS, NIAC, and Existence

The first step in the proof of theorem 2 uses NIAS (A2) and NIAC (A3) to arrive at a cost function $K(\mu, Q)$ that rationalizes all observed data. The proof joins the methods of Caplin and Martin [2015] and Caplin and Dean [2015] for finitely many observations with the corresponding methods for unrestricted data introduced by Rochet [1987] and Rockafellar [1970]. This step involves entirely different logic than the second step in which we demonstrate that the rationalizing cost function is unique, and is separated out as a lemma that may itself be of independent interest.

Lemma 1.1: (Existence of Rationalizing Cost Function) Given $C \in \mathcal{C}$ satisfying A2 and A3, there exists $K \in \mathcal{K}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$.

Proof. The first step in the proof is to define the maximal utility that can be obtained given an arbitrary set of available actions A and posterior distribution $Q \in \mathcal{Q}(\mu)$,

$$\hat{G}(A, Q) \equiv \sum_{\gamma \in \Gamma(Q)} Q(\gamma)\hat{u}(\gamma, A), \tag{1}$$

*Center for Experimental Social Science and Department of Economics, New York University. Email: andrew.caplin@nyu.edu

†Department of Economics, Columbia University. Email: mark.dean@columbia.edu

‡Department of Economics, Universtity of Michigan and NBER. Email: jvleahy@umich.edu

where $\hat{u}(\gamma, A) \equiv \max_{a \in A} \bar{u}(\gamma, a)$, as defined in equation (1) in the main document. We use a constructive procedure to find $K \in \mathcal{K}$ such that, given $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$,

$$\hat{G}(A, \mathbf{Q}_P) - K(\mu, \mathbf{Q}_P) \geq \hat{G}(A, Q) - K(\mu, Q), \tag{2}$$

all $Q \in \mathcal{Q}(\mu)$. The second step shows that this function is a rationalizing cost function in the sense of this Lemma.

It simplifies the construction of the function for which inequality (2) holds to introduce the set of all pairs of decision problems and associated chosen data,

$$\mathcal{B} = \{b = (A_b, P_b) | (\mu, A_b) \in \mathcal{D} \text{ and } P_b \in C(\mu, A_b)\}.$$

Since the proof works prior by prior, we simplify from now on by fixing $\mu \in \Gamma$ in the background and treating it implicitly unless confusion would result. For example we define $\mathcal{Q}(\mu) \equiv \mathcal{Q}$, $\Omega(\mu) = \Omega$, $C(\mu, A) = C(A)$, etc., and show how to identify the corresponding section of the cost function $K(Q) \equiv K(\mu, Q)$ on $Q \in \mathcal{Q}(\mu)$.

An important construction involves associating a value with switching choice data across decision problems. Specifically, for $b, c \in \mathcal{B}$ we define $G(b, c)$ to be the maximum value associated with action set A_b and $\mathbf{Q}_{P_c}$ the revealed posterior distribution in c,

$$G(b, c) \equiv \hat{G}(A_b, \mathbf{Q}_{P_c}). \tag{3}$$

so that $G : \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ due to the finiteness of all action sets. In what follows it is important to note from (1) that this connects directly to a utility calculation introduced in the definition of the NIAC axiom in equation (14) in the main document,

$$G(b, c) = \sum_{\gamma \in \Gamma(\mathbf{Q}_{Pc})} \mathbf{Q}_{P_c}(\gamma) \hat{u}(\gamma, A_b) = \hat{U}(A_b, P_c). \tag{4}$$

Rather than directly establish existence of a qualifying cost function, we use the indirect approach of Rochet [1987] and Rockafellar [1970]. What we identify is a function $t : \mathcal{B} \to \mathbb{R}$ such that $\forall\ b, b' \in \mathcal{B}$,

$$G(b, b) - t(b) \geq G(b, b') - t(b') \tag{5}$$

If such a function can indeed be identified, it allows us to define a candidate cost function $K(Q)$ on $Q \in \mathcal{Q}$ satisfying (2). Concretely,

$$K(Q) = \begin{cases} t(b) \text{ for any } b \in \mathcal{B} \text{ such that } Q = \mathbf{Q}_{P_b}; \\ \infty \text{ if } \nexists\ b \in \mathcal{B} \text{ such that } Q = \mathbf{Q}_{P_b}. \end{cases} \tag{6}$$

Note that existence of $t : \mathcal{B} \to \mathbb{R}$ satisfying (5) ensures that this is well-defined: $t(b) = t(c)$ for any $b, c \in \mathcal{B}$ with $\mathbf{Q}_{P_b}{=}\mathbf{Q}_{P_C}$. Assume to the contrary that there exists $b, c \in \mathcal{B}$ with $\mathbf{Q}_{P_b}{=}\mathbf{Q}_{Pc}$ yet with $t(b) > t(c)$. This implies that,

$$G(b, b) - t(b) = \hat{G}(A_b, \mathbf{Q}_{P_b}) - t(b) < \hat{G}(A_b, \mathbf{Q}_{P_c}) - t(c) = G(b, c) - t(c)$$

a contradiction of condition (5). Note that K is not necessarily a qualifying cost function, as inattentive strategies are not guaranteed to be zero cost. However, following Caplin and Dean [2015] it is always possible to renormalize any cost function to make this hold, so without loss of generality we proceed assuming that K has this property.

To establish that $K \in \mathcal{K}$ as defined in (6) satisfies inequality (2), the key observation relates to inequality (5). Note first that the infinite cost of posterior distributions that are not the revealed posteriors (6) for some observed data means that the only possible reversals of inequality (2) derive from a distinct observed posterior distribution. Now consider two observations $b = (A_b, P_b), c = (A_c, P_c) \in \mathcal{B}$ and note that, by direct substitution of the definitions of G and K into (3),

$$\hat{G}(A_b, \mathbf{Q}_{P_b}) - K(\mathbf{Q}_{P_b}) = G(b,b) - t(b) \geq G(b,c) - t(c) = \hat{G}(A_b, \mathbf{Q}_{Pc}) - K(\mathbf{Q}_{Pc}),$$

establishing (2).

To construct $t : \mathcal{B} \to \mathbb{R}$ satisfying (5), we first define, for any $x, y \in \mathcal{B}$ the set $L(x,y)$ of all finite sequences starting at $x \in \mathcal{B}$ and ending at y, which we refer to as chains from x to y. Generic element $l \in L(x,y)$, comprises an ordered list of finite length $N(l) + 1$ from $\mathcal{B}$, $(b_0^l, b_1^l, ...b_{N(l)}^l) \in \mathcal{B}^{N(l)+1}$ with $b_0^l = x$ and $b_{N(l)}^l = y$. We define the value of such a chain $v : L(x,y) \to \mathbb{R}$ as,

$$v(l) = \sum_{n=0}^{N(l)} \left[G(b_{n+1}^l, b_n^l) - G(b_n^l, b_n^l) \right]; \tag{7}$$

and define also the corresponding supremal value,

$$V(x,y) = \sup_{l \in L(x,y)} v(l).$$

The function $V(x,y)$ has important qualitative properties. First among these is that, due to NIAC (A2), the supremal value of all chains that have the same start and end point is zero,

$$V(x,x) = 0. \tag{8}$$

all $x \in \mathcal{B}$. To prove this, let $(A_n^l, P_n^l) = b_n^l$ be the corresponding action and choice set. Note first that the chain that goes directly from x to x gives a value of zero so that $V(x,x) \geq 0$. To show the opposite inequality, consider an arbitrary chain $(b_0^l, b_1^l, ...b_{N(l)}^l) \in L(\bar{b}_0)$ from x to x and substitute equation (4) into equation (7) to derive,

$$v(l) = \sum_{n=0}^{N(l)} \left[G(b_{n+1}^l, b_n^l)) - G(b_n^l, b_n^l) \right] = \sum_{n=0}^{N(l)} \left[\hat{U}((A_{n+1}^l, P_n^l) - \hat{U}(A_n^l, P_n^l) \right].$$

Note by construction that $A_0^l = A_{N(l)}^l = A_x$. Given that $P_n^l \in C(A_n^l)$, NIAC (A2) directly implies that,

$$\sum_{n=0}^{N(l)} \hat{U}((A_{n+1}^l, P_n^l) \leq \sum_{n=0}^{N(l)} \hat{U}(A_n^l, P_n^l).$$

Hence indeed

$$\sum_{n=0}^{N(l)} \left[G(b_{n+1}^l, b_n^l)) - G(b_n^l, b_n^l) \right] \leq 0.$$

As the inequality holds for every element in the set it must also hold for the supremum, completing the proof of (8).

We now show that $V(x,y)$ for general $x, y \in \mathcal{B}$ is real-valued by providing real upper and lower bounds. The lower bound derives from the direct chain $\bar{l} = (x,y) \in \mathcal{B}^2$ for which,

$$V(x,y) \geq v(\bar{l}) = G(y,x) - G(x,x) \in \mathbb{R}. \tag{9}$$

We adapt this reasoning to provide an upper bound on $V(x,y)$. Specifically, given an arbitrary $l \in L(x,y)$, we define $l' \in L(x,x)$ as the chain $(b_0^l, b_1^l, ...b_{N(l)}^l, x) \in \mathcal{B}^{N(l))+2}$. Note that since $l' \in L(x,x)$ we know that it achieves no higher than the supremal value of zero,

$$v(l') \leq V(x,x) = 0. \tag{10}$$

Note also that the difference between $v(l)$ and $v(l')$ is defined by function G as:

$$v(l) + G(x,y) - G(y,y) = v(l') \leq 0;$$

where the final inequality follows direction from (10). Hence,

$$v(l) \leq G(y,y) - G(x,y).$$

Since this applies to arbitrary $l \in L(x,y)$, it applies also to the supremal value,

$$V(x,y) \leq G(y,y) - G(x,y). \tag{11}$$

An extension of this reasoning shows that, for any $x, y, z \in \mathcal{B}$,

$$V(x,y) - V(x,z) \geq G(y,z) - G(z,z). \tag{12}$$

To see this, note that for, given an arbitrary $\bar{l} \in L(x,z)$, we can define the new chain $\bar{l}' \in L(x,y)$ as the chain $(b_0^l, b_1^l, ...b_{N(l)}^l, y) \in \mathcal{B}^{N(l))+2}$. Note that the difference between $v(\bar{l})$ and $v(\bar{l}')$ is defined by function G as:

$$v(\bar{l}') = v(\bar{l}) + G(z,y) - G(z,z); \tag{13}$$

Taking the supremum on $\bar{l} \in L(x,z)$ on the RHS we arrive at,

$$\sup_{\bar{l} \in L(x,z)} v(\bar{l}) + G(z,y) - G(z,z) = V(x,z) + G(z,y) - G(z,z).$$

Applying equation (13) to a sequence of strategies $\bar{l}(n) \in L(x,z)$ converging to the supremum, we note that the corresponding sequence $\bar{l}'(n) \in L(x,y)$ can achieve no more than the corresponding supremal value, $V(x,y)$. Hence,

$$V(x,y) \geq V(x,z) + G(y,z) - G(z,z),$$

establishing (12).

Finally, we are in position to define the sought for function $t : \mathcal{B} \to \mathbb{R}$ such that (5) holds. Specifically we fix a reference element $z \in \mathcal{B}$ and define,

$$t(b) \equiv G(b,b) - V(z,b), \tag{14}$$

on $b \in \mathcal{B}$. To validate (5), we set $x = b' \in \mathcal{B}$ and $y = b$ and note by direct substitution of (14) that,

$$t(b') - t(b) = G(b', b') - G(b, b) + V(z, b) - V(z, b')$$

We now apply inequality (12) to replace the value functions on the RHS,

$$\begin{aligned} t(b') - t(b) &\geq G(b', b') - G(b, b) + G(b, b') - G(b', b') \\ &= G(b, b') - G(b, b), \end{aligned}$$

establishing (5).

We have now completed construction of a qualifying cost function $K \in \mathcal{K}$ satisfying (2). Expanding the inequality out, note that, given $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$ and associated $\mathbf{Q}_P$,

$$\sum_{\gamma \in \Gamma(\mathbf{Q}_P)} \mathbf{Q}_P(\gamma) \hat{u}(\gamma, A) - K(\mathbf{Q}_P) \geq \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \hat{u}(\gamma, A) - K(Q),$$

all $Q \in \mathcal{Q}$.

To complete the proof of Lemma 1.1 we need to show that this cost function represents the data in the sense of the Lemma: given $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$ we can find $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$ such that $P = \mathbf{P}_\lambda$. The proof is constructive. Our candidate strategy is the revealed attention strategy $\boldsymbol{\lambda}(P) = (\mathbf{Q}_P, \mathbf{q}_P)$. Since (2) holds, it suffices first to show that $(\mathbf{Q}_P, \mathbf{q}_P)$ obtains $\hat{G}(A, \mathbf{Q}_P)$, and then that it generates the data, $P = \mathbf{P}_{\boldsymbol{\lambda}(P)}$.

With regard to the first step, note first that $(\mathbf{Q}_P, \mathbf{q}_P) \in \Lambda(\mu, A)$, since, given $\omega \in \Omega(\mu)$,

$$\begin{aligned} \sum_{\{\gamma \in \Gamma(P)\}} \mathbf{Q}_P(\gamma) \gamma(\omega) &= \sum_{\{\gamma \in \cup_{a \in \mathcal{A}(P)} \bar{\gamma}_P^a\}} \sum_{\{a \in \mathcal{A}(P) | \bar{\gamma}_P^a = \gamma\}} P(a) \gamma(\omega) = \sum_{a \in \mathcal{A}(P)} P(a) \bar{\gamma}_P^a(\omega) \\ &= \sum_{a \in \mathcal{A}(P)\}} \mu(\omega) P(a|\omega) = \mu(\omega) \sum_{a \in \mathcal{A}(P)\}} P(a|\omega) = \mu(\omega). \end{aligned}$$

Note further that for any $a \in A$ with $\mathbf{q}_P(a|\gamma) > 0$ for some $\gamma = \bar{\gamma}_P^a \in \Gamma(\mathbf{Q}_P)$, we know that $a \in \mathcal{A}(P)$. Hence by NIAS (A2) we know that,

$$\sum_{\omega \in \Omega(\mu)} \bar{\gamma}_P^a(\omega) u(a, \omega) = \hat{u}(\gamma, A).$$

Hence,

$$\begin{aligned} &\sum_{\gamma \in \Gamma(\mathbf{Q}_P)} \mathbf{Q}_P(\gamma) \sum_{a \in A} \mathbf{q}_P(a|\gamma) \left(\sum_{\omega \in \Omega} \gamma(\omega) u(a, \omega) \right) = \sum_{\gamma \in \Gamma(\mathbf{Q}_P)} \mathbf{Q}_P(\gamma) \hat{u}(\gamma, A) \\ &= \hat{G}(A, \mathbf{Q}_P). \end{aligned}$$

Hence indeed $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$.

It remains only to show that $P = \mathbf{P}_{\boldsymbol{\lambda}(P)}$, which is established in Lemma 2.13 in Appendix 2. ■

A1.2: Completeness and Uniqueness

Theorem 2: Given $C \in \mathcal{C}$ satisfying A2-A4, the function $K \in \mathcal{K}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$ is unique.

Proof. By Lemma 1.1, we know that with A2 and A3 there exists $K \in \mathcal{K}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$. With the addition of Completeness, A4, we will show that any function with this property is recoverable directly from the data, hence unique.

We show how to establish the costs associated with some arbitrary $\bar{Q} \in Q^C(\mu)$. The recovery method involves constructing a parametrized path of distributions over posteriors $\bar{Q}_t \in \mathcal{Q}^C(\mu)$ for $t \in [0, 1]$, all of which are also observed in the data. As a first step in the construction, we index by n for $1 \leq n \leq N = |\Gamma(\bar{Q})|$ the posteriors $\bar{\gamma}^n \in \Gamma(\bar{Q})$. For each n we define a linear path from prior to posterior,

$$\bar{\gamma}_t^n = t\bar{\gamma}^n + (1-t)\mu, \tag{15}$$

on $t \in [0, 1]$. By construction $\bar{\gamma}_0^n = \mu$ and $\bar{\gamma}_1^n = \bar{\gamma}^n$. We define $\bar{\Gamma}_t = \cup_{n=1}^N \bar{\gamma}_t^n$ as the corresponding set of such posteriors. Finally we define the parametrized path of posterior distribution of interest $\bar{Q}_t \in \mathcal{Q}(\mu)$ by setting,

$$\bar{Q}_t(\bar{\gamma}_t^n) = \bar{Q}(\bar{\gamma}^n) \equiv \bar{Q}^n \tag{16}$$

for all $t \in [0, 1]$ and $1 \leq n \leq N$.

We wish to show that $\bar{Q}_t \in \mathcal{Q}^C(\mu)$ all $t \in [0, 1]$. By construction $\bar{Q}_1 = \bar{Q} \in Q^C(\mu)$. Also by construction each posterior distribution satisfies the Bayesian constraint,

$$\sum_n \bar{Q}_t(\bar{\gamma}_t^n)\gamma_t^n = \sum_n \bar{Q}^n \left[t\bar{\gamma}^n + (1-t)\mu\right] = \mu.$$

Finally, note that for all $t \in [0, 1)$, $\Gamma(\bar{Q}_t) \subset \bar{\Gamma}(\mu)$, so that by Completeness (A4), $\bar{Q}_t \in Q^C(\mu)$, as required.

Given $K \in \mathcal{K}$ and our parameterized posterior distributions Q_t, we define corresponding cost functions,

$$\bar{K}(t) \equiv K(\mu, \bar{Q}_t).$$

By definition the posterior distribution at $t = 0$ is inattentive and that at $t = 1$ generates $\bar{Q}$. Hence, by normalization,

$$\bar{K}(0) = 0 \text{ and } \bar{K}(1) = K(\mu, \bar{Q}).$$

Beyond this, key observations are that $\bar{K}(t)$ is continuous and convex in t. In proving this it is convenient to simplify notation. As in the last proposition, since the proof works prior by prior, it can be suppressed in all ensuing statements unless this would cause confusion.

Since $\bar{Q}_t \in \mathcal{Q}^C$, there is, for each $t \in [0, 1]$, an action set that has the given posterior distribution as its revealed posterior distribution.

Technically, Completeness (A4) implies that given $t \in [0, 1]$ there exists $(\mu, \bar{A}_t) \in \mathcal{D}$ and $\bar{P}_t \in C$ $(\mu, \bar{A}_t)$ such that $\mathbf{Q}_{\bar{P}_t} = \bar{Q}_t$, and we introduce just such a path through action sets. We now compute expected utility that is derived by setting each action as in set $\bar{A}_t$ and using it at the posteriors corresponding to all different values $s \in [0, 1]$. Note first that, while there may be several actions in principle in each set $\bar{A}_s$ that have the same revealed posterior, they all must have the same expected utility. For any $\bar{\gamma}_s^n \in \Gamma(\bar{Q}_s)$ and any $a, a' \in \bar{A}_s$ that are possibly chosen, $\min\{q(a|\bar{\gamma}_s^n), q(a'|\bar{\gamma}_s^n)\} > 0$, NIAS (A2) implies that the corresponding expected utility is equal,

$$\bar{u}(a, \bar{\gamma}_s^n) = \sum_{\omega \in \Omega} u(a, \omega)\bar{\gamma}_s^n(\omega) = \sum_{\omega \in \Omega} u(a', \omega)\bar{\gamma}_s^n(\omega) = \bar{u}(a', \bar{\gamma}_s^n). \tag{17}$$

Hence with regard to the computation of expected utility we can WLOG select one chosen action and designate it as the unique chosen action $\bar{a}_t^n \in \bar{A}_t$ for computing maximized expected utility.

The specific path of expected utility that we compute involves fixing $t \in [0, 1]$ and using the action $\bar{a}_t^n \in \bar{A}_t$ at all posteriors γ_s^n for $s \in [0, 1]$. We compute the corresponding expected utility for all pairings of parameterized action sets $\bar{A}_t$ and posterior distributions $\bar{Q}_s$ on the defined path,

$$H(t, s) = \sum_n \bar{Q}_s(\bar{\gamma}_s^n) \left(\sum_{\omega \in \Omega} \bar{\gamma}_s^n(\omega) u(\bar{a}_t^n, \omega) \right) \equiv \sum_n \bar{Q}^n \bar{u}(\bar{a}_t^n, \bar{\gamma}_s^n). \tag{18}$$

A simple observation is that $H(t, s)$ is linear on $s \in [0, 1]$. Given $\alpha, s_1, s_2 \in [0, t]$,

$$\alpha H(t, s_1) + (1 - \alpha)H(t, s_2) = H(t, \alpha s_1 + (1 - \alpha)s_2). \tag{19}$$

This follows directly from (18) since,

$$\begin{aligned} H(t, \alpha s_1 + (1 - \alpha)s_2) &= \sum_n \bar{Q}^n \bar{u}(\bar{a}_t^n, \bar{\gamma}_{\alpha s_1 + (1-\alpha)s_2}^n) \\ &= \sum_n \bar{Q}^n \left[\alpha \bar{u}(\bar{a}_t^n, \bar{\gamma}_{s_1}^n) + (1 - \alpha)\bar{u}(\bar{a}_t^n, \bar{\gamma}_{s_2}^n) \right] = \alpha H(t, s_1) + (1 - \alpha)H(t, s_2) \end{aligned}$$

As a linear function, $H(t, s)$ is differentiable in $s \in [0, 1]$. The corresponding partial derivative is of interest since we will consider a related optimization problem,

$$H_2(t, s) = \sum_n \bar{Q}^n \left(\sum_{\omega \in \Omega} \frac{\partial \bar{\gamma}_s^n(\omega)}{\partial s} u(\bar{a}_t^n, \omega) \right)$$

Substituting in for $\bar{\gamma}_s^n(\omega)$ from definition (15) that

$$\frac{\partial \bar{\gamma}_s^n(\omega)}{\partial s} = \left[\bar{\gamma}^n(\omega) - \mu(\omega) \right].$$

Hence,

$$H_2(t, s) = \sum_n \bar{Q}^n \left(\sum_{\omega \in \Omega} \left[\bar{\gamma}^n(\omega) - \mu(\omega) \right] u(\bar{a}_t^n, \omega) \right).$$

Given its importance in what follows it is valuable to simplify the notation for the dot product between the state specific vector of changes from posterior to prior and the corresponding state specific vector of utilities.

$$[\bar{\gamma}^n - \mu] \cdot u(\bar{a}_t^n) \equiv \sum_{\omega \in \Omega} \left[\bar{\gamma}^n(\omega) - \mu(\omega) \right] u(\bar{a}_t^n, \omega). \tag{20}$$

The optimization problem that we study relies on that observation that, for all $t \in [0,1]$,

$$H(t,t) - \bar{K}(t) \geq H(t,s) - \bar{K}(s), \tag{21}$$

all $s \in [0,1]$. This follows directly from inequality (19), since we know that $(\mu, \bar{A}_t) \in \mathcal{D}$ and $\bar{P}_t \in C(\mu, \bar{A}_t)$, so that the corresponding revealed posterior $\mathbf{Q}_{P_t} = \bar{Q}_t$ maximizes expected utility net of costs of attention costs and

$$\begin{aligned} H(t,t) - \bar{K}(t) &= \hat{G}(\bar{A}_t, \bar{Q}_t) - K(\bar{Q}_t) \geq \hat{G}(\bar{A}_t, \bar{Q}_s) - K(\bar{Q}_s) \\ &\geq H(t,s) - \bar{K}(s). \end{aligned}$$

In essence the left-hand side expression is the optimized expected utility at t in the observed data under the CIR, while the RHS represents expected utility for a policy that would be feasible in this set of using posterior distribution Q_s and at each γ_s^n picking $\bar{a}_t^n \in \bar{A}_t$, which may or may not in fact be optimal.

We use the function H to prove continuity and convexity of $\bar{K}(t)$. With regard to convexity, consider $t_1 \neq t_2 \in [0,1]$ and their average

$$\bar{t} = \frac{t_1 + t_2}{2}.$$

Direct application of inequality (21) establishes that,

$$H(\bar{t},\bar{t}) - K(\bar{t}) \geq 0.5 \left[H(\bar{t}, t_1) - \bar{K}(t_1) + H(\bar{t}, t_2) - \bar{K}(t_2) \right] \tag{22}$$

From (19) it is clear that,

$$0.5 \left[H(\bar{t}, t_1) + H(\bar{t}, t_2) \right] = H(\bar{t}, \bar{t}).$$

Substitution in (22) yields,

$$\bar{K}(\bar{t}) \leq 0.5 \left[\bar{K}(t_1) + \bar{K}(t_2) \right]$$

implying that $\bar{K}(t)$ is convex as claimed.

Given that $\bar{K}(t)$ is convex, it is continuous on its interior, $t \in (0,1)$. Moreover the only possible discontinuities at the boundary point involve an increase in costs,

$$\bar{K}(0) > \lim_{t \downarrow 0} \bar{K}(t) \text{ or } \bar{K}(1) > \lim_{t \uparrow 1} \bar{K}(t). \tag{23}$$

To see that these cannot hold, apply (21) at the corresponding end-point,

$$\begin{aligned} H(0,0) - \bar{K}(0) &\geq H(0,\epsilon) - \bar{K}(\epsilon); \\ H(1,1) - \bar{K}(1) &\geq H(1, 1-\epsilon) - \bar{K}(1-\epsilon) \end{aligned}$$

all $\epsilon > 0$. (19) implies continuity of $H(0,\epsilon)$ and $H(0,\epsilon)$ in ϵ, we conclude therefore that $\bar{K}(0) \leq \lim_{\epsilon \downarrow 0} \bar{K}(\epsilon)$ and $\bar{K}(1) \leq \lim_{\epsilon \uparrow 1} \bar{K}(\epsilon)$, both of which directly contradict (23).

Given that $\bar{K}(t)$ is convex and continuous, it can have at most a countable number of non-differentiable points (Rockafellar [1970]) and hence is integrable. By the fundamental theorem of calculus $\bar{K}(t)$ can be reconstructed from its derivative, $\bar{K}'(t)$, which is defined except on a set of measure zero. Hence since $\bar{K}(0) = 0$, and we can recover its final value focusing only on points of differentiability as,

$$K(\bar{Q}) = \bar{K}(1) = \int_0^1 \bar{K}'(t)dt, \tag{24}$$

We now show how to characterize the derivative $\bar{K}'(t)$ from the path of utilities identified above, noting that $H(t,s)$ satisfies (19) and is hence everywhere differentiable in s. Consider a point $t \in (0,1)$ of differentiability of $\bar{K}$ and problem $(\mu, \bar{A}_t)$ for which t therefore maximizes,

$$\max_{s \in [0,1]} H(t,s) - \bar{K}(s) = \sum_n \bar{Q}^n \bar{u}(\bar{a}_t^n, \gamma_s^n) - \bar{K}(s). \tag{25}$$

Hence the corresponding first order condition for maximizing (25) at $s = t$ is,

$$\bar{K}'(t) = H_2(t,t) = \sum_n \bar{Q}^n \frac{\partial \bar{u}(\bar{a}_t^n, \gamma_s^n)}{\partial s}. \tag{26}$$

Substituting the definition in (17),

$$\bar{u}(\bar{a}_t^n, \gamma_s^n) = \sum_{\omega \in \Omega(\mu)} \bar{\gamma}_s^n(\omega) u(\bar{a}_t^n, \omega) = \sum_{\omega \in \Omega(\mu)} [s\bar{\gamma}^n(\omega) + (1-s)\mu(\omega)] u(\bar{a}_t^n, \omega).$$

Hence the chain rule yields,

$$\frac{\partial \bar{u}(\bar{a}_t^n, \gamma_s^n)}{\partial s} = \sum_{\omega \in \Omega} [\bar{\gamma}^n(\omega) - \mu(\omega)] \, u_2(\bar{a}_t^n, \omega) = [\bar{\gamma}^n - \mu] \cdot u(\bar{a}_t^n),$$

where the last equation uses the simpler notation for the dot product introduced in (20).

The corresponding first order condition for maximizing (25) at $s = t$ is,

$$\bar{K}'(t) = H_2(t,t) = \sum_n \bar{Q}^n \left([\bar{\gamma}^n - \mu] \cdot u(\bar{a}_t^n) \right). \tag{27}$$

Combining (27) and(24),

$$\begin{aligned} K(\mu, \bar{Q}) &= \int_0^1 \sum_n \bar{Q}^n \left\{ [\bar{\gamma}^n - \mu] \cdot u(\bar{a}_t^n) \right\} dt \\ &= \sum_n \bar{Q}^n \left[\bar{\gamma}^n - \mu \right] \cdot \int_0^1 u(\bar{a}_t^n) dt, \end{aligned}$$

since the dot product survives under integration. This completes the constructive procedure for computing the cost function. ■

We state the form of this computation as a corollary since it is the jumping off point for the proof of theorem 3.

Corollary 1 *Given $C \in \mathcal{C}$ satisfying A2-A4, the unique function $K \in \mathcal{K}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$ can be computed for each $(\mu, \bar{Q}) \in \mathcal{F}$ with $\bar{Q} \in \mathcal{Q}^C(\mu)$ by enumerating* *the support* $\Gamma(\bar{Q}) = \{\bar{\gamma}^n | 1 \leq n \leq N\}$ *and computing,*

$$K(\mu, \bar{Q}) \equiv \sum_n \bar{Q}(\bar{\gamma}^n) T^C_\mu(\bar{\gamma}^n, \bar{Q}) - T^C_\mu(\mu, \bar{Q}),$$

where $T^C_\mu(\mu, \bar{Q}) = 0$ *and,*

$$T^C_\mu(\bar{\gamma}^n, \bar{Q}) \equiv [\bar{\gamma}^n - \mu] \cdot \int_0^1 u(\bar{a}^n_t) dt.$$

Appendix 2: The PS Model and Convex Analysis

For the analysis of the PS model, we fix a specific prior $\bar{\mu} \in \Gamma$. Costs are then defined by a strictly convex function $T_{\bar{\mu}} : \Gamma(\bar{\mu}) \to \bar{\mathbb{R}}$, real valued on $\tilde{\Gamma}(\bar{\mu})$, such that, given $Q \in \mathcal{Q}(\bar{\mu})$,

$$K(\bar{\mu}, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) T_{\bar{\mu}}(\gamma) - T_{\bar{\mu}}(\mu).$$

We define a "$\bar{\mu}$-based" net utility function for strategies $\lambda \in \Lambda(\mu, A)$ for $\mu \in \Delta(\Gamma(\bar{\mu}))$ using the costs associated with $\bar{\mu}$,

$$N_{\bar{\mu}}(\mu, \lambda) \equiv U(\lambda) - \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) T_{\bar{\mu}}(\gamma) = V(\bar{\mu}, \lambda) - T_{\bar{\mu}}(\bar{\mu}). \tag{28}$$

Another key function is the "net utility" of choosing action $a \in A$ given posterior $\gamma \in \Gamma(\bar{\mu})$,

$$N^a_{\bar{\mu}}(\gamma) \equiv \sum_{\omega \in \Gamma(\bar{\mu})} u(a, \omega) \gamma(\omega) - T_{\bar{\mu}}(\gamma) = \bar{u}(\gamma, a) - T_{\bar{\mu}}(\gamma). \tag{29}$$

We introduce also a mixture operation.

Definition 1 *Given any finite set of strategies,* $\{\lambda(l) = (Q_l, q_l) \in \Lambda(\mu(l)\}_{1 \leq l \leq L}$ *, and strictly positive probability weights* $\{\alpha(l)\}_{1 \leq l \leq L}$*, define the corresponding* ***mixture strategy*** $\lambda(\alpha) = (Q_\alpha, q_\alpha) \in \Lambda$ *by:*

$$\begin{aligned} Q_\alpha(\gamma) &= \sum_l \alpha(l) Q_l(\gamma) \text{ all } \gamma \in \Gamma(Q_\alpha); \\ q_\alpha(a|\gamma) &= \frac{\sum_l \alpha(l) q_l(a|\gamma) Q_l(\gamma)}{Q_\alpha(\gamma)} \text{ all } \gamma \in \Gamma(Q_\alpha), \ a \in \mathcal{A}(\lambda(\alpha)); \end{aligned}$$

where $\Gamma(Q_\alpha) = \cup_l \Gamma(Q_l)$ *and* $\mathcal{A}(\lambda(\alpha)) = \cup_l \mathcal{A}(\lambda(l))$*. Define* $\mu(\alpha) = \sum_l \alpha(l) \mu(l)$ *as the corresponding weighted average of priors.*

A2.1: Linearity and Uniqueness

Lemma 2.1 (Linearity under Mixing): Given $K \in \mathcal{K}^{PS}$, $(\bar{\mu}, A) \in \mathcal{D}$, and, for $1 \leq l \leq L$, strategies $\lambda(l) = (Q_l, q_l) \in \Lambda(\mu(l), A)$ and probability weights $\alpha(l)$,

$$N_{\bar{\mu}}(\mu(\alpha), \lambda(\alpha)) = \sum_l \alpha(l) N_{\bar{\mu}}(\mu(l), \lambda(l)).$$

Proof. Note that,

$$\sum_{\gamma \in \Gamma(Q_\alpha)} \gamma Q_\alpha(\gamma) = \sum_{\gamma \in \Gamma(Q_\alpha)} \gamma \left[\sum_l \alpha(l) Q_l(\gamma) \right] = \sum_l \alpha(l) \mu(l) = \mu(\alpha),$$

so that $\lambda(\alpha) \in \Lambda(\mu(\alpha), A)$. With regard to utility,

$$\begin{aligned} U(\lambda(\alpha)) &= \sum_{\gamma \in \Gamma(Q_\alpha)} Q_\alpha(\gamma) \sum_{a \in A} q_\alpha(a|\gamma) \bar{u}(\gamma, a) \\ &= \sum_{\gamma \in \Gamma(Q_\alpha)} Q_\alpha(\gamma) \sum_{a \in A} \left[\frac{\sum_l \alpha(l) q_l(a|\gamma) Q_l(\gamma)}{Q_\alpha(\gamma)} \right] \bar{u}(\gamma, a) \\ &= \sum_{\gamma \in \Gamma(Q_\alpha)} \sum_l \alpha(l) Q_l(\gamma) \sum_{a \in A} q_l(a|\gamma) \bar{u}(\gamma, a) \\ &= \sum_l \alpha(l) \sum_{\gamma \in \Gamma(\lambda(l))} Q_l(\gamma) \sum_{a \in A} q_l(a|\gamma) \bar{u}(\gamma, a) = \sum_l \alpha(l) U(\lambda(l)). \end{aligned}$$

Likewise,

$$\begin{aligned} \sum_{\gamma \in \Gamma(\alpha)} Q_\alpha(\gamma) T_{\bar{\mu}}(\gamma) &= \sum_{\gamma \in \Gamma(Q_\alpha)} \left[\sum_l \alpha(l) Q_l(\gamma) \right] T_{\bar{\mu}}(\gamma) \\ &= \sum_l \alpha(l) \sum_{\gamma \in \Gamma(\lambda(l))} Q_l(\gamma) T_{\bar{\mu}}(\gamma). \end{aligned}$$

In combination these imply,

$$\begin{aligned} N_{\bar{\mu}}(\mu(\alpha), \lambda(\alpha)) &= U(\lambda(\alpha)) - \sum_{\gamma \in \Gamma(\alpha)} Q_\alpha(\gamma) T_{\bar{\mu}}(\gamma) \\ &= \sum_l \alpha(l) \left[U(\lambda(l)) - \sum_{\gamma \in \Gamma(\lambda(l))} Q_l(\gamma) T_{\bar{\mu}}(\gamma) \right] = \sum_l \alpha(l) N_{\bar{\mu}}(\mu(l), \lambda(l)). \end{aligned}$$

establishing the Lemma. ■

Lemma 2.2: (Mixing and Optimality) Given $K \in \mathcal{K}^{PS}$, $(\bar{\mu}, A) \in \mathcal{D}$, and strategies $\lambda \in \Lambda(\bar{\mu}, A)$ and $\lambda(l) = (Q_l, q_l) \in \Lambda(\bar{\mu}, A)$ for $1 \leq l \leq L$, together with strictly positive probability $\alpha(l) > 0$ such that $\lambda = \sum_{l=1}^{L} \alpha(l)\lambda(l)$,

$$\lambda \in \hat{\Lambda}(\bar{\mu}, A|K) \Longleftrightarrow \lambda(l) \in \hat{\Lambda}(\bar{\mu}, A|K) \text{ all } l.$$

Proof. If $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$ we know that,

$$N_{\bar{\mu}}(\bar{\mu}, \lambda) = \hat{V}(\bar{\mu}, A|K) + T_{\bar{\mu}}(\bar{\mu}).$$

We can then apply Lemma 2.1, which has particularly simple implications in this case since all strategies have prior $\bar{\mu}$, to conclude that,

$$N_{\bar{\mu}}(\bar{\mu}, \lambda) = \sum_{l} \alpha(l) N_{\bar{\mu}}(\bar{\mu}, \lambda(l)) = \hat{V}(\bar{\mu}, A|K) + T_{\bar{\mu}}(\bar{\mu}).$$

It directly follows that $\lambda(l) \in \hat{\Lambda}(\bar{\mu}, A)$ all l, since together $N_{\bar{\mu}}(\bar{\mu}, \lambda(l)) \leq \hat{V}(\bar{\mu}, A|K) + T_{\mu}(\bar{\mu})$ and $\sum_l \alpha(l) = 1$ imply $N_{\bar{\mu}}(\bar{\mu}, \lambda(l)) = \hat{V}(\bar{\mu}, A|K) + T_{\bar{\mu}}(\bar{\mu})$. Conversely, if all strategies $\lambda(l)$ are optimal,

$$N_{\bar{\mu}}(\bar{\mu}, \lambda(l)) = \hat{V}(\bar{\mu}, A|K) + T_{\bar{\mu}}(\bar{\mu}),$$

all l. Hence this applies also to $N_{\bar{\mu}}(\bar{\mu}, \lambda)$ which directly implies $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$. ■

Lemma 2.3 (Unique Posterior Lemma): Given $K \in \mathcal{K}^{PS}$, $(\bar{\mu}, A) \in \mathcal{D}$, $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$, and $a \in \mathcal{A}(\lambda)$, there exists a unique posterior $\gamma \in \Gamma(Q_\lambda)$ such that $q_\lambda(a|\gamma) > 0$. We denote this posterior γ_λ^a.

Proof. Suppose to the contrary that $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$ involves distinct posteriors $\gamma^1 \neq \gamma^2 \in \Gamma(\bar{\mu})$ with $q_\lambda(a|\gamma^l) > 0$ for $l = 1, 2$ for some action $a \in \mathcal{A}(\lambda)$. Define a new strategy $\bar{\lambda}$ that is unchanged except in that the unconditional probability of each of these posteriors is reduced accordingly by $Q_\lambda(\gamma^l) q(a|\gamma^l)$ with the probability of action a reduced to zero and that of all other chosen options expanded proportionately to fill the gap. In strategy $\bar{\lambda}$ the additional probability is assigned to the mean posterior $\bar{\gamma}$,

$$\bar{\gamma}(j) \equiv \frac{\sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) \gamma^l(j)}{\sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l)}$$

with $q_{\bar{\lambda}}(a|\bar{\gamma}) = 1$. Note that there is no change in gross utility,

$$U(\bar{\lambda}) - U(\lambda) = \sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) \bar{u}(\gamma^l, a) - \sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) \bar{u}(\gamma^l, a) = 0,$$

due to additivity of action-specific expected utility across posteriors: given $\alpha_1, \alpha_2 > 0$,

$$\begin{aligned} \alpha_1 \bar{u}(\gamma^1, a) + \alpha_2 \bar{u}(\gamma^2, a) &= \alpha_1 \sum_{\omega} u(a, \omega) \gamma^1(\omega) + \alpha_2 \sum_{\omega} u(a, \omega) \gamma^2(\omega) \\ &= \sum_{\omega} u(a, \omega) \left[\alpha_1 \gamma^1(\omega) + \alpha_2 \gamma^2(\omega) \right]. \end{aligned}$$

To establish the contradiction and with it complete the proof, it suffices that costs would reduced in the switch from λ to $\bar{\lambda}$. This follows directly from the strict convexity of T, which implies that,

$$\begin{aligned} K(\bar{\mu}, Q_\lambda) - K(\bar{\mu}, Q_{\bar{\lambda}}) &= \sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l)) \left[\left[\frac{\sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) T(\gamma^l)}{\sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l)} \right] - T_{\bar{\mu}}(\bar{\gamma}) \right] \\ &> \sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) \left[T_{\bar{\mu}} \left[\frac{\sum_{l=1,2} Q_\lambda(\gamma^l) q(a|\gamma^l) \gamma^l}{\bar{Q}(a)} \right] - T_{\bar{\mu}}(\bar{\gamma}) \right] = T_{\bar{\mu}}(\bar{\gamma}) - T_{\bar{\mu}}(\bar{\gamma}) = 0. \end{aligned}$$

■

Lemma 2.4 (Unique Optimal Strategy): Given $K \in \mathcal{K}^{PS}$, $(\bar{\mu}, A) \in \mathcal{D}$, and $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$ with $\Gamma(Q_\lambda)$ linearly independent,

$$|\Gamma(Q_\lambda)| = |A| \Longrightarrow \lambda = \hat{\Lambda}(\bar{\mu}, A|K).$$

Proof. The first observation is that, with $\Gamma(Q_\lambda)$ linearly independent, no strict subset of the posteriors is even feasible. To see this, consider $\lambda' \in \Lambda(\bar{\mu}, A)$ with $\Gamma(Q_{\lambda'}) \subset \Gamma(Q_\lambda)$. Note that for feasibility,

$$\Sigma_{\gamma \in \Gamma(Q_\lambda)} \gamma Q_\lambda(\gamma) = \Sigma_{\gamma \in \Gamma(Q_{\lambda'})} \gamma Q_{\lambda'}(\gamma) = \mu,$$

Subtraction yields,

$$\Sigma_{\gamma \in \Gamma(Q_\lambda)} \gamma \left[Q_\lambda(\gamma) - Q_{\lambda'}(\gamma) \right] = 0,$$

whereupon linear independence implies that $Q_\lambda(\gamma) = Q_{\lambda'}(\gamma)$ all $\gamma \in \Gamma(Q_\lambda)$.

By Lemma 2.3, note that no action is chosen at more than one posterior in an optimal strategy. Hence with $|\Gamma(\lambda)| = |A|$ this means that each action is chosen at only one posterior, at which it is chosen deterministically. Note that if there were two distinct deterministic strategies using the same set of posteriors, mixing them would be optimal by Lemma 2.2, yet would involve the same action at two distinct posteriors, which is inconsistent with Lemma 2.3. Hence changing action choices in any way at the given posteriors must strictly lower the payoff. Hence there is no alternative optimal strategy that involves retaining posterior set $\Gamma(Q_\lambda)$ yet changing action choices.

The final possibility for generating multiplicity is if there exists some $\lambda' \in \hat{\Lambda}(\bar{\mu}, A|K)$ with some posterior $\gamma' \in \Gamma(Q_{\lambda'})$ that is not in $\Gamma(Q_\lambda)$. Since $\mathcal{A}(\lambda) = A$, we can identify $a' \in A \cap \mathcal{A}(\lambda')$ with $q(a'|\gamma') > 0$. By Lemma 2.2, the strategy $\frac{\lambda}{2} + \frac{\lambda'}{2}$ is also optimal. This identifies a supposedly optimal attention strategy in which a' is chosen at two distinct posteriors, contradicting Lemma 2.3 and completing the proof. ■

A2.2: Lagrangian Analysis

Our next series of results relate to the lower epigraph of $N_{\bar{\mu}}(\mu, \lambda)$. To study this, we reduce the dimension of the state space by defining $J = |\Omega(\bar{\mu})|$, correspondingly labeling states, and letting

$\Gamma^{J-1}(\bar{\mu})$ denote the space of distributions of interest,

$$\Gamma^{J-1}(\bar{\mu}) = \left\{ \mu \in \mathbb{R}_{+}^{J-1} | \sum_{j=1}^{J-1} \mu(j) \leq 1 \right\};$$

with $\mu(_J) = 1 - \sum_{j=1}^{J-1} \mu(j)$ left as implicit.

It will also be use of to define the net utility of a strategy λ according to the costs associated with posterior $\bar{\mu}$.

$$N_{\bar{\mu}}(\mu, \lambda) \equiv U(\lambda) - \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) T_{\bar{\mu}}(\gamma).$$

Lemma 2.5 (Convexity): The lower epigraph of $N_{\bar{\mu}}(\mu, \lambda)$,

$$\mathcal{E}(\bar{\mu}, A) \equiv \left\{ (y, \mu) \in \mathbb{R} \times \Gamma^{J-1}(\bar{\mu}) \; | y \leq N_{\bar{\mu}}(\mu, \lambda) \text{ some } \lambda \in \Lambda(\mu, A) \right\},$$

is a convex set.

Proof. Consider $\mu(l) \in \Delta(\Omega)$ for $l \in \{0, 1\}$, with corresponding strategies $\lambda(l) = (Q_l, q_l) \in \Lambda(\mu(l), A)$ and supports $\Gamma(\lambda_l)$. By the Linearity Lemma we know that, for any $\alpha \in (0, 1)$, strategy $\lambda(\alpha)$ satisfies,

$$N_{\bar{\mu}}(\mu(\alpha), \lambda(\alpha)) = \alpha N_{\bar{\mu}}(\lambda(0)) + (1 - \alpha) N_{\bar{\mu}}(\lambda(1)).$$

Hence,

$$(\alpha N_{\bar{\mu}}(\lambda(0)) + (1 - \alpha) N_{\bar{\mu}}(\lambda(1)), \mu(\alpha)) \in \mathcal{E}(\bar{\mu}, A),$$

establishing the Lemma. ■

Lemma 2.6 (Lagrangean): Given $K \in \mathcal{K}^{PS}$ and $(\bar{\mu}, A) \in \mathcal{D}$, $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$ if and only if $\exists \theta \in \mathbb{R}^{J-1}$ s.t.,

$$N_{\bar{\mu}}^{a}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{a' \in A, \gamma' \in \Gamma(\bar{\mu})} N_{\bar{\mu}}^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j),$$

all $\gamma \in \Gamma(\bar{\mu})$ and $a \in A$, with equality if $\gamma \in \Gamma(Q_\lambda)$ and $q_\lambda(a|\gamma) > 0$.

Proof. If $\lambda \in \hat{\Lambda}(\bar{\mu}, A)$, we know that $(N_{\bar{\mu}}(\bar{\mu}, \lambda), \bar{\mu})$ is an upper boundary point of the convex set $\mathcal{E}(\bar{\mu}, A)$. Hence there exists a supporting hyperplane $\theta(j)$ for $0 \leq j \leq J - 1$ with $\theta(j) \neq 0$ for some j such that,

$$\theta(0)y - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \theta(0) N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j),$$

all $(y, \gamma) \in \mathcal{E}(\bar{\mu}, A)$.

We show now that $\theta(0) > 0$. Suppose to the contrary that $\theta(0) < 0$. In this case we can renormalize to $\theta(0) = -1$ to conclude that,

$$-y - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq -N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j), \tag{30}$$

which is a clear contradiction, since the left hand side is unbounded as we lower y arbitrarily. Finally we show that $\theta(0) \neq 0$. Note that by definition $\bar{\mu}(j) > 0$ all j,

$$\min \left\{ \min_{1 \leq j \leq J-1} \{\bar{\mu}(j)\}, 1 - \sum_{j=1}^{J-1} \bar{\mu}(j) \right\} > 0.$$

Finally suppose that $\theta(0) = 0$, so that,

$$\sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j) \leq \sum_{j=1}^{J-1} \theta(j)\gamma(j).$$

all $\gamma \in \Gamma(\bar{\mu})$. To minimize the expression on the RHS, one can set $\gamma(\bar{j}) = 1$, where the index $\bar{j}$ is chosen so that,

$$\theta(\bar{j}) = \min_{1 \leq j \leq J-1} \{\theta(j)\}.$$

Hence the inequality can be valid only if $\theta(j) = \theta(\bar{j}) = \bar{\theta}$ for all j. Hence what is required is,

$$\bar{\theta} \sum_{j=1}^{J-1} \bar{\mu}(j) \leq \bar{\theta} \sum_{j=1}^{J-1} \gamma(j),$$

all $\gamma \in \Gamma(\bar{\mu})$. Since $0 < \sum_{j=1}^{J-1} \bar{\mu}(j) < 1$ while $\sum_{j=1}^{J-1} \gamma(j)$ has a range that includes 0 and 1, this is impossible unless $\bar{\theta} = 0$, a contradiction to the non-zero separating plane.

Given that $\theta(0) > 0$, we renormalize to $\theta(0) = 1$ and conclude that,

$$y - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j).$$

Given $\gamma' \in \Gamma$ and $a' \in A$, we know that $(N_{\bar{\mu}}^{a'}(\gamma'), \gamma') \in \mathcal{E}(\bar{\mu}, A)$, so that,

$$N_{\bar{\mu}}^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \leq N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j). \tag{31}$$

Now consider any decomposition of the optimal strategy, $\lambda = \sum_{l=1}^{L} \alpha(l)\lambda(l)$ for a finite set

$\{\lambda(l) = (Q_l, q_l) \in \Lambda(\mu(l), A)\}_{1\leq l\leq L}$, and probability weights $\{\alpha(l)\}_{1\leq l\leq L}$. Lemma 2.1 implies,

$$\begin{aligned} N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j) &= \sum_{l} \alpha(l) N_{\bar{\mu}}(\mu(l), \lambda(l)) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j) \\ &= \sum_{l} \alpha(l) \left[N_{\bar{\mu}}(\mu(l), \lambda(l)) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) \right], \end{aligned}$$

since $\bar{\mu}(j) = \sum_l \alpha(l)\mu_l(j)$. By inequality (31) none of the terms in the weighted average on the RHS can be higher than the LHS since $(N_{\bar{\mu}}(\mu(l), \lambda(l)) \in \mathcal{E}(\bar{\mu}, A)$. Hence they are all equal to it,

$$N_{\bar{\mu}}(\mu_l, \lambda(l)) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) = N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j). \tag{32}$$

We now provide a simple decomposition of strategy λ using only inattentive strategies. We index possible posteriors $\gamma \in \Gamma(Q_\lambda)$ as γ^l and define the inattentive strategy $\lambda(l) \in I(\gamma^l)$ by setting $q_{\lambda(l)}(a|\gamma^l) = q_\lambda(a|\gamma^l)$. Setting the weights as $\alpha(l) = Q_\lambda(\gamma^l)$ accomplishes this decomposition. The special feature of such inattentive strategies is that,

$$N_{\bar{\mu}}(\mu_l, \lambda(l)) = \sum_{a\in A} q_\lambda(a|\gamma^l) \left[\bar{u}(a, \gamma^l) - T_{\bar{\mu}}(\gamma^l)\right] = \sum_{a\in A} q_\lambda(a|\gamma^l) N^a_{\bar{\mu}}(\gamma^l).$$

Hence,

$$\begin{aligned} N_{\bar{\mu}}(\mu_l, \lambda(l)) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) &= \sum_{a\in A} q_\lambda(a|\gamma^l) N^a_{\bar{\mu}}(\gamma^l) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) \\ &= \sum_{a\in A} q_\lambda(a|\gamma^l) \left[N^a_{\bar{\mu}}(\gamma^l) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) \right] \end{aligned}$$

where the first line follows directly, the second follows because $\sum_{a\in A} q_\lambda(a|\gamma^l) = 1$. Hence equation (32) implies,

$$\sum_{a\in A} q_\lambda(a|\gamma^l) \left[N^a_{\bar{\mu}}(\gamma^l) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) \right] = N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j);$$

Hence by (31), given $\gamma^l \in \Gamma(Q_\lambda)$ and $q_\lambda(a|\gamma^l) > 0$,

$$N^a_{\bar{\mu}}(\gamma^l) - \sum_{j=1}^{J-1} \theta(j)\mu_l(j) = N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j),$$

and again applying (31) completes the proof of necessity.

With regard to sufficiency, consider $\lambda \in \Lambda(\bar{\mu}, A)$ for which there exists $\theta(j)$ such that, given

$\gamma \in \Gamma(Q_\lambda)$ and $q_\lambda(a|\gamma) > 0$,

$$N_{\bar{\mu}}^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \leq N_{\bar{\mu}}^{a}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j),$$

all $\gamma' \in \Gamma$ and $a' \in A$. Carry out the decomposition called for above into inattentive strategies indexing $\gamma \in \Gamma(Q_\lambda)$ as γ^l, defining $\lambda(l) \in I(\gamma^l)$ by setting $q_{\lambda(l)}(a|\gamma^l) = q_\lambda(a|\gamma^l)$, and using the linearity lemma to conclude that,

$$N_{\bar{\mu}}(\bar{\mu}, \lambda) = \sum_l Q_\lambda(\gamma^l) \sum_a q_{\lambda(l)}(a|\gamma^l) N_{\bar{\mu}}^a(\gamma^l).$$

Hence, since at all possible posteriors $\gamma \in \Gamma(Q_\lambda)$ and corresponding actions achieve the same value of $N_{\bar{\mu}}^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j)$, this precise value also applies to strategy λ, so that,

$$\begin{aligned} N_{\bar{\mu}}^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) &= \sum_l Q_\lambda(\gamma^l) \sum_a q_{\lambda(l)}(a|\gamma^l) \left[N_{\bar{\mu}}^a(\gamma^l) - \sum_{j=1}^{J-1} \theta(j)\gamma^l(j) \right] \\ &= N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j) \left[\sum_l Q_\lambda(\gamma^l)\gamma^l(j) \right] = N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j), \end{aligned}$$

where $\sum_l Q_\lambda(\gamma^l)\gamma^l(j) = \bar{\mu}(j)$ by Bayes' rule. Hence,

$$N_{\bar{\mu}}^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \leq N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j)$$

all $\gamma' \in \Gamma$ and $a' \in A$.

Now consider an arbitrary strategy $\eta \in \Lambda(\bar{\mu}, A)$. Repeat precisely the corresponding decomposition into inattentive strategies indexing $\gamma \in \Gamma(Q_\eta)$ as $\tilde{\gamma}^l$, defining $\eta(l) \in I(\tilde{\gamma}^l)$ by setting $q_{\eta(l)}(a|\tilde{\gamma}^l) = q_\eta(a|\gamma^l)$ to conclude that,

$$N_{\bar{\mu}}(\bar{\mu}, \eta) = \sum_l Q_\eta(\tilde{\gamma}^l) \sum_a q_{\eta(l)}(a|\tilde{\gamma}^l) N_{\bar{\mu}}^a(\tilde{\gamma}^l).$$

Using a similar decomposition to the above we have that

$$N_{\bar{\mu}}(\bar{\mu}, \eta) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j) = \sum_l Q_\eta(\tilde{\gamma}^l) \sum_a q_{\eta(l)}(a|\tilde{\gamma}^l) \left[N_{\bar{\mu}}^a(\tilde{\gamma}^l) - \sum_{j=1}^{J-1} \theta(j)\tilde{\gamma}^l(j) \right].$$

Since in addition,

$$N_{\bar{\mu}}^a(\tilde{\gamma}^l) - \sum_{j=1}^{J-1} \theta(j)\tilde{\gamma}^l(j) \leq N_{\bar{\mu}}(\bar{\mu}, \lambda) - \sum_{j=1}^{J-1} \theta(j)\bar{\mu}(j),$$

for all a and $\tilde{\gamma}^l$ we conclude that,

$$N_{\bar{\mu}}(\bar{\mu},\eta) - \sum_{j=1}^{J-1}\theta(j)\bar{\mu}(j) \leq N_{\bar{\mu}}(\bar{\mu},\lambda) - \sum_{j=1}^{J-1}\theta(j)\bar{\mu}(j).$$

Hence $N_{\bar{\mu}}(\bar{\mu},\eta) \leq N_{\bar{\mu}}(\bar{\mu},\lambda)$, establishing optimality. ■

Lemma 2.7 (Feasibility Implies Optimality): Given $\bar{\mu} \in \Gamma$ and $K \in \mathcal{K}^{PS}$, $\tilde{\Gamma}(\mu) \subset \hat{\Gamma}(\mu|K)$, and there exists a one-to-one function on the optimal posterior set $\hat{\Gamma}(\mu|K)$, $f_{\bar{\mu}} : \hat{\Gamma}(\mu|K) \to \mathcal{A}$, with range $\mathcal{A}_{\bar{\mu}}$ such that, given a set $A \subset \mathcal{A}_{\bar{\mu}}$ with $(\bar{\mu}, A) \in \mathcal{D}$ and given $\lambda = (Q_\lambda, q_\lambda) \in \Lambda(\bar{\mu}, A)$,

$$q_\lambda(f_{\bar{\mu}}(\gamma)|\gamma) = 1 \text{ for all } \gamma \in \Gamma(Q_\lambda) \Longrightarrow \lambda \in \hat{\Lambda}(\bar{\mu}, A|K).$$

Proof. Given $K \in \mathcal{K}^{PS}$ and $\bar{\mu} \in \Gamma$, we define for each $\bar{\gamma} \in \hat{\Gamma}(\mu|K)$ a particular action $f_{\bar{\mu}}(\bar{\gamma})$ with the defining property that, using a strictly convex function $T_{\bar{\mu}} : \Gamma(\bar{\mu}) \to \mathbb{R}$ associated with $K \in \mathcal{K}^{PS}$, the corresponding function $N_{\bar{\mu}}$ has maximal value of zero at $\bar{\gamma}$:

$$N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\phi) = \bar{u}(\phi, f_{\bar{\mu}}(\bar{\gamma})) - T_{\bar{\mu}}(\phi) \leq N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\bar{\gamma}) = 0, \tag{33}$$

all $\phi \in \Gamma(\bar{\mu})$. Note that this is sufficient to establish the result, defining $\mathcal{A}_{\bar{\mu}}$ to be the union of $f_{\bar{\mu}}(\bar{\gamma})$. In this case if we consider any set $A \subset \mathcal{A}_{\bar{\mu}}$ with $(\bar{\mu}, A) \in \mathcal{D}$ and a strategy $\lambda \in \Lambda(\bar{\mu}, A)$ such that $q_\lambda(f_{\bar{\mu}}(\gamma)|\gamma) = 1$ for all $\gamma \in \Gamma(Q_\lambda)$, we can conclude that this strategy satisfies the sufficient conditions for optimality in Lemma 2.6 when we set all multipliers to zero, $\theta(j) = 0$, establishing that $\lambda \in \hat{\Lambda}(\bar{\mu}, A)$.

We define the function satisfying equation (33) in two phases. First we consider interior beliefs $\bar{\gamma}$ with $\bar{\gamma}(\omega) > 0$ all $\omega \in \Omega(\bar{\mu})$. We note first that since $-T_{\bar{\mu}}(\gamma)$ is a strictly concave function, there exist multipliers $\bar{\beta}(j)$ on $0 \leq j \leq J-1$, not all zero, such that,

$$-\bar{\beta}(0)T_{\bar{\mu}}(\phi) - \sum_{j=1}^{J-1}\bar{\beta}(j)\phi(j) \leq -\bar{\beta}(0)T_{\bar{\mu}}(\bar{\gamma}) - \sum_{j=1}^{J-1}\bar{\beta}(j)\bar{\gamma}(j), \tag{34}$$

all $\phi \in \Gamma(\bar{\mu})$. Given that $\bar{\gamma}$ satisfies $\bar{\gamma}(\omega) > 0$ all $\omega \in \Omega(\bar{\mu})$ we can mimic the proof in the second paragraph of Lemma 2.6 above to establish that $\bar{\beta}(0) \neq 0$. It is also not possible that $\bar{\beta}(0) < 0$. To see this suppose this were so. In this case we could renormalize to $\bar{\beta}(0) = -1$ in (34),

$$T_{\bar{\mu}}(\phi) - \sum_{j=1}^{J-1}\bar{\beta}(j)\phi(j) \leq T_{\bar{\mu}}(\bar{\gamma}) - \sum_{j=1}^{J-1}\bar{\beta}(j)\bar{\gamma}(j), \tag{35}$$

all $\phi \in \Gamma(\bar{\mu})$. Given that $\bar{\gamma}(\omega) > 0$ all $\omega \in \Omega(\bar{\mu})$, we can find two distinct beliefs $\phi_1, \phi_2 \in \Gamma(\bar{\mu})$ such that $\bar{\gamma} = \frac{\phi_1+\phi_2}{2}$. Averaging inequality (35) applied to each of ϕ_1, ϕ_2 separately, we conclude that,

$$\begin{aligned}\frac{T_{\bar{\mu}}(\phi_1) - \sum_{j=1}^{J-1}\bar{\beta}(j)\phi_1(j) + T_{\bar{\mu}}(\phi_2) - \sum_{j=1}^{J-1}\bar{\beta}(j)\phi_2(j)}{2} &= \frac{T_{\bar{\mu}}(\phi_1) + T_{\bar{\mu}}(\phi_2)}{2} - \sum_{j=1}^{J-1}\bar{\beta}(j)\bar{\gamma}(j) \\ &\leq T_{\bar{\mu}}(\bar{\gamma}) - \sum_{j=1}^{J-1}\bar{\beta}(j)\bar{\gamma}(j).\end{aligned}$$

We conclude therefore that,

$$0.5T_{\bar{\mu}}(\phi_1) + 0.5T_{\bar{\mu}}(\phi_2) \leq T_{\bar{\mu}}(\frac{\phi_1+\phi_2}{2}), \tag{36}$$

which contradicts strict convexity of T.

With $\bar{\beta}(0) > 0$, we can renormalize to $\bar{\beta}(0) = 1$ in (34) and flip signs to conclude that,

$$T_{\bar{\mu}}(\phi) + \sum_{j=1}^{J-1} \bar{\beta}(j)\phi(j) \geq T_{\bar{\mu}}(\bar{\gamma}) + \sum_{j=1}^{J-1} \bar{\beta}(j)\bar{\gamma}(j). \tag{37}$$

Now define action $f_{\bar{\mu}}(\bar{\gamma})$ so that, for $1 \leq k \leq J$,

$$u(f_{\bar{\mu}}(\bar{\gamma}), k) = \begin{cases} \left[T_{\bar{\mu}}(\bar{\gamma}) + \sum_{j=1}^{J-1} \bar{\beta}(j)\bar{\gamma}(j)\right] - \bar{\beta}(k) \text{ for } 1 \leq k \leq J-1; \\ \left[T_{\bar{\mu}}(\bar{\gamma}) + \sum_{j=1}^{J-1} \bar{\beta}(j)\bar{\gamma}(j)\right] \text{ for } k = J. \end{cases}.$$

By construction, given $\phi \in \Gamma(\bar{\mu})$ and so

$$\begin{aligned} N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\phi) &\equiv \sum_{k=1}^{J} u(f_{\bar{\mu}}(\bar{\gamma}), k)\phi(k) - T_{\bar{\mu}}(\phi) = \\ &= \left[T_{\bar{\mu}}(\bar{\gamma}) + \sum_{j=1}^{J-1} \bar{\beta}(j)\bar{\gamma}(j)\right] - \left[T_{\bar{\mu}}(\phi) + \sum_{k=1}^{J-1} \bar{\beta}(k)\phi(k)\right] \leq 0. \end{aligned}$$

with $N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\bar{\gamma}) = 0$, where the last inequality derives directly from (37).

Given a boundary posterior $\bar{\gamma} \in \Gamma(Q_\lambda)$ with $\bar{\gamma}(\omega) = 0$ some $\omega \in \Omega(\bar{\mu})$ we cannot guarantee that the multiplier $\beta(0)$ in (34) is non-zero (Shannon is a counterexample). The remaining cases therefore involve boundary posteriors that are part of an optimal strategy for some decision problem - i.e. $\gamma \in \hat{\Gamma}(\bar{\mu}|K)$. By definition there exists an optimal strategy $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\bar{\mu}, A)$ with $\bar{\gamma} \in \Gamma($ $Q_\lambda)$, and so by Lemma 2.6, there exists $\bar{\theta} \in \mathbb{R}^{J-1}$ such that, for $\bar{a}$ with $q_{\bar{\lambda}}(\bar{a}|\bar{\gamma}) > 0$, then,

$$\begin{aligned} \sum_{j=1}^{J} u(\bar{a}, j)\phi(j) - T_{\bar{\mu}}(\phi) - \sum_{j=1}^{J-1} \bar{\theta}(j)\phi(j) &= N_{\bar{\mu}}^{\bar{a}}(\phi) - \sum_{j=1}^{J-1} \bar{\theta}(j)\phi(j) = \\ &\leq N_{\bar{\mu}}^{\bar{a}}(\bar{\gamma}) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j) = \sum_{j=1}^{J} u(\bar{a}, j)\bar{\gamma}(j) - T_{\bar{\mu}}(\bar{\gamma}) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j) \end{aligned}$$

all $\phi \in \Gamma(\bar{\mu})$. Rearrangement yields,

$$\sum_{j=1}^{J} u(\bar{a}, j)\phi(j) - \sum_{j=1}^{J-1} \bar{\theta}(j)\phi(j) - \left[\sum_{j=1}^{J} u(\bar{a}, j)\bar{\gamma}(j) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j)\right] \bar{\gamma}(j) + T_{\bar{\mu}}(\bar{\gamma}) - T_{\bar{\mu}}(\phi) \leq 0.$$

Now define action $f_{\bar{\mu}}(\bar{\gamma}) \in \mathcal{A}$ so that, for $1 \leq k \leq J$,

$$u(f_{\bar{\mu}}(\bar{\gamma}), k) = \begin{cases} u(\bar{a}, k) - \bar{\theta}(k) - \left[\sum_{j=1}^{J} u(\bar{a}, j)\bar{\gamma}(j) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j) - T_{\bar{\mu}}(\bar{\gamma}) \right] & \text{for } 1 \leq k \leq J-1; \\ u(\bar{a}, J) - \left[\sum_{j=1}^{J} u(\bar{a}, j)\bar{\gamma}(j) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j) - T_{\bar{\mu}}(\bar{\gamma}) \right] & \text{for } k = J. \end{cases}$$

By construction, given $\phi \in \Gamma(\bar{\mu})$ and defining $\phi(J) = 1 - \sum_{j=1}^{J-1} \phi(j) \geq 0$ on $\phi \in \Gamma(\bar{\mu})$,

$$N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\phi) = \sum_{k=1}^{J} u(\bar{a}, k)\phi(k) - \sum_{k=1}^{J-1} \bar{\theta}(k)\ \phi(k) - \left[\sum_{j=1}^{J} u(\bar{a}, j)\bar{\gamma}(j) - \sum_{j=1}^{J-1} \bar{\theta}(j)\bar{\gamma}(j) - T_{\bar{\mu}}(\bar{\gamma}) \right] - T_{\bar{\mu}}(\phi) \leq 0,$$

with $N_{\bar{\mu}}^{f_{\bar{\mu}}(\bar{\gamma})}(\bar{\gamma}) = 0$.

To complete the proof, note that the 1-1 nature of $f_{\bar{\mu}}(\bar{\gamma})$ follows since otherwise there exists an optimal strategy that selects the same action at two different posteriors, which would contradict Lemma 2.3. ■

A2.3: Preservation of Optimality

We have already seen that mixing preserves optimality. There are other important operations that ensure preservation of optimality.

Lemma 2.8: (Perturbed Payoffs and Optimality) Consider $(\bar{\mu}, A) \in \mathcal{D}$ and $\lambda \in \hat{\Lambda}(\bar{\mu}, A|K)$. Given any unchosen action $b \in A \backslash \mathcal{A}(\lambda)$, consider $h(b) \in \mathcal{A}$ with $u(h(b), \omega) < u(b, \omega)$ all $\omega \in \Omega(\bar{\mu})$ and define $A' = \mathcal{A}(\lambda) \cup_{b \in A/\mathcal{A}(\lambda)} h(b)$. Then

$$\lambda' \in \hat{\Lambda}(\bar{\mu}, A'|K) \Longrightarrow \mathcal{A}(\lambda') \subset \mathcal{A}(\lambda)$$

Proof. This follows by a direct contradiction. Note that the strategy λ remains feasible, so that the value must be no lower in the new decision problem,

$$\hat{V}(\bar{\mu}, A'|K) \geq \hat{V}(\bar{\mu}, A|K).$$

Now suppose that there was a strategy $\lambda' = (Q', q') \in \hat{\Lambda}(\bar{\mu}, A'|K)$ with $h(b) \in \mathcal{A}(\lambda')$ some $b \in A \backslash \mathcal{A}(\lambda)$. Then this would have to achieve $\hat{V}(\bar{\mu}, A'|K)$ so that,

$$V(\bar{\mu}, \lambda'|K) \geq \hat{V}(\bar{\mu}, A|K). \tag{38}$$

Now define strategy $\lambda'' = (Q'', q'') \in \Lambda(\bar{\mu}, A)$ with $Q'' = Q'$ and any chosen actions $h(b)$ for $b \in A \backslash \mathcal{A}(\lambda)$ replaced by a corresponding b that maps to it (easiest to make h $1-1$)

$$q''(a|\gamma) = \begin{cases} q'(a|\gamma) \text{ if } a \in \mathcal{A}(\lambda); \\ q'(b|\gamma) \text{ if } a = h(b) \text{ some } b \in A \backslash \mathcal{A}(\lambda); \end{cases}$$

Note that this action achieves strictly higher utility than λ' but cannot achieve more than optimal strategy λ,

$$\hat{V}(\bar{\mu}, A|K) \geq V(\bar{\mu}, \lambda''|K) > V(\bar{\mu}, \lambda'|K),$$

contradicting (38) and establishing the Lemma. ■

Lemma 2.9: (Intersecting Posteriors and Intersecting Actions) Given $(\bar{\mu}, A_1) \in \mathcal{D}$, $\lambda_1 = (Q_1, q_1) \in \hat{\Lambda}(\bar{\mu}, A_1|K)$, and $Q_2 \in \hat{\mathcal{Q}}(\bar{\mu})$ with $\Gamma(Q_1) \cap \Gamma(Q_2) \neq \emptyset$, there exists $(\bar{\mu}, A_2) \in \mathcal{D}$ and $\bar{\lambda}(2) = (\bar{Q}_2, \bar{q}_2) \in \hat{\Lambda}(\bar{\mu}, A_2)$ with $\bar{Q}_2(\gamma) = Q_2(\gamma)$ and,

$$\bar{q}_2(a|\gamma) = q_1(a|\gamma),$$

all $\gamma \in \Gamma(Q_1) \cap \Gamma(Q_2)$.

Proof. Consider $(\bar{\mu}, A_1) \in \mathcal{D}$, and $\lambda(1) = (Q_1, q_1) \in \hat{\Lambda}(\bar{\mu}, A_1|K)$. By the Lagrangian Lemma there exists $\theta \in \mathbb{R}^{J-1}$ s.t.,

$$N_{\bar{\mu}}^{a}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{a' \in A_1, \gamma' \in \Gamma(\bar{\mu})} N_{\bar{\mu}}^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \equiv \bar{N}, \tag{39}$$

all $\gamma \in \Gamma(\bar{\mu})$ and $a \in A_1$, with equality if $\gamma \in \Gamma(Q_1)$ and $q_{\lambda(1)}(a|\gamma) > 0$. To simplify notation in this step we define subsets $A_1(\gamma) \subset A_1$ on $\gamma \in \Gamma(Q_1)$ by the condition,

$$a \in A_1(\gamma) \Longleftrightarrow q_{\lambda(1)}(a|\gamma) > 0,$$

By (39), we know that, given $\gamma, \gamma' \in \Gamma(Q_1)$,

$$N_{\bar{\mu}}^{a_1(\gamma)}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) = N_{\bar{\mu}}^{a_1(\gamma')}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) = \bar{N}.$$

for any $a_1(\gamma) \in A_1(\gamma)$ and $a_1(\gamma') \in A_1(\gamma')$.

We now associate with each remaining possible posterior $\gamma \in \Gamma(Q_2)/\Gamma(Q_1)$ an action $a_2(\gamma)$. In defining these payoffs, we make essential use of the function $f_{\bar{\mu}}(\gamma) \in \mathcal{A}$ from Lemma 2.7 which is well defined on $\Gamma(Q_2)$ since $Q_2 \in \hat{\mathcal{Q}}(\bar{\mu})$. We make use also of the Lagrangians $\theta(j)$ and the net utility functions and value $\bar{N}$ in (39). Specifically, we define $a_2(\gamma)$ on $\gamma \in \Gamma(Q_2)/\Gamma(Q_1)$ to have state dependent payoffs,

$$u(a_2(\gamma), j) = \begin{cases} \bar{N} + u(f_{\bar{\mu}}(\gamma), j) + \theta(j) \text{ for } 1 \leq j \leq J-1 \\ \bar{N} + u(f_{\bar{\mu}}(\gamma), J) \end{cases}$$

We define the set of such actions, as well as their union with actions selected in the first step:

$$\begin{aligned} B_2 &= \{a_2(\gamma) | \gamma \in \Gamma(Q_2)/\Gamma(Q_1)\}; \\ A_2 &= B_2 \cup \{A_1(\gamma) | \gamma \in \Gamma(Q_1) \cap \Gamma(Q_2)\}. \end{aligned}$$

We now construct the strategy of interest $\bar{\lambda}(2) = (\bar{Q}_2, \bar{q}_2)$ according to the prescription in the statement of the Lemma. We first specify $\bar{Q}_2(\gamma) = Q_2(\gamma)$, so that $\Gamma(\bar{Q}_2) = \Gamma(Q_2)$. With regard to

$\bar{q}_2(\gamma)$, it is specified differently according to whether or not $\gamma \in \Gamma(Q_1) \cap \Gamma(Q_2)$:

$$\bar{q}_2(a|\gamma) = \begin{cases} q_1(a|\gamma) \text{ for } \gamma \in \Gamma(Q_1) \cap \Gamma(Q_2); \\ 1 \text{ if } \gamma \in \Gamma(Q_2)/\Gamma(Q_1) \text{ and } a = a_2(\gamma); \\ 0 \text{ if } \gamma \in \Gamma(Q_2)/\Gamma(Q_1) \text{ and } a \neq a_2(\gamma). \end{cases}$$

Note by construction that $\mathcal{A}(\bar{\lambda}(2)) = A_2$, and also that, since $Q_2 \in \hat{\mathcal{Q}}(\bar{\mu})$,

$$\sum_{\gamma \in \Gamma(\bar{Q}_2)} \gamma \bar{Q}_2(\gamma) = \sum_{\gamma \in \Gamma(Q_2)} \gamma Q_2(\gamma) = \bar{\mu},$$

so that $\bar{\lambda}(2) \in \Lambda(\bar{\mu}, A_2)$.

It remains to show that $\bar{\lambda}(2) \in \hat{\Lambda}(\bar{\mu}, A_2)$. To establish this we use the sufficiency aspect of the Lagrangian Lemma. Specifically, we use the original Lagrangians $\theta(j)$ in (39) and show that,

$$N^a_{\bar{\mu}}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{a' \in A_2, \gamma' \in \Gamma(\bar{\mu})} N^{a'}_{\bar{\mu}}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) = \bar{N} \tag{40}$$

all $\gamma \in \Gamma(\bar{\mu})$ and $a \in A_2$, with equality if $\gamma \in \Gamma(\bar{Q}_2)$ and $\bar{q}_2(a|\gamma) > 0$.

The relevant equality for $a \in A_1(\gamma)$ for $\gamma \in \Gamma(Q_1) \cap \Gamma(Q_2)$ is directly implied by (39). We now consider $\gamma \in \Gamma(Q_2)/\Gamma(Q_1)$ and the corresponding chosen action $a_2(\gamma)$. By construction,

$$N^{a_2(\gamma)}_{\bar{\mu}}(\gamma) = \bar{N} + N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}(\gamma) + \sum_{j=1}^{J-1} \theta(j)\gamma(j).$$

Hence,

$$\begin{aligned} N^{a_2(\gamma)}_{\bar{\mu}}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) &= \bar{N} + N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}(\gamma) + \sum_{j=1}^{J-1} \theta(j)\gamma(j) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \\ &= \bar{N} + N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}(\gamma) = \bar{N}, \end{aligned}$$

since $N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}(\gamma) = 0$, confirming the requisite equality.

It remains to show that the inequality aspect of (40) holds,

$$N^a_{\bar{\mu}}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \bar{N},$$

all $a \in A_2$ and $\gamma \in \Gamma(\bar{\mu})$. That this holds for $a \in A_1(\gamma)$ for $\gamma \in \Gamma(Q_1) \cap \Gamma(Q_2)$ is directly implied by (39). It remains to confirm this for $a = a_2(\gamma) \in B_2$ for $\gamma \in \Gamma(Q_2)/\Gamma(Q_1)$ and $\gamma' \in \Gamma(\bar{\mu})$. In this case, the result follows from the defining properties of $N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}$. Given $\gamma' \in \Gamma(\bar{\mu})$,

$$N^{a_2(\gamma)}_{\bar{\mu}}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) = \bar{N} + N^{f_{\bar{\mu}}(\gamma)}_{\bar{\mu}}(\gamma') \leq \bar{N},$$

since $N_{\bar{\mu}}^{f_{\bar{\mu}}(\gamma)}(\gamma') \leq 0$ all $\gamma' \in \Gamma(\bar{\mu})$. This completes the proof. ■

Lemma 2.10 (Decomposition and Uniqueness): Given $\lambda \in \hat{\Lambda}(\bar{\mu}, A)$ there exist strategies $\lambda^*(l) = (Q_l^*, q_l^*) \in \hat{\Lambda}(\bar{\mu}, A)$ for $1 \leq l \leq L$ and corresponding probability weights $\alpha(l) > 0$ such that,

$$\lambda \equiv \sum_{l=1}^{L} \alpha(l)\lambda^*(l),$$

with each strategy $\lambda^*(l)$ uniquely optimal with regard to the chosen actions $\mathcal{A}[\lambda^*(l)]$.

$$\hat{\Lambda}(\bar{\mu}, \mathcal{A}[\lambda^*(l)]) = \lambda^*(l).$$

Proof. We first show that there exists a decomposition $\lambda(l) = (Q_{\lambda(l)}, q_{\lambda(l)}) \in \hat{\Lambda}(\bar{\mu}, A)$ for $1 \leq l \leq L$ and corresponding probability weights $\alpha(l)$ such that,

$$\lambda \equiv \sum_{l=1}^{L} \alpha(l)\lambda(l),$$

with each set $\Gamma(Q_{\lambda(l)})$ linearly independent. The proof is constructive. If $\Gamma(Q_\lambda)$ is linearly independent, we are done. If not, then we know from Caratheodory's theorem that since $\Gamma(Q_\lambda)$ contains $\bar{\mu}$ in its convex hull, there exists a linearly independent set $\Gamma(1) \subset \Gamma(Q_\lambda)$ with $|\Gamma(1)| < |\Gamma(Q_\lambda)|$ that also has $\bar{\mu}$ in its convex hull. Hence there exist strictly positive probability weights $Q_1^{LI}(\gamma) > 0$ on $\gamma \in \Gamma(1)$ (extended to $\Gamma(Q_\lambda)$ by setting probabilities of excluded posteriors to zero) such that $\bar{\mu} = \sum_{\gamma \in \Gamma(1)} \gamma Q_1^{LI}(\gamma)$. We define strategy $\lambda(1) \in \Lambda(\bar{\mu}, A)$ to satisfy $\Gamma(Q_{\lambda(1)}) = \Gamma(1)$ with precisely this distribution of posteriors,

$$Q_{\lambda(1)}(\gamma) = Q_1^{LI}(\gamma),$$

and with the same mixed strategy action choice as in strategy λ,

$$q_{\lambda(1)}(a|\gamma) = q_\lambda(a|\gamma).$$

We now identify the smallest scalar $\pi(1) \in (0, 1)$ such that,

$$\pi(1)Q_1^{LI}(\gamma) = Q_\lambda(\gamma),$$

some $\gamma \in \Gamma(Q_{\lambda(1)})$. That such a scalar exists follows from the fact that,

$$\sum_{\gamma \in \Gamma(Q_{\lambda(1)}))} Q_1^{LI}(\gamma) = \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) = 1,$$

with all components in both sums strictly positive and with $\left|\Gamma(Q_{\lambda(1)})\right| < |\Gamma(Q_\lambda)|$.

Define $\hat{\Gamma}(1) = \Gamma(Q_\lambda)$ and $Q_1 = Q_\lambda$ to start the iteration. We now define function $Q_2(\gamma)$ on $\gamma \in \Gamma(Q_\lambda)$ by,

$$Q_2(\gamma) = \frac{Q_1(\gamma) - \pi(1)Q_{\lambda(1)}(\gamma)}{1 - \pi(1)} \geq 0,$$

Note that these define a probability distribution on $\Gamma(Q_\lambda)$,

$$\sum_{\gamma\in\Gamma(Q_\lambda)} Q_2(\gamma) = \frac{\sum_{\gamma\in\Gamma(Q_\lambda)} Q_1(\gamma) - \pi(1)\sum_{\gamma\in\Gamma(Q_\lambda)} Q_1^{LI}(\gamma)}{1-\pi(1)} = 1.$$

Correspondingly, we define,

$$\tilde{\Gamma}(2) = \{\gamma \in \Gamma | Q_2(\gamma) > 0\},$$

noting that $\left|\tilde{\Gamma}(2)\right| < \left|\hat{\Gamma}(1)\right|$, since by construction there exists $\gamma \in \Gamma(Q_\lambda)$ with $\pi(1)Q_{\lambda(1)}(\gamma) = Q_\lambda(\gamma)$ so that $Q_2(\gamma) = 0$. Note that the mean is preserved,

$$\begin{aligned}
\sum_{\gamma\in\tilde{\Gamma}(2)} \gamma Q_2(\gamma) &= \sum_{\gamma\in\Gamma(Q_\lambda)} \gamma Q_2(\gamma) = \sum_{\gamma\in\Gamma(Q_\lambda)} \gamma \left[\frac{Q_1(\gamma) - \pi(1)Q_1^{LI}(\gamma)}{1-\pi(1)}\right] \\
&= \frac{1}{1-\pi(1)}\left[\sum_{\gamma\in\Gamma(Q_\lambda)} \gamma Q_1(\gamma) - \pi(1)\sum_{\gamma\in\Gamma(Q_\lambda)} \gamma Q_1^{LI}(\gamma)\right] \\
&= \frac{\mu}{1-\pi(1)}\left[1-\pi(1)\right] = \bar{\mu}.
\end{aligned}$$

We define strategy $\tilde{\lambda}(2) \in \Lambda(\bar{\mu}, A)$ to involve precisely these posteriors, $Q_{\tilde{\lambda}(2)}(\gamma) = Q_2(\gamma)$ on $\gamma \in \Gamma(2) = \Gamma(Q_{\tilde{\lambda}(2)})$, with the same mixed action strategies as in λ,

$$q_{\tilde{\lambda}(2)}(a|\gamma) = q_\lambda(a|\gamma).$$

If set $\tilde{\Gamma}(2)$ is linearly independent we define $\lambda(2) = \tilde{\lambda}(2) \in \Lambda(\bar{\mu}, A)$ and stop the iteration. If not, we reapply Caratheodory's theorem and identify a linearly independent set $\Gamma(2) \subset \tilde{\Gamma}(2)$ that retains $\bar{\mu}$ in its convex hull, hence for which there exist strictly positive probability weights $Q_2^{LI}(\gamma) > 0$ on $\gamma \in \Gamma(2)$ such that $\bar{\mu} = \sum_{\gamma\in\Gamma(2)} Q_2^{LI}(\gamma)\gamma$. In this case, we define strategy $\lambda(2) \in \Lambda(\bar{\mu}, A)$ to involve precisely these posteriors, $\Gamma(\lambda(2)) = \Gamma(2)$ with the corresponding probability weights, $Q_{\lambda(2)}(\gamma) = Q_2^{LI}(\gamma)$, and again with the same mixed action choice as in strategy λ, $q_{\lambda(2)}(a|\gamma) = q_\lambda(a|\gamma)$. Rounding out the iterative process, we then define $\pi(2) \in (0,1)$ as the smallest number such that,

$$\pi(2)Q_{\lambda(2)}(\gamma) = Q_2(\gamma),$$

some $\gamma \in \Gamma(Q_{\lambda(2)})$. Finally, we define $Q_3(\gamma)$ on $\gamma \in \Gamma(Q_\lambda)$ by,

$$Q_3(\gamma) = \frac{Q_2(\gamma) - \pi(2)Q_{\lambda(2)}(\gamma)}{1-\pi(2)} \geq 0,$$

and $\tilde{\Gamma}(3) = \{\gamma \in \Gamma | Q_3(\gamma) > 0\}$. We continue in iterative fashion defining non-empty sets of posterior $\tilde{\Gamma}(l)$, linearly independent subsets $\Gamma(l) \subset \tilde{\Gamma}(l)$ and corresponding strategies $\lambda(l) \in \Lambda(\bar{\mu}, A)$. This iteration is completed in a finite number of steps, $L \in \mathbb{N}$, since $\left|\tilde{\Gamma}(l+1)\right| < \left|\tilde{\Gamma}(l)\right|$.

The above construction provides us with a set of strategies $\{\lambda(l)\}_{1\leq l\leq L}$ that are feasible, $\lambda(l) \in \Lambda(\bar{\mu}, A)$, and that have linearly independent posteriors. By construction, the distribution of the posteriors for Lth strategy are given by

$$Q_{\lambda(L)}(\gamma) = \frac{Q_\lambda(\gamma)}{\Pi_{l=1}^{L-1}(1-\pi(l))} - \sum_{l=1}^{L-1} \frac{\pi(l) Q_{\lambda(l)}(\gamma)}{\Pi_{k=l}^{L-1}(1-\pi(l))}$$

We now show that we can reverse engineer the construction to identify probability weights $\alpha(l)$ such that

$$\lambda(\alpha) = \sum_{l=1}^{L} \alpha(l)\lambda(l) = \lambda.$$

Specifically, we define

$$\alpha(l) = \Pi_{k=1}^{l-1} [1-\pi(k)] \pi(l)$$

for $1 \leq l \leq L-1$ (using the convention that $\alpha(1) = \Pi_{k=1}^{0} [1-\pi(k)] \pi(1) = \pi(1)$), and

$$\alpha(L) = \Pi_{k=1}^{L-1} [1-\pi(k)]$$

First note that these weights sum to 1, as

$$\begin{aligned} & \alpha(1) + \alpha(2) + \ldots + \alpha(L-1) + \alpha(L) \\ = \; & \pi(1) + \pi(2)(1-\pi(1)) + \ldots + \pi(L-1)\Pi_{k=1}^{L-2} [1-\pi(k)] + (1-\pi(L-1))\Pi_{k=1}^{L-2} [1-\pi(k)] \end{aligned}$$

The final two terms collapse to $\Pi_{k=1}^{L-2} [1-\pi(k)]$, which can then be combined with the term from $\alpha(L-2)$ in order to give $\Pi_{k=1}^{L-3} [1-\pi(k)]$. Iterating on this process leaves eventually

$$\pi(1) + (1-\pi(1)) = 1$$

To confirm that indeed

$$\lambda(\alpha) = \sum_{l=1}^{L} \alpha(l)\lambda(l) = \lambda,$$

we need to show only that the unconditional posterior probabilities are the same,

$$Q_{\lambda(\alpha)}(\gamma) = \sum_{l} \alpha(l) Q_{\lambda(l)}(\gamma) = Q_\lambda(\gamma);$$

since the construction ensures that all conditional action strategies are identical,

$$q_{\lambda(l)}(a|\gamma) = q_\lambda(a|\gamma),$$

all l. Note that,

$$\begin{aligned} & \sum_{l=1}^{L} \alpha(l) Q_{\lambda(l)}(\gamma) \\ = \; & \sum_{l=1}^{L-1} \Pi_{k=1}^{l-1} [1-\pi(k)] \pi(l) Q_{\lambda(l)}(\gamma) + \Pi_{k=1}^{L-1}(1-\pi(k)) \left[\frac{Q_\lambda(\gamma)}{\Pi_{l=1}^{L-1}(1-\pi(l))} - \sum_{l=1}^{L-1} \frac{\pi(l) Q_{\lambda(l)}(\gamma)}{\Pi_{k=l}^{L-1}(1-\pi(l))} \right] \\ = \; & Q_\lambda(\gamma). \end{aligned}$$

Given that $\lambda \equiv \sum_{l=1}^{L} \alpha(l)\lambda(l)$ and $\lambda \in \hat{\Lambda}(\bar{\mu}, A)$, we apply Lemma 2.2 to conclude that $\lambda(l) \in \hat{\Lambda}(\bar{\mu}, A)$ all l. We now take each strategy $\lambda(l)$ in turn. We move to a pure strategy versions $\lambda^*(l,m) = (Q^*_{l,m}, q^*_{l,m})$. For each such strategy we set $Q^*_{l,m}(\gamma) = Q_{\lambda(l)}(\gamma)$ and then take each possible posterior $\gamma \in \Gamma(l)$, selecting one action $a \in A$ that is chosen with positive probability at that posterior, $q_{\lambda(l)}(a|\gamma) > 0$, and setting its probability to 1,

$$q^*_{l,m}(a|\gamma) = 1.$$

Note that $N_{\bar{\mu}}(\bar{\mu}, \lambda(l)) = N_{\bar{\mu}}(\bar{\mu}, \lambda^*(l,m))$ since optimality implies that all options chosen at any given posterior produce the same value, so that $\lambda^*(l,m) \in \hat{\Lambda}(\bar{\mu}, A)$ all l and m. We repeat this exercise for all possible combinations of actions chosen according to $\lambda(l)$ at posteriors $\gamma \in \Gamma(Q_l)$, using M to denote the number of such combinations, then appropriately weight these strategies together with weights such that $\sum_{m=1}^{M} \alpha^*(l,m) = \alpha(l)$ and

$$\sum_{m=1}^{M} \frac{\alpha^*(l,m)}{\alpha(l)} \lambda^*(l,m) = \lambda(\lambda).$$

To complete the proof we consider the cardinality of the set of chosen actions, $\mathcal{A}[\lambda^*(l)] \subset A$. Note by the Lemma 2.3 that since $\lambda^*(l) \in \hat{\Lambda}(\bar{\mu}, A)$, each chosen action is associated with a unique posterior, so that,

$$|\Gamma(Q^*_l)| = |\mathcal{A}[\lambda^*(l)]|.$$

Together these put us in position to apply Lemma 2.4 to complete the proof: since $\lambda^*(l) \in \hat{\Lambda}(\bar{\mu}, A)$, $\Gamma(Q_{\lambda^*(l)}) = \Gamma(Q_{\lambda(l)}) \subset \Gamma$ is linearly independent, and $\left|\Gamma\left(Q_{\lambda^*(l)}\right)\right| = \mathcal{A}[\lambda^*(l)]$, the optimal strategy is unique,

$$\lambda^*(l) = \hat{\Lambda}(\bar{\mu}, \mathcal{A}[\lambda^*(l)]).$$

■

A2.4: Generalized PS Models

From this point forward through the remainder of this appendix, there is no need to consider changes in the prior within a given proof, so that we remove the over-bar, using μ in place of $\bar{\mu}$, and correspondingly defining the generic decision problem to be (μ, A) rather than $(\bar{\mu}, A)$.

At a key point in the proof of theorem 2 we need to consider variants of the PS cost function in which the T function is not strictly convex. To simplify the proof, it is convenient to consider T functions that take infinite value on unchosen posteriors.

Definition 2 *We define $K \in \mathcal{K}$ to be* ***generalized*** *PS, $K \in \mathcal{K}^{GPS}$, if, given $\mu \in \Gamma$ and $Q \in \mathcal{Q}(\mu)$,*

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) T_{\mu}(\gamma) - T_{\mu}(\mu),$$

if $\hat{\Gamma}(\mu|K)$ is a convex set and $T_{\mu} : \Gamma(\mu) \to \bar{\mathbb{R}}$ is real-valued $\hat{\Gamma}(\mu|K)$ and infinite-valued on $\gamma \notin \hat{\Gamma}(\mu|K)$. We define the corresponding convexified cost function $K^{CONV} \in \mathcal{K}$ by defining T^{CONV}_{μ} to *be the* ***convex hull*** *of* T_μ*. This is defined as the greatest convex function majorized by* T_μ*, and is shown by Rockafellar [1970] (page 36) to be equal to,*

$$T_\mu^{CONV}(\gamma) = \inf\{\sum_{m=1}^{M} \alpha(m)T(\gamma(m))| \sum_{m=1}^{M} \alpha(m)\gamma(m) = \gamma\},$$

over weights $\alpha(m) > 0$ *satisfying* $\sum \alpha(m) = 1$*. We then define the cost function*

$$K^{CONV}(\mu, Q) = \sum_{\gamma\in\Gamma(Q)} Q(\gamma)T_\mu^{CONV}(\gamma) - T_\mu^{CONV}(\mu).$$

Lemma 2.11: Convexification and Optimal Strategies Given $C \in \mathcal{C}$ with a CIR representation $K \in \mathcal{K}^{GPS}$,

$$\hat{P}(\mu, A|K) = \hat{P}(\mu, A|K^{CONV})$$

all $(\mu, A) \in \mathcal{D}$.

Proof. To show that $\hat{P}(\mu, A|K) \subset \hat{P}(\mu, A|K^{CONV})$, it suffices to show that any strategy $\lambda \in \hat{\Lambda}(\mu, A|K)$ is also optimal with the convexified cost function, $\lambda \in \hat{\Lambda}(\mu, A|K^{CONV})$, since then the corresponding data $\mathbf{P}_\lambda$ is in both sets. The first step is to show that the value function is no higher for the convexified function,

$$\hat{V}(\mu, A|K^{CONV}) \leq \hat{V}(\mu, A|K). \tag{41}$$

Consider an arbitrary strategy $\eta \in \Lambda(\mu, A)$ and index the finite set of possible posteriors $\bar{\eta}(n) \in \Gamma(\eta)$ for $1 \leq n \leq N$. By construction of the lower semi-continuous hull of T_μ (Rockafellar page 36),

$$T_\mu^{CONV}(\gamma) = \inf\{\sum_{m=1}^{M} \alpha(m)T(\gamma(m))| \sum_{m=1}^{M} \alpha(m)\gamma(m) = \gamma\}.$$

Hence we know that for each posterior $\bar{\eta}(n)$ and $p \in \mathbb{N}$ there exists a finite set of posteriors $\eta(n, m, p)$ for $1 \leq m \leq M(p)$ and corresponding weights $\alpha(n, m, p) > 0$ with $\sum_{m=1}^{M_n} \alpha(n, m, p) = 1$ such that.

$$\sum_{m=1}^{M(n,p)} \sum_{n=1}^{N} \alpha(n, m, p)\eta(n, m, p) = \bar{\eta}(n);$$

and such that the corresponding weighted average value of $T[\eta(n, m, p)]$ is no more than $\frac{1}{p}$ above $T_\mu^{CONV}[\bar{\eta}(n)]$, so that, for $1 \leq n \leq N$,

$$T_\mu^{CONV}[\bar{\eta}(n)] \geq \sum_{m=1}^{M_n} \alpha(n, m, p)T[\eta(n, m, p)] - \frac{1}{p}.$$

For each $p \in \mathbb{N}$ we introduce a corresponding strategy $F(\eta, p) = (Q_{F(\eta,p)}, q_{F(\eta,p)})$ with possible posteriors,

$$\Gamma(Q_{F(\eta,p)}) = \{\eta(n, m, p)|1 \leq n \leq N \text{ and } 1 \leq m \leq M(n, p)\}.$$

Specifically, the strategy is defined by:

$$\begin{aligned} Q_{F(\eta,p)}\left[\eta(n,m,p)\right] &= \alpha(n,m,p)Q_\eta(\bar{\eta}(n)); \\ q_{F(\eta,p)}\left[a|\eta(n,m,p)\right] &= q_\eta\left[a|\bar{\eta}(n)\right]. \end{aligned}$$

The first key observation is that this strategy is feasible, $F(\eta,p) \in \Lambda(\mu, A)$. It is immediate that $\mathcal{A}(F(\eta,p)) = \mathcal{A}(\eta) \subset A$. That $F(\eta,p) \in \Lambda(\mu,A)$ requires first that it is a strategy, which means that the probabilities over possible posteriors add to 1. This follows directly from the definition,

$$\begin{aligned} \sum_{\eta \in \Gamma(Q_{F(\eta,p)})} Q_{F(\eta,p)}\left[\eta(n,m,p)\right] &= \sum_{n=1}^{N}\sum_{m=1}^{M_n} \alpha(n,m,p)Q_\eta(\bar{\eta}(n)) \\ &= \sum_{n=1}^{N} Q_\eta(\bar{\eta}(n)) \sum_{m=1}^{M_n} \alpha(n,m,p) = \sum_{n=1}^{N} Q_\eta(\bar{\eta}(n)) = 1 \end{aligned}$$

To complete this part of the proof requires confirmation of Bayes' rule. This again is definitional,

$$\begin{aligned} \sum_{\eta \in \Gamma(Q_{F(\eta,p)})} \eta(n,m,p)Q_{F(\eta,p)}\left[\eta(n,m,p)\right] &= \sum_{n=1}^{N}\sum_{m=1}^{M_n} \alpha(n,m,p)\eta(n,m,p)Q_\eta(\bar{\eta}(n)) \\ &= \sum_{n=1}^{N} Q_\eta(\bar{\eta}(n)) \sum_{m=1}^{M_n} \alpha(n,m,p)\eta(n,m,p) \\ &= \sum_{n=1}^{N} \bar{\eta}(n) Q_\eta(\bar{\eta}(n)) = \mu. \end{aligned}$$

The second key observation is that $F(\eta,p)$ using K achieves utility net of attention costs within $\frac{1}{p}$ of that η achieves using K^{CONV}. To see this, consider first the expected prize utility as defined by the probability distribution over rewards,

$$U(F(\eta,p)) = \sum_{\gamma \in \Gamma(Q_{F(\eta,p)})} \sum_{a \in A} Q_{F(\eta,p)}(\gamma) q_{F(\eta,p)}(a|\gamma)\bar{u}(\gamma,a).$$

Defining the relevant set of indices,

$$\mathcal{I} \equiv \{(n,m,p)|1 \le n \le N \text{ and } 1 \le m \le M(n,p)$$

Note by direct substitution that,

$$\begin{aligned}
\sum_{\gamma\in\Gamma(Q_{F(\eta,p)})}\sum_{a\in A} Q_{F(\eta,p)}(\gamma)q_{F(\eta,p)}(a|\gamma)\bar{u}(\gamma,a) &= \sum_{\mathcal{I}}\sum_{a\in A}\alpha(n,m,p)Q_\eta(\bar{\eta}(n))q_{F(\eta,p)}\left[a|\eta(n,m,p)\right]\bar{u}(\bar{\eta}(n),a)\\
&= \sum_{\mathcal{I}}\sum_{a\in A}\alpha(n,m,p)Q_\eta(\bar{\eta}(n))q_\eta\left[a|\bar{\eta}(n)\right]\bar{u}(\bar{\eta}(n),a)\\
&= \sum_{1\le n\le N}\sum_{a\in A}\sum_{1\le m\le M(n,p)\}}\alpha(n,m,p)Q_\eta(\bar{\eta}(n))q_\eta\left[a|\bar{\eta}(n)\right]\bar{u}(\bar{\eta}(n),a)\\
&= \sum_{1\le n\le N}\sum_{a\in A}Q_\eta(\bar{\eta}(n))q_\eta\left[a|\bar{\eta}(n)\right]\bar{u}(\bar{\eta}(n),a) = U(\eta).
\end{aligned}$$

With regard to the costs, note by construction that,

$$T^{CONV}\left[\bar{\eta}(n)\right] \ge \sum_{m=1}^{M(n,p)}\alpha(n,m,p)T\left[\eta(n,m,p)\right]-\frac{1}{p}.$$

Hence,

$$\begin{aligned}
K^{CONV}(Q_\eta)+T^{CONV}\left[\mu\right] &= \sum_{n=1}^{N}Q_\eta(\bar{\eta}(n))T^{CONV}\left[\bar{\eta}(n)\right]\\
&\ge \sum_{n=1}^{N}\sum_{m=1}^{M_n}Q_\eta(\bar{\eta}(n))\alpha(n,m,p)T\left[\eta(n,m,p)\right]-\frac{1}{p}\\
&= \sum_{n=1}^{N}\sum_{m=1}^{M_n}Q_{F(\eta)}\left[\eta(n,m,p)\right]T\left[\eta(n,m,p)\right]-\frac{1}{p} = K(Q_{F(\eta,p)})-\frac{1}{p}.
\end{aligned}$$

Hence,

$$V(\mu,\eta|K^{CONV}) = U(\eta)-K^{CONV}(Q_\eta) \le U(F(\eta,p))-K(Q_{F(\eta,p)})+\frac{1}{p} = V(\mu,F(\eta,p)|K)+\frac{1}{p}.$$

By increasing p without bound, we establish that,

$$V(\mu,\eta|K^{CONV}) \le \sup_{\lambda\in\Lambda(\mu,A)}V(\mu,\lambda|K) = \hat{V}(\mu,A|K).$$

Since $\eta\in\Lambda(\mu,A)$ is arbitrary, this ensures that the supremum of all values is correspondingly bounded above,

$$\hat{V}(\mu,A|K^{CONV}) \equiv \sup_{\eta\in\Lambda(\mu,A)}V(\mu,\eta|K^{CONV}) \le \hat{V}(\mu,A|K),$$

completing the proof of (41).

To show that $\hat{P}(\mu,A|K)\subset\hat{P}(\mu,A|K^{CONV})$, note that since C has a CIR, we know that $\hat{\Lambda}(\mu,A|K)\neq\emptyset$ for any $(\mu,A)\in\mathcal{D}$. Now consider any optimal strategy $\lambda\in\hat{\Lambda}(\mu,A|K)$ which therefore achieves the value

$$\hat{V}(\mu,A|K) = V(\mu,\lambda|K) = U(\lambda)-K(\mu,Q_\lambda)$$

By definition of the convexification operation, note that $T_{\mu}^{CONV}(\gamma) \leq T_{\mu}\ (\gamma)$ all $\gamma \in \Gamma^{C}(\mu)$, so that,

$$K^{CONV}(\mu, Q_{\lambda}) \leq K(\mu, Q_{\lambda}).$$

Hence,

$$\begin{aligned}\hat{V}(\mu, A|K) &= U(\lambda) - K(\mu, Q_{\lambda}) \leq U(\lambda) - K^{CONV}(\mu, Q_{\lambda}) \leq V(\mu, A|K^{CONV}) \\ &\leq \hat{V}(\mu, A|K^{CONV}) \leq \hat{V}(\mu, A|K),\end{aligned}$$

making all equalities, so that indeed $\lambda \in \hat{\Lambda}(\mu, A|K^{CONV})$, completing the proof that $\hat{P}(\mu, A|K) \subset \hat{P}(\mu, A|K^{CONV})$.

To complete the proof, we establish the converse set inclusion. The first key observation is that $\hat{\Gamma}(\mu|K^{CONV}) \subset \hat{\Gamma}(\mu|K)$. To see this, first note that by definition, $\gamma' \notin \hat{\Gamma}(\mu|K)$ implies $T(\gamma') = \infty$. Moreover, because $\hat{\Gamma}(\mu|K)$ is convex by the definition of a generalized CIR, any weighted average of posteriors for which $\sum_{m=1}^{M} \alpha(m)\gamma(m) = \gamma'$ must involve at least one posterior $\gamma(m) \in \Gamma(\mu)/\hat{\Gamma}(\mu|K)$ for which $T(\gamma(m)) = \infty$. Hence $\sum_{m=1}^{M} \alpha(m)T(\gamma(m)) = \infty$. Thus,

$$T_{\mu}^{CONV}(\gamma') = \inf\{\sum_{m=1}^{M} \alpha(m)T(\gamma(m))| \sum_{m=1}^{M} \alpha(m)\gamma(m) = \gamma'\} = \infty,$$

where the weights $\alpha(m)$ are positive and sum to 1.

Finally, this implies that for any $Q \in \mathcal{Q}(\mu)$ with $\gamma' \in \Gamma(Q)$, the cost $K^{CONV}(\mu, Q) = \infty$, and so cannot be optimal (as the inattentive strategy will always provide a higher payoff). Hence $\gamma' \notin \hat{\Gamma}(\mu|K^{CONV})$.

Next, we show that, given any $\gamma' \in \hat{\Gamma}(\mu|K)$, it must be the case that $T^{CONV}\ [\gamma'] = T[\gamma']$. Assume not, then by definition of $\hat{\Gamma}(\mu|K)$ and K there exists $(\mu, A) \in \mathcal{D}$ and $\lambda = (q_{\lambda}, Q_{\lambda}) \in \hat{\Lambda}(\mu, A|K)$ such that $\gamma' \in \Gamma(Q)$. Note that, by construction $T^{CONV}\ [\gamma'] \leq T[\gamma']$, so assume by way of contradiction that $T^{CONV}\ [\gamma'] < T[\gamma']$. By definition of $T^{CONV}\ [\gamma']$ as the infimum of $\sum_{\xi \in \Gamma(\bar{Q})} T(\xi)\bar{Q}(\xi)$ on the set of posterior distributions that generate γ', this implies that there exists some alternative distribution of posteriors $\bar{Q}(\xi)$ that satisfies two conditions:

$$\begin{aligned}\sum_{\xi \in \Gamma(\bar{Q})} \xi\bar{Q}(\xi) &= \gamma'; \\ \sum_{\xi \in \Gamma(\bar{Q})} T(\xi)\bar{Q}(\xi) &< T[\gamma'].\end{aligned}$$

If such a distribution existed, one could amend strategy $\lambda = (q_{\lambda}, Q_{\lambda}) \in \hat{\Lambda}(\mu, A|K)$ and construct an alternative strategy $\lambda^{*} = (q^{*}, Q^{*}) \in \Lambda(\mu, A)$ that produced strictly higher net utility, contradicting the assumed optimality. One would simply reduce the probability of $\gamma' \in \Gamma(Q_{\lambda})$ by $Q_{\lambda}(\gamma')$ and increase the probability of each $\gamma \in \Gamma(\bar{Q})$ by $\gamma \in Q_{\lambda}(\gamma')\bar{Q}(\gamma)$, adjusting action choice probabilities appropriately to generate the same revealed posteriors as in λ. Effectively the new strategy uses $\sum_{\xi \in \Gamma(\bar{Q})} \xi\bar{Q}(\xi)$ rather than γ', but is otherwise identical. It is straightforward to show that $\lambda^{*} \in \Lambda(\mu, A)$, $U(\lambda) = U(\lambda^{*})$ but $K(\mu, Q^{*}) < K(\mu, Q_{\lambda})$.

To complete the proof of the lemma, we need to show that $\hat{P}(\mu, A|K^{CONV}) \subset \hat{P}(\mu, A|K)$. By construction, for any $P \in \hat{P}(\mu, A|K^{CONV})$, there exists an optimal strategy $\lambda = (q_\lambda, Q_\lambda) \in \hat{\Lambda}(\mu, A|K^{CONV})$ such that,

$$P = \mathbf{P}_\lambda.$$

First, note that, by the first claim above, for any $\gamma \in \Gamma(Q_\lambda)$, it must be the case that $\gamma \in \hat{\Gamma}(\mu|K)$, and so, by the second claim, $T^{CONV}[\gamma'] = T[\gamma']$. This directly implies that

$$\begin{aligned} K^{CONV}(\mu, Q_\lambda) &= \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) T^{CONV}(\gamma) - T^{CONV}(\mu) \\ &= \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) T(\gamma) - T(\mu) \\ &= K(\mu, \bar{Q}), \end{aligned}$$

and so $V(\mu, \lambda|K^{CONV}) = V(\mu, \lambda|K)$, and as $\hat{V}(\mu, A|K) \leq \hat{V}(\mu, A|K^{CONV}) = V(\mu, \lambda|K^{CONV})$, we have $\lambda \in \bar{\Lambda}(\mu, A|K)$ and so $\mathbf{P}_\lambda \in \hat{P}(\mu, A|K)$, completing the proof. ■

A2.5: Linking Strategies with Data

Lemma 2.12 (Strategies and Revealed Posteriors): Given $\lambda \in \hat{\Lambda}(\mu, A|K)$ for $K \in \mathcal{K}^{PS}$ and some $(\mu, A) \in \mathcal{D}$:

1. $\mathcal{A}(\mathbf{P}_\lambda) = \mathcal{A}(\lambda)$;
2. Given for $\gamma \in \Gamma(Q_\lambda)$ and a such that $q_\lambda(a|\gamma) > 0$

$$\bar{\gamma}^a_{\mathbf{P}_\lambda} = \gamma.$$

Proof. Given $\lambda = (Q_\lambda, q_\lambda) \in \Lambda(\mu, A)$ it is definitional that, for all $a \in \mathcal{A}(\lambda)$ and $\omega \in \Omega(\mu)$,

$$\mathbf{P}_\lambda(a|\omega) = \frac{\sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) q_\lambda(a|\gamma) \gamma(\omega)}{\mu(\omega)},$$

so that indeed $\mathcal{A}(\mathbf{P}_\lambda) = \mathcal{A}(\lambda)$. Note also, that, by Lemma 2.3 and the fact that $\lambda \in \hat{\Lambda}(\mu, A)$, for any, $\gamma \in \Gamma(Q_\lambda)$ and a such that $q_\lambda(a|\gamma) > 0$, $q_\lambda(a|\gamma') = 0$ for all other $\gamma' \in \Gamma(Q_\lambda)$. Thus, for any such a, and using the definition of the revealed posterior

$$\begin{aligned} \bar{\gamma}^a_{\mathbf{P}_\lambda}(\omega) &= \frac{\mu(\omega) \mathbf{P}_\lambda(a|\omega)}{\mathbf{P}_\lambda(a)} \\ &= \frac{\mu(\omega) \frac{Q_\lambda(\gamma) q_\lambda(a|\gamma) \gamma(\omega)}{\mu(\omega)}}{\sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) q_\lambda(a|\gamma)} \\ &= \gamma(\omega). \end{aligned}$$

■

Lemma 2.13 (Inverse Operation on Data): Given $P \in \mathcal{P}(\mu, A)$ for some $(\mu, A) \in \mathcal{D}$, $\mathcal{A}(\mathbf{Q}_P) = \mathcal{A}(P)$, and:

1. $\mathbf{P}_{\boldsymbol{\lambda}(P)} = P$.
2. $\gamma^a_{\boldsymbol{\lambda}(P)} = \bar{\gamma}^a_P$.

Proof. Given $P \in \mathcal{P}$, it is definitional that $\mathcal{A}(\mathbf{Q}_P) = \mathcal{A}(P)$ and that, given $a \in \mathcal{A}(P)$ and $\omega \in \Omega(\mu)$,

$$\mathbf{P}_{\boldsymbol{\lambda}(P)}(a|\omega) = \frac{\sum\limits_{\gamma \in \Gamma(\mathbf{Q}_P)} \mathbf{Q}_P(\gamma)\mathbf{q}_P(a|\gamma)\gamma(\omega)}{\mu(\omega)}.$$

It is also definitional that $\boldsymbol{\lambda}(P) = (\mathbf{Q}_P, \mathbf{q}_P)$ has support comprising the revealed posteriors $\Gamma(P) = \cup_{a \in \mathcal{A}(P)} \bar{\gamma}^a_P(\omega)$,

$$\bar{\gamma}^a_P(\omega) \equiv \frac{\mu(\omega)P(a|\omega)}{\sum\limits_{\nu \in \Omega(\mu)} \mu(\nu)P(a|\nu)} \equiv \frac{\mu(\omega)P(a|\omega)}{P(a)};$$

and that,

$$\begin{aligned} \mathbf{Q}_P(\gamma) &= \sum_{\{a \in \mathcal{A}(P) | \bar{\gamma}^a_P = \gamma\}} P(a); \\ \mathbf{q}_P(a|\gamma) &= \begin{cases} \frac{P(a)}{\mathbf{Q}_P(\gamma)} & \text{if } \bar{\gamma}^a_P = \gamma; \\ 0 & \text{if } \bar{\gamma}^a_P \neq \gamma. \end{cases} \end{aligned}$$

Substitution yields,

$$\begin{aligned} \mathbf{P}_{\boldsymbol{\lambda}(P)}(a|\omega) &= \frac{\sum\limits_{a \in \mathcal{A}(P)} \mathbf{Q}_P(\bar{\gamma}^a_P)\mathbf{q}_P(a|\bar{\gamma}^a_P)\bar{\gamma}^a_P(\omega)}{\mu(\omega)} \\ &= \frac{\sum\limits_{\bar{\gamma}^a_P \in \mathcal{A}(P)} P(a)\bar{\gamma}^a_P(\omega)}{\mu(\omega)} = \frac{\sum\limits_{\bar{\gamma}^a_P \in \mathcal{A}(P)} P(a)\left[\frac{\mu(\omega)P(a|\omega)}{P(a)}\right]}{\mu(\omega)} = P(a|\omega), \end{aligned}$$

completing the proof of (1).

For part (2), we apply part (2) in the Lemma 2.12 to conclude that $\gamma^a_{\boldsymbol{\lambda}(P)} = \bar{\gamma}^a_{\mathbf{P}_{\boldsymbol{\lambda}(P)}}$, whereupon the fact that $\mathbf{P}_{\boldsymbol{\lambda}(P)} = P$ establishes that,

$$\gamma^a_{\boldsymbol{\lambda}(P)} = \bar{\gamma}^a_P,$$

completing the proof. ∎

Lemma 2.14 (Optimal Strategies and Data): Given $\lambda \in \hat{\Lambda}(\mu, A|K)$ for $K \in \mathcal{K}^{PS}$ and some $(\mu, A) \in \mathcal{D}$:

1. $\Gamma(\mathbf{Q}_{\mathbf{P}_\lambda}) = \Gamma(Q_\lambda)$.
2. $\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma) = Q_\lambda(\gamma)$ all $\gamma \in \Gamma(\mathbf{Q}_{\mathbf{P}_\lambda})$.

3. $\mathbf{q}_{\mathbf{P}_\lambda}(a|\gamma) = q_\lambda(a|\gamma)$ all $a \in \mathcal{A}(\mathbf{P}_\lambda)$, $\gamma \in \Gamma(\mathbf{Q}_{\mathbf{P}_\lambda})$.
4. $\boldsymbol{\lambda}(\mathbf{P}_\lambda) = \lambda$.

Proof. To prove (1), note directly that the possible posteriors satisfy,

$$\Gamma(\mathbf{Q}_{\mathbf{P}_\lambda}) = \cup_{a \in \mathcal{A}(\mathbf{P}_\lambda)} \bar{\gamma}^a_{\mathbf{P}_\lambda} = \cup_{a \in \mathcal{A}(\lambda)} \gamma^a_\lambda = \Gamma(Q_\lambda),$$

where we use Lemma 2.12 in equating the posteriors between strategy and data. By Lemma 2.12 we know that for an action $a \in \mathcal{A}(\lambda)$ is chosen if and only if $\gamma = \gamma^a_\lambda$ for some $\gamma \in \Gamma(Q_\lambda)$, allowing us to compute choice probabilities in the data from the only strictly positive term of the summation,

$$\begin{aligned} \mathbf{P}_\lambda(a|\omega) &= \frac{Q_\lambda(\gamma^a_\lambda) q_\lambda(a|\gamma^a_\lambda) \gamma^a_\lambda(\omega)}{\mu(\omega)}; \\ \mathbf{P}_\lambda(a) &= Q_\lambda(\gamma^a_\lambda) q_\lambda(a|\gamma^a_\lambda). \end{aligned}$$

Substitution in the definitions for $\mathbf{Q}_{\mathbf{P}_\lambda}$ implies that, given $\gamma \in \Gamma(Q_\lambda)$ with $Q_\lambda(\gamma) > 0$ there exists $a \in \mathcal{A}(\lambda)$ such that $\gamma = \gamma^a_\lambda$ and that,

$$\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma) = \sum_{\{a \in \mathcal{A}(\lambda) | \gamma^a_\lambda = \gamma\}} \mathbf{P}_\lambda(a) = Q_\lambda(\gamma) \sum_{\{a \in \mathcal{A}(\lambda) | \gamma^a_\lambda = \gamma\}} q_\lambda(a|\gamma) = Q_\lambda(\gamma),$$

confirming (2). With regard to part (3), note that given any posterior $\gamma = \gamma^a_\lambda \in \Gamma(Q_\lambda)$ for $a \in \mathcal{A}(\lambda)$, and any action $b \in A$,

$$\mathbf{q}_{\mathbf{P}_\lambda}(b|\gamma) = q_\lambda(b|\gamma).$$

By Lemma 2.3, both are zero when $\gamma^b_\lambda \neq \gamma^a_\lambda = \gamma$. Applying the definitions and the just-established equality of $\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma)$ and $Q_\lambda(\gamma)$, we conclude that, with $\gamma^b_\lambda = \gamma^a_\lambda = \gamma \in \Gamma(Q_\lambda)$,

$$\begin{aligned} \mathbf{q}_{\mathbf{P}_\lambda}(b|\gamma) &= \frac{\mathbf{P}_\lambda(b)}{\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma)} = \frac{q_\lambda(b|\gamma^a_\lambda) Q_\lambda(\gamma^a_\lambda)}{Q_\lambda(\gamma^a_\lambda)} \\ &= q_\lambda(b|\gamma^a_\lambda) = q_\lambda(b|\gamma). \end{aligned}$$

rounding out the proof. Note that (2) and (3) together directly establish (4):

$$\boldsymbol{\lambda}(\mathbf{P}_\lambda) \equiv (\mathbf{Q}_{\mathbf{P}_\lambda}, \mathbf{q}_{\mathbf{P}_\lambda}) = (Q_\lambda, q_\lambda) = \lambda.$$

■

Lemma 2.15 (Data and Optimal Strategies): Given $C \in \mathcal{C}$ with a PS representation $K \in \mathcal{K}^{PS}$ and $P \in C(\mu, A)$ some $(\mu, A) \in \mathcal{D}$,

$$\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K).$$

Proof. By definition $P \in C(\mu, A)$ when C has a PS representation $K \in \mathcal{K}^{PS}$ implies that there exists $\lambda \in \hat{\Lambda}(\mu, A|K)$ such that,

$$P = \mathbf{P}_\lambda.$$

By part (2) of Lemma 2.14 we know that, since $\lambda \in \hat{\Lambda}(\mu, A|K)$, $\boldsymbol{\lambda}(\mathbf{P}_\lambda) \in \hat{\Lambda}(\mu, A|K)$, so that the just established fact that $\lambda = \boldsymbol{\lambda}(\mathbf{P}_\lambda)$ when $\lambda \in \hat{\Lambda}(\mu, A|K)$ implies that,

$$\lambda(P) = \boldsymbol{\lambda}(\mathbf{P}_\lambda) = \lambda \in \hat{\Lambda}(\mu, A|K),$$

completing the proof. ■

Lemma 2.16 (Identical Posteriors): Given $C \in \mathcal{C}$ with a PS representation $K \in \mathcal{K}^{PS}$,

$$\hat{\Gamma}(\mu|K) = \Gamma^C(\mu)$$

all $\mu \in \Gamma$.

Proof. First pick $\gamma \in \Gamma^C(\mu)$. By definition, this means that there exists $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$ with $\gamma \in \Gamma(P)$. Now consider strategy $\boldsymbol{\lambda}(P)$ and note by Lemma 2.15 that $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K)$ and by Lemma 2.13 that $\Gamma(\boldsymbol{\lambda}(P)) = \Gamma(P)$, so $\gamma \in \Gamma(\boldsymbol{\lambda}(P))$, so that definitionally $\gamma \in \hat{\Gamma}(\mu|K)$.

To go in the converse direction, pick $\gamma \in \hat{\Gamma}(\mu|K)$ and note therefore that there exists $(\mu, A) \in \mathcal{D}$ and $\lambda \in \hat{\Lambda}(\mu, A|K)$ with $Q_\lambda(\gamma) > 0$. Now consider data $\mathbf{P}_\lambda$ and note by definition of a PS representation that $\mathbf{P}_\lambda \in C(\mu, A)$. By Lemma 2.14 note that $\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma) = Q_\lambda(\gamma)$, so that definitionally $\gamma \in \Gamma^C(\mu)$. ■

Lemma 2.17 (Identical Posterior Distributions): Given $C \in \mathcal{C}$ with a PS representation $K \in \mathcal{K}^{PS}$,

$$\Delta(\hat{\Gamma}(\mu|K)) = \mathcal{Q}^C(\mu),$$

all $\mu \in \Gamma$.

Proof. First pick $Q \in \mathcal{Q}^C(\mu)$ and note definitionally that there exists $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$ s.t. $\mathbf{Q}_P = Q$. Note by Lemma 2.14 that $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K)$. Hence definitionally the corresponding posterior distribution satisfies $Q_P \in \Delta(\hat{\Gamma}(\mu|K)$.

To go in the converse direction, pick $Q \in \Delta(\hat{\Gamma}(\mu|K)$ and apply FIO to find $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$ such that $Q_\lambda(\gamma) = Q(\gamma)$ all $\gamma \in \Gamma(Q)$. Define $\mathbf{P}_\lambda \in \mathcal{P}(\mu, A)$ and note by definition of CIR that $\mathbf{P}_\lambda \in C(\mu, A)$. By Lemma 21.4 note that, since $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$,

$$\mathbf{Q}_{\mathbf{P}_\lambda}(\gamma) = Q_\lambda(\gamma) = Q(\gamma),$$

all $\gamma \in \Gamma(\mathbf{Q}_{\mathbf{P}_\lambda}) = \Gamma(Q_\lambda) = \Gamma(Q)$. Hence definitionally, $\mathbf{Q}_{\mathbf{P}_\lambda} \in \mathcal{Q}^C(\mu)$, completing the proof. ■

Lemma 2.18: (Mixtures and Data) Given $\bar{\mu} \in \Gamma$, if strategies $\lambda, \{\lambda(l)\}_{1 \le l \le L} \in \Lambda(\bar{\mu}, A)$ are such that $\lambda = \sum_{l=1}^{L} \alpha(l)\lambda(l)$ for probability weights $\{\alpha(l)\}_{1 \le l \le L}$, then,

$$\mathbf{P}_\lambda = \sum_{l=1}^{L} \alpha(l)\mathbf{P}_{\lambda(l)}.$$

Proof. Given $a \in \mathcal{A}(\lambda)$ and $\omega \in \Omega(\bar{\mu})$,

$$\mathbf{P}_\lambda(a|\omega) = \frac{\sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma) q_\lambda(a|\gamma)\gamma(\omega)}{\bar{\mu}(\omega)}.$$

By definition of the mixture strategy,

$$\begin{aligned} Q_\lambda(\gamma) &= \sum_l \alpha(l) Q_l(\gamma) \text{ all } \gamma \in \Gamma(Q_\lambda) = \cup_{\{l|\alpha(l)>0\}} \Gamma(Q_l); \\ q_\lambda(a|\gamma) &= \frac{\sum_l \alpha(l) q_l(a|\gamma) Q_l(\gamma)}{Q_\lambda(\gamma)} \text{all } \gamma \in \Gamma(Q_\lambda),\ a \in \mathcal{A}(\lambda) = \cup_{\{l|\alpha(l)>0\}} \mathcal{A}(\lambda(l)). \end{aligned}$$

Directly substitution establishes the result,

$$\begin{aligned} \mathbf{P}_\lambda(a|\omega) &= \frac{\sum_{\gamma \in \Gamma(Q_\lambda)} \sum_{l=1}^{L} \alpha(l) q_l(a|\gamma) Q_l(\gamma)\gamma(\omega)}{\bar{\mu}(\omega)} \\ &= \sum_{l=1}^{L} \alpha(l) \left[\frac{\sum_{\gamma \in \Gamma(Q_{\lambda(l)})} q_l(a|\gamma) Q_l(\gamma)\gamma(\omega)}{\bar{\mu}(\omega)} \right] = \sum_{l=1}^{L} \alpha(l) \mathbf{P}_{\lambda(l)}(a|\omega). \end{aligned}$$

■

Lemma 2.19: (Irrelevance of Impossible Payoffs) Given $C \in \mathcal{C}$ with CIR $K \in \mathcal{K}$, consider $a \in \mathcal{A}(P)$ and $a' \neq a \in \mathcal{A}$ with identical payoffs in possible states $\Omega(\mu)$ according to some μ and define A' to be A with a' replacing a:

$$\begin{aligned} u(a', \omega) &= u(a, \omega) \text{ all } \omega \in \Omega(\mu); \\ A' &= a' \cup A/a. \end{aligned}$$

Then,

$$C(\mu, A') = \left\{ P' \in \mathcal{P}(\mu, A') | \exists P \in C(\mu, A) \text{ s.t } P'(a'|\omega) = P(a|\omega) \text{ and } P'(b|\omega) = P(b|\omega) \text{ all } b \in A/a. \right\}$$

Proof. Given a CIR, $P \in C(\mu, A)$ implies $\boldsymbol{\lambda}_P \in \hat{\Lambda}(\mu, A|K)$. Given that there is no change in feasible payoffs in possible states from replacing a by a', we know that the corresponding strategy induced by P' is optimal for decision problem (μ, A'),

$$\boldsymbol{\lambda}_{P'} \in \hat{\Lambda}(\mu, A'|K).$$

Since this is a CIR, we know that the corresponding data is observed,

$$\mathbf{P}(\boldsymbol{\lambda}_{P'}) \in C(\mu, A').$$

Note by Lemma 2.12 that $\mathbf{P}(\boldsymbol{\lambda}_{P'}) = P'$, so that $P' \in C(\mu, A')$. The argument applies in both directions, establishing Lemma 2.19. ■

Appendix 3: Theorem 3

In this Appendix we prove theorem 3. We start by proving necessity of the axioms, then sufficiency.

Theorem 3: Data set $C \in \mathcal{C}$ has a PS representation if and only if it satisfies A2 through A8.

Proof. Necessity of A2 through A8 for a PS representation:

Necessity of NIAS (A2) and NIAC (A3) is immediate following Caplin and Martin (2015) and Caplin and Dean (2015) respectively. Necessity of Completeness (A4) follows from combining Lemma 2.7 (FIO) with Lemma 2.16 (Identical Posteriors) and 2.17 (Identical Posterior Distribution). First, FIO implies that $\tilde{\Gamma}(\mu) \subset \hat{\Gamma}(\mu|K)$. Given the Identical Posteriors Lemma, it is therefore immediate that $\tilde{\Gamma}(\mu) \subset \Gamma^C(\mu)$, establishing necessity of the first clause in the Completeness axiom. The Identical Posterior Lemma further establishes that $\hat{\Gamma}(\mu|K) = \Gamma^C(\mu)$, whereupon the assumed convexity of $\hat{\Gamma}(\mu|K)$ in a PS representation establishes the convexity clause of the Completeness axiom. Finally, the Identical Posterior Distribution Lemma establishes the third clause

$$\Delta(\Gamma^C(\mu)) = \Delta(\hat{\Gamma}(\mu|K)) = \mathcal{Q}^C(\mu),$$

completing the proof that A4 is necessary,

We now establish necessity of Separability (A5). Consider $(\mu, A(1)) \in \mathcal{D}$, and $P(1) \in C(\mu, A(1))$. Note by Lemma 2.15 that $\boldsymbol{\lambda}(P(1)) = (\mathbf{Q}_{P(1)}, \mathbf{q}_{P(1)}) \in \hat{\Lambda}(\mu, A_1|K)$ and by Lemma 2.13 that $\mathbf{P}_{\boldsymbol{\lambda}(P(1))} = P(1)$.

Now consider $Q_2 \in \mathcal{Q}^C(\mu)$ with $\Gamma(\mathbf{Q}_{P(1)}) \cap \Gamma(Q_2) \neq \emptyset$. By Lemma 2.17, $Q_2 \in \hat{Q}(\mu)$. Now apply Lemma 2.9 to conclude that there exists $(\mu, A_2) \in \mathcal{D}$ and $\lambda(2) = (Q_2, q_2) \in \hat{\Lambda}(\mu, A_2)$ with,

$$q_2(a|\gamma) = \mathbf{q}_{P(1)}(a|\gamma),$$

for $\gamma \in \Gamma(\mathbf{Q}_{P(1)}) \cap \Gamma(Q_2)$. Now apply the $\mathbf{P}$ operator to $\lambda(2)$ conclude that, noting that since this is a PS representation and $\lambda(2) = (Q_2, q_2) \in \hat{\Lambda}(\mu, A_2)$,

$$\mathbf{P}_{\boldsymbol{\lambda}(2)} \equiv P(2) \in C(\mu, A_2).$$

We can apply Lemma 2.14 to conclude that, since $\lambda(2) = (Q_2, q_2) \in \hat{\Lambda}(\mu, A_2)$ that,

$$\mathbf{q}_{P(2)}(a|\gamma) = q_2(a|\gamma).$$

Stringing these together we conclude that indeed,

$$\mathbf{q}_{P(1)}(a|\gamma) = \mathbf{q}_{P(2)}(a|\gamma),$$

all $\gamma \in \Gamma(\mathbf{Q}_{P(1)}) \cap \Gamma(Q_2)$, establishing necessity of separability, and with it the proof of necessity of A5.

To prove necessity of Non-linearity (A6), we consider $(\mu, A) \in \mathcal{D}$, $P \in C(\mu, A)$, and pick a_1, a_2, $a_3 \in A$ such that $\bar{\gamma}_P^{a_1} \neq \bar{\gamma}_P^{a_3}$ and such that,

$$\bar{\gamma}_P^{a_2} = \alpha\bar{\gamma}_P^{a_1} + (1-\alpha)\bar{\gamma}_P^{a_3},$$

for some $\alpha \in (0,1)$. By Lemma 2.15, we know that $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K)$. We know also that the corresponding revealed posteriors are identical to the posteriors used in the strategy,

$$\gamma^{a_i}_{\boldsymbol{\lambda}(P)} = \bar{\gamma}^{a_i}_P \text{ for } 1 \leq i \leq 3.$$

Thus, by the Lagrangian Lemma there exists $\theta \in \mathbb{R}^{J-1}$ such that:

$$\begin{aligned}
\sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_2}_P(j) + \bar{u}(\bar{\gamma}^{a_2}_P, a_2) - T_\mu(\bar{\gamma}^{a_2}_P) &= \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_1}_P(j) + \bar{u}(\bar{\gamma}^{a_1}_P, a_1) - T_\mu(\bar{\gamma}^{a_1}_P) \\
&= \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_3}_P(j) + \bar{u}(\bar{\gamma}^{a_3}_P, a_3) - T_\mu(\bar{\gamma}^{a_3}_P).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_2}_P(j) + \bar{u}(\bar{\gamma}^{a_2}_P, a_2) - T_\mu(\bar{\gamma}^{a_2}_P) \\
= \ &\alpha\left(\sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_1}_P(j) + \bar{u}(\bar{\gamma}^{a_1}_P, a_1) - T_\mu(\bar{\gamma}^{a_1}_P)\right) + (1-\alpha)\left(\sum_{j=1}^{J-1} \theta(j)\bar{\gamma}^{a_3}_P(j) + \bar{u}(\bar{\gamma}^{a_3}_P, a_3) - T_\mu(\bar{\gamma}^{a_3}_P)\right).
\end{aligned}$$

Rearrangement yields,

$$\begin{aligned}
&\sum_{j=1}^{J-1} \theta(j)\left[\bar{\gamma}^{a_2}_P(j) - \alpha\bar{\gamma}^{a_1}_P(j) + (1-\alpha)\bar{\gamma}^{a_3}_P(j)\right] + \bar{u}(\bar{\gamma}^{a_2}_P, a_2) - \left(\alpha\bar{u}(\bar{\gamma}^{a_1}_P, a_1) + (1-\alpha)\bar{u}(\bar{\gamma}^{a_3}_P, a_3)\right) \\
= \ &T_\mu(\bar{\gamma}^{a_2}_P) - \left(\alpha T_\mu(\bar{\gamma}^{a_1}_P) + (1-\alpha)T_\mu(\bar{\gamma}^{a_3}_P)\right).
\end{aligned}$$

Since $\bar{\gamma}^{a_2}_P = \alpha\bar{\gamma}^{a_1}_P + (1-\alpha)\bar{\gamma}^{a3}_P$ the first term is equal to zero, and if $\bar{u}(\bar{\gamma}^{a_2}_P, a_2) = \alpha\bar{u}(\bar{\gamma}^{a_1}_P, a_1) + (1-\alpha)\bar{u}(\bar{\gamma}^{a_3}_P, a_3)$, the second term would also be zero. Hence,

$$T_\mu(\bar{\gamma}^{a_2}_P) = \alpha T_\mu(\bar{\gamma}^{a_1}_P) + (1-\alpha)T_\mu(\bar{\gamma}^{a_3}_P),$$

in contradiction of strict convexity of T_μ. This establishes necessity of A6.

To establish necessity of Convexity (A7), consider $(\mu, A) \in \mathcal{D}$, $P_l \in C(\mu, A)$ for $1 \leq l \leq L$, and probability weights $\alpha(l) > 0$. Define the mixture data $P_\alpha \in P(\mu, A)$ by,

$$P_\alpha(a|\omega) \equiv \sum_{l=1}^{L} \alpha(l)P_l(a|\omega).$$

Convexity requires that $P_\alpha \in C(\mu, A)$. Note that since the C is a CIR, there exists $\lambda(l) = (Q_l, q_l) \in \hat{\Lambda}(\mu, A|K)$ for which,

$$\mathbf{P}_{\lambda(l)} = P_l \text{ all } l\ .$$

Define the strategy $\lambda(\alpha)$ to be the corresponding mixture strategy, as defined in Appendix 2. By the Mixing and Optimality Lemma (2.2), $\lambda(\alpha) \in \hat{\Lambda}(\mu, A|K)$, and since this is a CIR,

$$\mathbf{P}_{\lambda(\alpha)} \in C(\mu, A).$$

To complete the proof, we apply the Mixtures and Data Lemma (2.18) to confirm that $\lambda(\alpha)$ correspondingly mixes the SDSC data: given $a \in A$ and $\omega \in \Omega(\mu)$,

$$\mathbf{P}_{\lambda(\alpha)}(a|\omega) = \sum_{l=1}^{L} \alpha(l)\mathbf{P}_{\lambda(l)} = \sum_{l=1}^{L} \alpha(l)P_l a|\omega) = P_\alpha(a|\omega) \in C(\mu, A)$$

We conclude the necessity proof by establishing necessity of Continuity (A8). To this end, we consider $\mu \in \Gamma$ and $K \in \mathcal{K}^{PS}$ together with $I \geq 1$ sequences of actions $a^i(m)$ with $\lim_{m\to\infty} a^i(m) = \bar{a}^i$ for $1 \leq i \leq I$, with $A(m) = \cup_{i=1}^{I} a^i(m)$ and $\bar{A} = \cup_{i=1}^{I} \bar{a}^i$. Suppose that there exists $P \in \cap_{m=1}^{\infty} C(\mu, A(m))$ with $\mathcal{A}(P) \subset \bar{A}$. Then by Lemma 2.8, the revealed attention strategy satisfies $\boldsymbol{\lambda}(P) \in \cap_{m=1}^{\infty} \hat{\Lambda}(\mu, A(m)|K)$ and since $\mathcal{A}(P) \subset \bar{A}$, it is also feasible in the limit problem, $\boldsymbol{\lambda}(P) \in \Lambda(\mu, \bar{A})$. What we need to prove is that it is optimal.

Since $\boldsymbol{\lambda}(P) = (\mathbf{Q}_P, \mathbf{q}_P) \in \hat{\Lambda}(\bar{\mu}, A(m)|K)$, we know that it achieves optimal value, which we denote $\hat{V}$,

$$V(\boldsymbol{\lambda}_P) = U(\boldsymbol{\lambda}_P) - K_{\bar{\mu}}(\bar{\mu}, \mathbf{Q}_P) = \hat{V}(\bar{\mu}, A(m)) \equiv \hat{V}.$$

Since $\boldsymbol{\lambda}_P \in \Lambda(\bar{\mu}, \bar{A})$, we know that $\hat{V}(\bar{\mu}, \bar{A}|K) \geq \hat{V}$. Strictness of this inequality is not possible. To see this, consider a purported strategy $\bar{\lambda}' = (\bar{Q}', \bar{q}') \in \Lambda\left(\bar{\mu}, \bar{A}\right)$ that achieves higher value in the limit problem,

$$V(\bar{\lambda}', \bar{\mu}, \bar{A}|K) > \hat{V}.$$

We now construct the corresponding strategy $\lambda'_m = (Q'_m, q'_m) \in \Lambda\left(\bar{\mu}, A(m)\right)$ that uses precisely the same posterior distribution and the correspondingly indexed action choices,

$$Q'_m(\gamma) = \bar{Q}'(\gamma) \text{ and } q'_m(a^i(m)|\gamma^i) = \bar{q}'(\bar{a}^i|\gamma^i).$$

Since $\lim_{m\to\infty} a^i(m) = \bar{a}^i$, we know that

$$\lim_{m\to\infty} U(\lambda'_m) = U(\bar{\lambda}'),$$

hence, as costs depend only on the unchanging distribution over posteriors, the corresponding holds for the valuation,

$$\lim_{m\to\infty} V(\lambda'_m, \mu, A(m)|K) = V(\bar{\lambda}', \mu, \bar{A}|K) > \hat{V},$$

contradicting $\boldsymbol{\lambda}_P \in \hat{\Lambda}(\bar{\mu}, A(m)|K)$. We conclude that $\boldsymbol{\lambda}_P$ is optimal in the limit problem, $\boldsymbol{\lambda}_P \in \hat{\Lambda}(\bar{\mu}, \bar{A}|K)$. Hence, given that this is a CIR, $\mathbf{P}_{\boldsymbol{\lambda}_P} \in C(\bar{\mu}, \bar{A})$. That $\mathbf{P}_{\boldsymbol{\lambda}_P} = P$ follows from Lemma 2.13, completing the proof that $P \in C(\bar{\mu}, \bar{A})$. ■

Proof. Sufficiency of A2 through A8 for a PS representation:

There are three key steps in the sufficiency proof. In the first, we invoke corollary 1 to theorem 2 as established in Appendix 1 which gives a cost function of the PS form, yet in which the distribution of posteriors impacts the computed cost of each posterior. The first point that we establish is that A5, Separability allows us to remove the dependence of the function $T_{\mu}^{C}(\cdot, \bar{Q})$ on the particular distribution of posteriors. With this we will know that, given $\mu \in \Gamma$, A2 through A5 imply that there exists Generalized PS cost function $\bar{K} \in \mathcal{K}^{GPS}$ (as defined prior to the proof of Lemma 2.11 in Appendix 2) such that $C(\mu, A) \subset \hat{P}(\mu, A|\bar{K})$ all $(\mu, A) \in \mathcal{D}$ and such that, given $(\mu, Q) \in \mathcal{F}$ with

$Q \in \mathcal{Q}^C(\mu)$,

$$\bar{K}(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)\, \bar{T}_\mu(\gamma) - \bar{T}_\mu(\mu). \tag{42}$$

for some $\bar{T}_\mu : \Gamma(\mu) \to \bar{\mathbb{R}}$ (with $\bar{T}_\mu$ real-valued on $\Gamma^C(\mu)$). The second step in the proof applies Lemma 2.11 to show that $\bar{T}_\mu$ can be assumed weakly convex without any loss of generality, and that with A6 (Non-linearity), it must be strictly rather than weakly convex. The final step shows that addition of Axiom A7 (Convexity) and A8 (Continuity) allow us to generate all data, $C(\mu, A) = \hat{P}(\mu, A|\bar{K})$. In this final step, a key role is played by Lemmas that are established in Appendix 2, which is to be expected given that the cost function at this stage is of precisely the PS form and the only remaining question relates to ensuring that all optima are observed in the data.

As noted, in the first step of the proof, we invoke corollary 1 to theorem 2 which we state as follows for current purposes. Given $C \in \mathcal{C}$ satisfying A2-A4, there exists a unique function $K \in \mathcal{K}$ such that, given $\mu \in \Gamma$, $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$ where $K \in \mathcal{K}$ can be computed for $(\mu, \bar{Q}) \in \mathcal{F}$ with $\bar{Q} \in \mathcal{Q}^C(\mu)$ by enumerating the support $\Gamma(\bar{Q}) = \{\bar{\gamma}^n | 1 \leq n \leq N\}$ and using the definitions of $T^C_\mu(\bar{\gamma}^n, \bar{Q})$ and $T^C_\mu(\mu, \bar{Q})$ in the proof of theorem 2

$$T^C_\mu(\bar{\gamma}^n, \bar{Q}) \equiv [\bar{\gamma}^n - \mu] \cdot \int_0^1 u(\bar{a}^n_t)dt.$$

and computing,

$$K(\mu, \bar{Q}) \equiv \sum_n \bar{Q}(\bar{\gamma}^n) T^C_\mu(\bar{\gamma}^n, \bar{Q}) - T^C_\mu(\mu, \bar{Q}). \tag{43}$$

In this stage of the proof we show that A5 enables us to remove the dependence on $\bar{Q}$ and find $\bar{T}^C_\mu$ in the data such that we can set,

$$T^C_\mu(\gamma, Q) = \bar{T}^C_\mu(\gamma),$$

all $Q \in \mathcal{Q}^C(\mu)$.

We set up our candidate function $\bar{T}^C_\mu$ in two steps. In the first step, we select $\bar{Q} \in \mathcal{Q}^C(\mu)$ with $\Gamma(\bar{Q}) = \cup_{1 \leq n \leq N} \bar{\gamma}^n$ with $N = |\Gamma(\bar{Q})|$ with the $\{\bar{\gamma}^n\}_{n=1}^N$ being affine independent vectors in $\mathbb{R}^N$ thereby forming a basis for $\Gamma(\mu)$: that this is possible follows from Completeness, whereby the full-dimensional interior posteriors is observed, $\tilde{\Gamma} \subset \Gamma^C(\mu)$. We then apply corollary 1 to establish existence of $T^C_\mu(\bar{\gamma}^n, \bar{Q})$ such that,

$$K(\mu, \bar{Q}) = \sum_{n=1}^N \bar{Q}(\gamma^n) T^C_\mu(\gamma^n, \bar{Q}),$$

where,

$$T^C_\mu(\gamma^n, \bar{Q}) = [\gamma^n - \mu] \cdot \int_0^1 u(\bar{a}^n_t)dt \tag{44}$$

and the $\{\bar{a}^n_t\}_{n=1}^N$ for $t \in [0, 1]$ are the actions that support this construction.

We cost all other posteriors by embedding them in decision problems that include this fixed basis $\Gamma(\bar{Q})$. Picking any other observed posterior $\eta \in \Gamma^C(\mu) \backslash \Gamma(\bar{Q})$, we identify a corresponding distribution $\bar{Q}^\eta$ with full support

$$\Gamma(\bar{Q}^\eta) = \Gamma(\bar{Q}) \cup \eta,$$

that satisfies the Bayesian constraint,

$$\sum_{\gamma \in \Gamma(\bar{Q}^{\eta})} \gamma \bar{Q}^{\eta}(\gamma) = \mu.$$

That this is possible follows from the fact that $\Gamma(\bar{Q})$ forms a basis for $\Gamma(\mu)$ so that an arbitrarily small probability added to η can be precisely off-set by corresponding reductions in the weights on the basis vectors $\bar{\gamma}^n \in \Gamma(\bar{Q})$ while retaining strict positivity (see Lemma 4.5 for details).

Let $\bar{A}_t = \{\bar{a}_t^n\}_{n=1}^N$ and let $(\mu, \bar{A}_t) \in \mathcal{D}$ and $\bar{P}_t \in C\ (\mu, \bar{A}_t)$ such that $\mathbf{Q}_{\bar{P}_t} = \bar{Q}_t$ as in the proof to theorem 2. Let,

$$\eta_t = t\eta + (1-t)\mu$$

on $t \in [0,1]$ and let $\bar{Q}_t(\bar{\gamma}_t^n) = \bar{Q}(\bar{\gamma}^n) \equiv \bar{Q}^n$. By construction $\bar{Q}_t^{\eta} \in \mathcal{Q}^C(\mu)$ all $t \in [0,1]$.

We can now apply the Separability axiom, A5. Since $(\mu, \bar{A}_t) \in \mathcal{D}$, $\bar{P}_t \in C(\mu, \bar{A}_t)$ with $\mathbf{Q}_{\bar{P}_t} = \bar{Q}_t$, and $\bar{Q}_t^{\eta} \in \mathcal{Q}^C(\mu)$ satisfies $\Gamma(\bar{Q}_t^{\eta}) \cap \Gamma(\bar{Q}_t) = \Gamma(\bar{Q}_t)$, the axiom asserts existence for each $t \in [0,1]$ of $A_t(\eta) \subset \mathcal{A}$ and $P_t(\eta) \in C(\mu, A_t(\eta))$ with $\mathbf{Q}_{P_t(\eta)} = \bar{Q}_t^{\eta}$ such that for each $\bar{\gamma}_t^n \in \Gamma(\bar{Q}_t)$ there exists an action $a \in \bar{A}_t \cap A_t(\eta)$ with $\bar{\gamma}^a_{P_t(\eta)} = \bar{\gamma}^a_{\bar{P}_t} = \bar{\gamma}_t^n$. We identify just such a choice set and data combination, with $P_t(\eta) \in C(\mu, A_t(\eta))$, and define $a_t(\eta) \in A_t(\eta)$ as the action associated with the new revealed posterior,

$$\eta_t = \bar{\gamma}^{a_t(\eta)}_{P_t(\eta)}$$

According to the prescription in theorem 2 we can now compute the cost function from this data using the posterior-by-posterior approach as,

$$K(\mu, \bar{Q}^{\eta}) = \sum_{\gamma \in \Gamma(\bar{Q}) \cup \eta}^{N} \bar{Q}^{\eta}(\gamma) T_{\mu}^{C}(\gamma, \bar{Q}^{\eta}),$$

where

$$T_{\mu}^{C}(\gamma, \bar{Q}^{\eta}) = [\gamma - \mu] \cdot \int_0^1 u(\bar{a}_t^n) dt. \tag{45}$$

for each $\gamma \in \Gamma(\bar{Q})$, and

$$T_{\mu}^{C}(\gamma, \bar{Q}^{\eta}) = [\gamma - \mu] \cdot \int_0^1 u(a_t(\gamma)) dt. \tag{46}$$

for $\gamma = \eta$. Note that since the $\bar{a}_t^n$ are the same in (45) and (44), $T_{\mu}^{C}(\gamma, \bar{Q}^{\eta}) = T_{\mu}^{C}(\gamma, \bar{Q})$ for all $\gamma \in \Gamma(\bar{Q})$.

We now define our candidate cost function $\bar{T}_{\mu}^{C}(\gamma)$. Specifically, we repeat the above for all $\eta \in \Gamma(Q) \backslash \Gamma(\bar{Q})$ and set

$$\bar{T}_{\mu}(\gamma) = \begin{cases} T_{\mu}^{C}(\gamma, \bar{Q}) & \text{for } \gamma \in \Gamma(\bar{Q}); \\ T_{\mu}^{C}(\gamma, \bar{Q}^{\gamma}) & \text{for } \gamma \in \Gamma^{C}(\mu) \backslash \Gamma(\bar{Q}). \end{cases} \tag{47}$$

The claim is that, for any $Q \in \mathcal{Q}^C(\mu)$,

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \bar{T}_{\mu}(\gamma) \tag{48}$$

We establish (48) first for all $Q \in \mathcal{Q}^C(\mu)$ such that $\Gamma(\bar{Q}) \subset \Gamma(Q)$. The proof is inductive on cardinality. We first establish that the result holds for any $Q \in \mathcal{Q}^C(\mu)$ with cardinality $N+1$ and $\Gamma(\bar{Q}) \subset \Gamma(Q)$. In this case, there is a unique $\eta \in \Gamma(Q)\backslash\Gamma(\bar{Q})$ and $\Gamma(Q^\eta) = \Gamma(\bar{Q}^\eta)$ where $\bar{Q}^\eta$ is the distribution used the construction of (47). Since $\Gamma(Q^\eta) = \Gamma(\bar{Q}^\eta)$, the Separability axiom says that we can use the same acts to calculate $T_\mu^C(\gamma, Q^\eta)$ as we used to calculate $T_\mu^C(\gamma, \bar{Q}^\eta)$, and since the construction of $T_\mu^C(\gamma, \bar{Q}^\eta)$ in (45) and (46) depends on $\bar{Q}^\eta$ only through the actions, it follows that

$$T_\mu^C(\gamma, Q^\eta) = \bar{T}_\mu(\gamma)$$

for all $\gamma \in \Gamma(Q^\eta)$ and

$$K(\mu, Q) = \sum_{n=1}^{N} Q(\gamma^n)\bar{T}_\mu(\gamma^n) + Q(\eta)\bar{T}_\mu(\eta) \tag{49}$$

as required.

We now suppose that (48) holds for all $Q \in \mathcal{Q}^C(\mu)$ with $\Gamma(\bar{Q}) \subset \Gamma(Q)$ and cardinality $N+m$ and for $m \geq 1$ and show that this extends to $N+m+1$. Consider $\hat{Q} \in \mathcal{Q}^C$ such that $\Gamma(\bar{Q}) \subset \Gamma(\hat{Q})$ and $|\Gamma(\hat{Q})| = N+m+1$. Since $|\Gamma(\hat{Q})| > |\Gamma(\bar{Q})| + 1$, we can find η_1,$\eta_2 \in \Gamma(\hat{Q})\backslash\Gamma(\bar{Q})$. Note that by assumption (48) holds for all $Q_1 \in \mathcal{Q}^C(\mu)$ such that $\Gamma(Q_1) = \Gamma(\hat{Q})\backslash\eta_1$ and some path of actions $A_{\Gamma(\hat{Q})\backslash\eta_1}(t)$ for $t \in [0,1]$. By Separability (A5) we can find, for any $t \in [0,1]$ a corresponding set of action paths $\hat{A}_1(t) = \{A_{\Gamma(\hat{Q})\backslash\eta_1}(t), \hat{a}_{\eta_1}(t)\}$ such that,

$$\begin{aligned} K(\hat{Q}) &= \sum_{\gamma\in\Gamma(\hat{Q})} \hat{Q}(\gamma)T_\mu^C(\gamma, \hat{Q}) \\ &= \sum_{\gamma\in\Gamma(\hat{Q})\backslash\eta_1} \hat{Q}(\gamma)\bar{T}_\mu(\gamma) + \hat{Q}(\eta_1)T_\mu^C(\eta_1, \hat{Q}) \end{aligned}$$

where,

$$T^C(\eta_1, \hat{Q}) = [\eta_1 - \mu] \cdot \int_0^1 \hat{a}_{\eta_1}(t)dt.$$

Similarly, we can find $Q_2 \in \mathcal{Q}^C(\mu)$ such that $\Gamma(Q_2) = \Gamma(\hat{Q})\backslash\eta_2$ and path $\hat{A}_2(t) = \{A_{\Gamma(\hat{Q})\backslash\eta_2}(t), \hat{a}_{\eta_2}(t)\}$ defined by the action paths associated with $\bar{T}(\gamma)$ for $\gamma \neq \eta_2$ such that,

$$K(\hat{Q}) = \sum_{\gamma\in\Gamma(\hat{Q})\backslash\eta_2} \hat{Q}(\gamma)\bar{T}_\mu(\gamma) + \hat{Q}(\eta_2)T_\mu^C(\eta_2, \hat{Q})$$

with

$$T^C(\eta_2, \hat{Q}) = [\eta_2 - \mu] \cdot \int_0^1 \hat{a}_{\eta_2}(t)dt.$$

Comparing the two different expressions for precisely the same cost, we conclude that,

$$\hat{Q}(\eta_1)\bar{T}_\mu(\eta_1) + \hat{Q}(\eta_2)T_\mu^C(\eta_2, \hat{Q}) = \hat{Q}(\eta_1)T_\mu^C(\eta_1, \hat{Q}) + \hat{Q}(\eta_2)\bar{T}_\mu(\eta_2),$$

or,

$$\hat{Q}(\eta_1)\left[\bar{T}_\mu(\eta_1) - T_\mu^C(\eta_1, \hat{Q})\right] = \hat{Q}(\eta_2)\left[\bar{T}_\mu(\eta_2) - T_\mu^C(\eta_2, \hat{Q})\right]. \tag{50}$$

Since $|\Gamma(\hat{Q})| > |\Omega(\mu)|$, and contains a basis, so $\Gamma(\bar{Q}) \subset \Gamma(\hat{Q})$, we can find a distinct distribution of posteriors $\hat{Q}' \in \mathcal{Q}^C$ with $\Gamma(\hat{Q}') = \Gamma(\hat{Q})$, $\hat{Q}'(\eta_1) \neq \hat{Q}(\eta_1)$ yet $\hat{Q}'(\eta_2) = \hat{Q}(\eta_2)$ and run corresponding logic to conclude that,

$$\hat{Q}'(\eta_1)\left[\bar{T}_\mu(\eta_1) - T_\mu^C(\eta_1, \hat{Q}')\right] = \hat{Q}'(\eta_2)\left[\bar{T}_\mu(\eta_2) - T_\mu^C(\eta_2, \hat{Q}')\right]. \tag{51}$$

Since $\Gamma(\hat{Q}') = \Gamma(\hat{Q})$, direct application of Separability to the case in which $\Gamma(Q_{P(1)}) \cap \Gamma(Q_2)$ means that we may use the same action set $\hat{A}_1(t) = \{A_{\Gamma(\hat{Q})\backslash\eta_1}(t), \hat{a}_{\eta_1}(t)\}$ to calculate both $T_\mu^C(\eta_1, \hat{Q}')$ and $T_\mu^C(\eta_1, \hat{Q})$. Similarly for η_2. Hence $T_\mu^C(\eta_1, \hat{Q}') = T_\mu^C(\eta_1, \hat{Q})$ and $T_\mu^C(\eta_2, \hat{Q}') = T_\mu^C(\eta_2, \hat{Q})$. Since $\hat{Q}'(\eta_2) = \hat{Q}(\eta_2)$ by assumption, subtracting (50) from (51) yields,

$$\left[\hat{Q}'(\eta_1) - \hat{Q}(\eta_1)\right]\left[\bar{T}_\mu(\eta_1) - T_\mu^C(\eta_1, \hat{Q})\right] = 0.$$

Since $\hat{Q}'(\eta_1) \neq \hat{Q}(\eta_1)$, it follows that

$$T_\mu^C(\eta_1, \hat{Q}) = \bar{T}_\mu(\eta_1),$$

Since η_1 is arbitrary, this establishes the induction step.

We now consider arbitrary $Q \in \mathcal{Q}^C(\mu)$. In particular, we consider Q such that $\Gamma(\bar{Q}) \not\subset \Gamma(Q)$. The preceding establishes that (48) holds for $Q' = Q \cup \bar{Q}$. Separability (A5) ensures that can use the same actions for Q to finally produce the representation of the desired form,

$$K(Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)\bar{T}_\mu(\gamma)$$

for all $Q \in \mathcal{Q}^C$, completing the proof that there exists $K \in \mathcal{K}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$ and such that,

$$K(\mu, Q) = \sum_{\Gamma(Q)} Q(\gamma)\,\bar{T}_\mu^C(\gamma).$$

This completes the first step of the proof.

At this stage we set $\bar{T}_\mu^C(\gamma)$ to infinity outside the convex set $\Gamma^C(\mu)$. Note that the construction of K as forming a CIR of the data assumes that subjects use the revealed attention strategy in each decision problem, meaning both that increasing the cost of posteriors outside $\Gamma^C(\mu)$ maintains the CIR and that $\Gamma^C(\mu) \subset \hat{\Gamma}(\mu|K)$. Setting $K(\mu, \gamma) = \infty$ for $\gamma \notin \Gamma^C(\mu)$ therefore makes $\Gamma^C(\mu) = \hat{\Gamma}(\mu|K)$ and so both are convex, by Completeness (A4). This means that $K \in \mathcal{K}^{GPS}$. Direct application of Lemma 2.11 shows that we can replace T_μ with its convexification, T_μ^{CONV}, then define K^{CONV} correspondingly, with assurance that the data is unchanged,

$$\hat{P}(\mu, A|K) = \hat{P}(\mu, A|K^{CONV}).$$

This then implies that,

$$C(\mu, A) \subset \hat{P}(\mu, A|K) \Longrightarrow C(\mu, A) \subset \hat{P}(\mu, A|K^{CONV}).$$

To complete the second step we show that, since $C \in \mathcal{C}$ satisfies A3 and A6, and there exists a convex function $T_\mu : \Gamma(\mu) \to \bar{\mathbb{R}}$ such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$, where,

$$K(\mu, Q) = \sum_{\Gamma(Q)} Q(\gamma)\, T_\mu(\gamma),$$

then T_μ is strictly convex.

Assume by way of contradiction that T_μ is convex but not strictly so. Then there exists γ_1, $\gamma_3 \in dom\ T_\mu$ such that $\gamma_2 = \alpha\gamma_1 + (1-\alpha)\gamma_3$ and $T_\mu(\gamma_2) = \alpha T_\mu(\gamma_1) + (1-\alpha)T_\mu(\gamma_3)$. By construction above, $T_\mu(\gamma)$ is real valued only on $\Gamma^C(\mu)$ and so by Axiom A4 (Completeness) we can conclude both that $\gamma_2 \in \Gamma^C(\mu)$ and that there exists a decision problem (μ, A) such that, for some a_1, a_2, $a_3 \in A$, we have $\bar{\gamma}_P^{a_1} = \gamma_1$, $\bar{\gamma}_P^{a_2} = \gamma_2$ and $\bar{\gamma}_P^{a_3} = \gamma_3$. By the Lagrangian Lemma, this implies that there is a strategy $\lambda \in \bar{\Lambda}(\mu, A)$ and corresponding multipliers $\theta \in \mathbb{R}^{J-1}$ such that, since $\gamma_1, \gamma_2, \gamma_3 \in \Gamma(\lambda)$,

$$\begin{aligned}
\sum_{j=1}^{J-1} \theta(j)\gamma_1(j) + \bar{u}(\gamma_1, a_1) - T_\mu(\gamma_1) &= \sum_{j=1}^{J-1} \theta(j)\gamma_2(j) + \bar{u}(\gamma_2, a_2) - T_\mu(\gamma_2);\\
\sum_{j=1}^{J-1} \theta(j)\gamma_3(j) + \bar{u}(\gamma_3, a_3) - T_\mu(\gamma_3) &= \sum_{j=1}^{J-1} \theta(j)\gamma_2(j) + \bar{u}(\gamma_2, a_2) - T_\mu(\gamma_2).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\sum_{j=1}^{J-1} \theta(j)\gamma_2(j) + \bar{u}(\gamma_2, a_2) - T_\mu(\gamma_2)\\
= \;& \alpha \left(\sum_{j=1}^{J-1} \theta(j)\gamma_1(j) + \bar{u}(\gamma_1, a_1) - T_\mu(\gamma_1) \right) + (1-\alpha) \left(\sum_{j=1}^{J-1} \theta(j)\gamma_3(j) + \bar{u}(\gamma_3, a_3) - T_\mu(\gamma_3) \right).
\end{aligned}$$

Equivalently,

$$\begin{aligned}
&\sum_{j=1}^{J-1} \theta(j)\left(\gamma_2(j) - (\alpha\gamma_1(j) + (1-\alpha)\gamma_3(j))\right) + \bar{u}(\gamma_2, a_2) - (\alpha\bar{u}(\gamma_1, a_1) + (1-\alpha)\bar{u}(\gamma_3, a_3))\\
= \;& T_\mu(\gamma_2) - (\alpha T_\mu(\gamma_1) + (1-\alpha)T_\mu(\gamma_3)).
\end{aligned}$$

Since $\gamma_2 = \alpha\gamma_1 + (1-\alpha)\gamma_3$ the first term is equal to zero, and by assumption $T_\mu(\gamma_2) = \alpha T_\mu(\gamma_1) + (1-\alpha)T_\mu(\gamma_3)$, and so the RHS equals zero,

$$\bar{u}(\bar{\gamma}_P^{a_2}, a_2) = \alpha\bar{u}(\bar{\gamma}_P^{a_1}, a_1) + (1-\alpha)\bar{u}(\bar{\gamma}_P^{a_3}, a_3).$$

which directly contradicts A6 (Nonlinearity). This contradiction establishes strict convexity of T^{CONV}, completing the second part of the proof.

The final step is to show that, since we have found the strictly convex function for which $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$, A7 and A8 imply that all optima are seen,

$$C(\mu, A) = \hat{P}(\mu, A|K).$$

Suppose that we have indeed found for $C \in \mathcal{C}$ a strictly convex function function $T_\mu : \Gamma(\mu) \to \bar{\mathbb{R}}$ (real valued on $\tilde{\Gamma}(\mu)$) and corresponding PS cost function **on** (μ, Q) **with** $Q \in Q^C(\mu)$**,**

$$K(\mu, Q) = \sum_{\Gamma(Q)} Q(\gamma) T_\mu(\gamma) - T_\mu(\mu),$$

such that $C(\mu, A) \subset \hat{P}(\mu, A|K)$ all $(\mu, A) \in \mathcal{D}$. What we show now is that, if C satisfies A7 and A8, then, given arbitrary $(\mu, A) \in \mathcal{D}$ and corresponding optimal strategy $\lambda \in \hat{\Lambda}(\mu, A|K)$, $P_\lambda \in C(\mu, A)$. To prove this we first invoke Lemma 2.10 (Decomposition and Uniqueness), which implies that there exist strategies $\lambda^*(l) = (Q_l^*, q_l^*) \in \hat{\Lambda}(\mu, A)$ for $1 \leq l \leq L$ and corresponding probability weights $\alpha(l)$ such that,

$$\lambda \equiv \sum_{l=1}^{L} \alpha(l) \lambda^*(l),$$

with each strategy $\lambda^*(l)$ uniquely optimal with regard to the chosen actions $\mathcal{A}[\lambda^*(l)] \subset A$,

$$\hat{\Lambda}(\mu, \mathcal{A}[\lambda^*(l)]) = \{\lambda^*(l)\}.$$

For each l and $a \in A$ we now construct sequences of actions $a(l, m)$ for $1 \leq m \leq \infty$ as follows:

$$u(a(l,m), j) = \begin{cases} u(a, j) \text{ if } a \in \mathcal{A}[\lambda^*(l)]; \\ u(a, j) - \frac{1}{m} \text{ if } a \in A/\mathcal{A}[\lambda^*(l)]. \end{cases}$$

We now define action sets $A(l, m)$ as the corresponding unions,

$$A(l, m) = \cup_{a \in A} a(l, m).$$

Note that this addition of new actions with lowered payoffs does not expand the set of optimal strategies beyond those in $\mathcal{A}[\lambda^*(l)]$ by Lemma 2.8. Hence,

$$\hat{\Lambda}(\mu, A(l, m)) = \lambda^*(l).$$

Existence of a CIR and uniqueness of the optimal strategy in the perturbed problems implies that the corresponding data is observed all the way to the limit.

$$\mathbf{P}_{\lambda^*(l)} \in \cap_{m=1}^{\infty} C(\mu, A(l, m)).$$

We are now in position to apply the Axiom A8 (Continuity). By construction, $\lim_{m\to\infty} a(l, m) = a$, $A(m) = \cup_{a \in A} a(l, m)$, and $\mathcal{A}(\mathbf{P}_{\lambda^*(l)}) \subset A$, so that this axiom implies that the data is also observed in the limit problem,

$$\mathbf{P}_{\lambda^*(l)} \in C(\mu, A).$$

Since this is true for all l, we can apply the Axiom A7 (Convexity) to conclude that the convex combination of data corresponding to the given strategy λ is also observed,

$$\sum_{l=1}^{L} \alpha(l)\mathbf{P}_{\lambda^*(l)} \in C(\mu, A).$$

To complete the proof, we note from Lemma 2.18 that, since $\lambda \equiv \sum_{l=1}^{L} \alpha(l)\lambda^*(l)$,

$$\mathbf{P}_\lambda = \sum_{l=1}^{L} \alpha(l)\mathbf{P}_{\lambda^*(l)} \in C(\mu, A),$$

as required. This completes the proof that $C(\mu, A) = \hat{P}(\mu, A|K)$ and with it the overall sufficiency proof. ■

A3.1: Recoverability

A second recoverability results follow from the above logic. It establishes a simple method of recovering this cost function.

Corollary 2: If $C \in \mathcal{C}$ has a PS representation $K \in \mathcal{K}^{PS}$, then, given $\mu \in \Gamma$ and non-degenerate $\bar{Q} \in \mathcal{Q}^C(\mu)$, there exists $\bar{A}$ for which there is both an inattentive optimal strategy, $\eta \in \Lambda^I(\mu, \bar{A}) \cap \hat{\Lambda}(\mu, \bar{A})$, and an attentive optimal strategy $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, \bar{A})$ with $Q_\lambda(\gamma) = \bar{Q}(\gamma)$ all $\gamma \in \Gamma(\bar{Q})$, so that,

$$K(\mu, Q_\lambda) = U(\lambda) - U(\eta).$$

Proof. If $C \in \mathcal{C}$ has a PS representation $K \in \mathcal{K}^{PS}$, consider $\mu \in \Gamma$ and non-degenerate $\bar{Q} \in \mathcal{Q}^C(\mu)$, and note that $\sum_{\gamma \in \Gamma(Q)} \gamma\bar{Q}(\gamma) = \mu$. By Lemma 2.17, the PS representation ensures $\mathcal{Q}^C(\mu) = \Delta(\hat{\Gamma}(\mu|K))$. The construction of $\bar{A}$ is based entirely on $f_\mu(\gamma)$, the function identified in the FIO Lemma:

$$\bar{A} = \{\cup_{\gamma \in \Gamma(\bar{Q})} f_\mu(\gamma)\} \cup \{f_\mu(\mu)\}.$$

The two strategies are defined precisely by deterministic selection of $f_\mu(\gamma)$ at posterior γ, $q_\lambda(f_\mu(\gamma)|\gamma) = 1$. By construction both strategies above satisfy $\lambda, \eta \in \Lambda(\mu, \bar{A})$. Hence we can apply FIO directly to conclude that $\lambda, \eta \in \hat{\Lambda}(\mu, \bar{A}|K)$. Hence they have equal expected utility net of attention costs. Hence the cost difference must be the same as the difference in expected utility,

$$K(\mu, \bar{Q}) - K(\mu, \eta) = U(\lambda) - U(\eta).$$

By construction, the inattentive strategy is free, $K(\mu, \eta) = 0$, completing the proof. ■

Appendix 4: Theorem 4

In this section of the appendix we take A2 through A8 and Theorem 3 as our starting point and show that addition of Locally Invariant Posteriors (A9) validates Theorem 4. We first establish some useful Lemmas. We then provide the necessity proof, which follows directly, and finally the sufficiency proof. Before doing this, we restate key definitions to ease reading of the Appendix. A PS cost function $K \in \mathcal{K}^{PS}$ is UPS $K \in \mathcal{K}^{UPS}$, if there exists a strictly convex function $T : \Gamma \to \mathbb{R}$ such that,

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)T(\gamma) - T(\mu),$$

all $(\mu, Q) \in \mathcal{F}$ such that $Q \in \hat{\mathcal{Q}}(\mu|K) \equiv \mathcal{Q}(\mu) \cap \Delta(\hat{\Gamma}(\mu|K))$, where $\hat{\Gamma}(\mu|K)$ is the optimal posterior set,

$$\hat{\Gamma}(\mu|K) = \{\gamma \in \Gamma | \exists\, (\mu, A) \in \mathcal{D} \text{ and } \lambda \in \hat{\Lambda}(\mu, A|K) \text{ with } \gamma \in \Gamma(Q_\lambda)\}.$$

Some additional notation simplifies proofs. We introduce special notation for the subset of $\mathcal{F}(\mu)$ consistent with optimality,

$$\hat{\mathcal{F}}(\mu|K) = \left\{(\mu, Q) \in \mathcal{F}(\mu) | Q \in \mathcal{Q}(\mu) \cap \Delta(\hat{\Gamma}(\mu|K))\right\}.$$

Given a UPS function, we define also net utility using the common cost function,

$$N^a(\gamma) \equiv \bar{u}(\gamma, a) - T(\gamma). \tag{52}$$

Finally, recall the notation γ_λ^a defined in Lemma 2.3: this is the unique posterior at which any chosen action $a \in \mathcal{A}(\lambda)$ may be chosen, $q_\lambda(a|\ \gamma_\lambda^a) > 0$, in an optimal strategy for a given PS cost function $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}\,(\mu, A|K)$.

A4.1: Lemmas

Lemma 4.1: (UPS Lagrangean Lemma) Given $\lambda = (Q_\lambda, q_\lambda) \in \Lambda\,(\mu, A)$ for $K \in \mathcal{K}^{UPS}$, $\lambda \in \hat{\Lambda}\,(\mu, A|K)$ if and only if $\exists \theta \in \mathbb{R}^{J-1}$ s.t.,

$$N^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{a' \in A, \gamma' \in \Gamma(\mu)} N^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j), \tag{53}$$

all $\gamma \in \Gamma(\mu)$, $b \in A$, and $a \in \mathcal{A}(\lambda)$.with equality if $\gamma \in \Gamma(Q_\lambda)$ and $q_\lambda(a|\gamma) > 0$.

Proof. By the standard Lagrangian Lemma 2.6 above, since $K \in \mathcal{K}^{UPS}$ implies that $K \in \mathcal{K}^{PS}$, we know that, given $(\mu, A) \in \mathcal{D}$, $\lambda \in \hat{\Lambda}(\mu, A|K)$ if and only if $\exists \theta \in \mathbb{R}^{J-1}$ s.t.,

$$N_\mu^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{a' \in A, \gamma' \in \Gamma(\mu)} N_\mu^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j),$$

all $\gamma \in \Gamma(\mu)$ and $a \in A$, with equality if $\gamma \in \Gamma(Q_\lambda)$ and $q_\lambda(a|\gamma) > 0$. But with $K \in \mathcal{K}^{UPS}$, we know that, for all μ, this holds also for the fixed function $N^a(\gamma)$ defined in (52), confirming (53). ■

Lemma 4.2: LIP in Optimal Strategies Given $K \in \mathcal{K}^{UPS}$, consider $(\mu, A) \in \mathcal{D}$ and $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$. Now consider $\rho(a) > 0$ on $A' \subset \mathcal{A}(\lambda)$ with $\sum_{a \in A'} \rho(a) = 1$ and define $\mu' = \sum_{a \in A'} \rho(a)\gamma_\lambda^a$ and $\lambda' = (Q', q') \in \Lambda(\mu', A')$ with $\Gamma(Q') \subset \hat{\Gamma}(\mu'|K)$ by:

$$
\begin{aligned}
Q'(\gamma) &= \begin{cases} \sum_{\{a \in A' | \gamma_\lambda^a = \gamma\}} \rho(a) \text{ if } \gamma \in \Gamma(Q'); \\ 0 \text{ else.} \end{cases} \\
q'(a|\gamma) &= \begin{cases} \frac{\rho(a)}{Q'(\gamma)} \text{ if } \gamma = \gamma_\lambda^a; \\ 0 \text{ else.} \end{cases}
\end{aligned}
$$

Then $\lambda' \in \hat{\Lambda}(\mu', A'|K)$.

Proof. Consider $(\mu, A) \in \mathcal{D}$ and $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$ for $K \in \mathcal{K}^{UPS}$. Given $K \in \mathcal{K}^{UPS}$ and that $\lambda \in \hat{\Lambda}(\mu, A|K)$, we know that $\Gamma(Q_\lambda) \subset \hat{\Gamma}(\mu|K)$ so that $Q_\lambda \in \Delta(\hat{\Gamma}(\mu|K))$. Hence we can use the common strictly convex function $T : \Gamma \to \mathbb{R}$ in computing the corresponding costs,

$$K(\mu, Q_\lambda) = \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma)T(\gamma) - T(\mu).$$

We define N^a to be the corresponding net utility using the common cost function as in (52). Given that $\lambda \in \hat{\Lambda}(\mu, A|K)$ for $K \in \mathcal{K}^{PS}$, we can apply Lemma 4.1, the UPS Lagrangian Lemma, to identify multipliers $\theta(j)$ such that,

$$N^b(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq N^a(\gamma_\lambda^a) - \sum_{j=1}^{J-1} \theta(j)\gamma_\lambda^a(j),$$

all $\gamma \in \Gamma(\mu)$, $b \in A$, and $a \in \mathcal{A}(\lambda)$. We rewrite the above as an equation and a set of inequalities. Given $a \in \mathcal{A}(\lambda)$,

$$
\begin{aligned}
N^c(\bar{\gamma}^c) - \sum_{j=1}^{J-1} \theta(j)\gamma_\lambda^c &= N^a(\gamma_\lambda^a) - \sum_{j=1}^{J-1} \theta(j)\gamma_\lambda^a(j) \text{ for } c \in \mathcal{A}(\lambda); \\
N^b(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) &\leq N^a(\gamma_\lambda^a) - \sum_{j=1}^{J-1} \theta(j)\gamma_\lambda^a(j) \text{ for } b \in A \text{ and } \gamma \in \Gamma(\mu).
\end{aligned}
$$

We now consider $\rho(a) > 0$ on $A' \subset \mathcal{A}(\lambda)$ with $\sum_{a \in B} \rho(a) = 1$ and define $\mu' = \sum_{a \in A'} \rho(a)\gamma_\lambda^a$ and $\lambda' = (Q', q') \in \Lambda(\mu', A')$ as in the statement of this Lemma. Given that $\Gamma(Q') \subset \hat{\Gamma}(\mu'|K)$ and $K \in \mathcal{K}^{UPS}$, we know that we can again use the common T function in expressing all net utilities, so that,

$$K(\mu', Q') = \sum_{\gamma \in \Gamma(Q')} Q'(\gamma)T(\gamma) - T(\mu').$$

We now apply the UPS Lagrangian Lemma using the same multipliers. Note that all equalities and inequalities defining of optimality remain valid. The subtlety is that there may be different state spaces, $\Omega(\mu') \neq \Omega(\mu)$, hence the summation and relevant multipliers $\theta(j)$ are only those in the smaller space, $j \in \Omega(\mu')$. The key observation that makes this irrelevant and validates the corresponding inequalities restricted to that subspace is that, for all $\gamma \in \Gamma(Q')$, the posteriors $\gamma(j)$ on all states $j \notin \Omega(\mu')$ are zero. Hence the corresponding terms add nothing to any of the terms on either the left-hand side or right-hand side, leaving the inequalities valid on the smaller state space to conclude that $\lambda' \in \hat{\Lambda}(\mu', A'|K)$, completing the proof. ■

Lemma 4.3: Invariance Under Affine Transforms Consider $K \in \mathcal{K}^{PS}$, $\mu \in \Gamma$, and $T_\mu : \Gamma(\mu) \to \bar{\mathbb{R}}$, such that, given $Q \in \hat{Q}(\mu|K)$,

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) T_\mu(\gamma) - T_\mu(\mu).$$

Then if $\tilde{T}_\mu(\gamma) = T_\mu(\gamma) + \alpha + \beta.\gamma$ some $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^{|\Omega(\mu)|}$, then

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \tilde{T}_\mu(\gamma) - \tilde{T}_\mu(\mu).$$

Proof. Pick $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^{|\Omega(\mu)|}$ and define the corresponding affine transforms,

$$\begin{aligned} \tilde{T}_\mu(\gamma) &= T_\mu(\gamma) + \alpha + \beta.\gamma; \text{ and} \\ \tilde{K}(\mu, Q) &= \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \tilde{T}_\mu(\gamma) - \tilde{T}_\mu(\mu). \end{aligned}$$

To see that the cost function is unchanged by this, note that,

$$\begin{aligned} \tilde{K}(\mu, Q) &= \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \tilde{T}_\mu(\gamma) - \tilde{T}_\mu(\mu) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) \left[T_\mu(\gamma) + \alpha + \beta.\gamma \right] - \left[T_\mu(\mu) + \alpha + \beta.\mu \right] \\ &= \sum_{\gamma \in \Gamma(Q)} Q(\gamma) T_\mu(\gamma) - T_\mu(\mu) - \beta. \left[\sum_{\gamma \in \Gamma(Q)} \gamma Q(\gamma) - \mu \right] = K(\mu, Q), \end{aligned}$$

since $\sum_{\gamma \in \Gamma(Q)} \gamma Q(\gamma) = \mu$. ■

Lemma 4.4: UPS Regularity If $C \in \mathcal{C}$ has a PS representation $K \in \mathcal{K}^{PS}$, and A9 is satisfied, then C is regular, $C \in \mathcal{C}^R$.

Proof. Since $C \in \mathcal{C}$ has a PS representation $K \in \mathcal{K}^{PS}$ we know by Theorem 3 that it satisfies A2 through A8. To establish that satisfaction in addition of A9 yields regularity, we need to show that, in this case, given $\mu_1 \in \Gamma$ and $Q \in \Delta(\Gamma(\mu_1))$ with $\Gamma(Q) \subset \Gamma^C(\mu_1)$,

$$\sum_{\gamma \in \Gamma(\mu_2)} \gamma Q(\gamma) = \mu_2 \Longrightarrow \Gamma(Q) \subset \Gamma^C(\mu_2).$$

Given such $\mu_1 \in \Gamma$ and $Q \in \Delta(\Gamma(\mu_1))$ with $\Gamma(Q) \subset \Gamma^C(\mu_1)$, we know from Completeness (A4) that there exists a corresponding $(\mu_1, A_1) \in \mathcal{D}$, $P \in C(\mu_1, A_1)$, and $\boldsymbol{\lambda}_P = (\mathbf{Q}_P, \mathbf{q}_P)$ such that $\Gamma(Q)$ is a subset of the support,

$$\Gamma\left(Q\right) \subset \Gamma\left(\mathbf{Q}_P\right).$$

By construction the probability that action $a \in \mathcal{A}(P) \subset A_1$ is chosen at revealed posterior $\bar{\gamma}_P^a \in \Gamma\left(\mathbf{Q}_P\right)$ in strategy $\boldsymbol{\lambda}_P = (\mathbf{Q}_P, \mathbf{q}_P)$ is greater than 0,

$$\mathbf{q}_P(a|\bar{\gamma}_P^a) > 0.$$

Now suppose that $\sum\limits_{\gamma \in \Gamma(\mu_1)} \gamma Q(\gamma) = \mu_2$ and define $A_2 \subset \mathcal{A}(P)$ to comprise the actions chosen at the posteriors in set $\Gamma\left(Q\right)$,

$$A_2 = \{a \in A_1 | \bar{\gamma}_P^a \in \Gamma\left(Q\right)\}.$$

With LIP (A9) we know that, since $P \in C(\mu_1, A_1)$, $A_2 \subset \mathcal{A}(P)$, and we have found probabilities $Q(\bar{\gamma}_P^a) > 0$ all $a \in A_2$ with $\sum\limits_{a \in A_2} Q(\bar{\gamma}_P^a) = 1$, that the data set $P_2 \in \mathcal{P}\left(\sum\limits_{a \in A_2} \bar{\gamma}_P^a Q(\bar{\gamma}_P^a), A'\right)$ that satisfies $\mathcal{A}(P_2) = A_2$, $\mathbf{Q}_{P_2}(a) = Q(\bar{\gamma}_P^a)$, and $\bar{\gamma}_{P_2}^a = \bar{\gamma}_P^a$ is observed at the corresponding prior $\mu_2 = \sum\limits_{a \in A_2} \bar{\gamma}_P^a Q(\bar{\gamma}_P^a)$,

$$P_2 \in C\left(\mu_2, A'\right).$$

Hence $\Gamma\left(Q\right) \subset \Gamma^C(\mu_2)$, establishing regularity. ■

Lemma 4.5: Given $C \in \mathcal{C}^R$ with a UPS representation $K \in \mathcal{K}^{UPS}$,

$$\Omega(\mu_1) = \Omega(\mu_2) \Longrightarrow \Gamma^C(\mu_1) = \Gamma^C(\mu_2).$$

Proof. Given $C \in \mathcal{C}$ with a UPS representation it also has a PS representation, hence by theorem 3 satisfies A2 through A8. For the purposes of this proof we set $\Omega(\mu_1) = \Omega(\mu_2) = \Omega$, and know by Completeness (A4) that both contain the common set of interior vectors, from which we correspondingly remove the subscript:

$$\tilde{\Gamma} = \tilde{\Gamma}(\mu_1) = \tilde{\Gamma}(\mu_2) \subset \Gamma^C(\mu_1) \cap \Gamma^C(\mu_2)$$

We now consider an arbitrary posterior $\eta \in \Gamma^C(\mu_1)$ with $\eta(j) = 0$ for some $j \in \Lambda(\mu_1)$ and show that, if $C \in \mathcal{C}^R$, then $\eta \in \Gamma^C(\mu_2)$.

For $1 \leq k \leq J = |\Omega|$ we create corresponding set of interior "basis" vectors constructed in such a manner that they allow us to construct distributions $\bar{Q}_1$ and $\bar{Q}_1$ over them that generate both μ_1 and μ_2. To do this, weight together the unit posteriors $e_k \in \Gamma_1 = \Gamma_2 = \Gamma_{12}$ with 1 in position k and zeroes elsewhere and their average $\bar{e} = \sum\limits_{k=1}^{J} e_k/J$, to arrive at a set of interior posteriors $\bar{\gamma}_k \in \Gamma_{12}$ that span (in the linear algebra sense) the set Γ_{12} and that contain μ_1 and μ_2 in the interior of their convex hull, which is possible since we know that μ_1 and μ_2 are both interior to Γ_{12}. Technically, we find $\delta \in (0, 1)$ such that, when we define the corresponding posteriors $\bar{\gamma}_k^\delta$ for $1 \leq k \leq J$,

$$\bar{\gamma}_k^\delta(j) = \begin{cases} \frac{\delta}{J} + (1-\delta) \text{ if } k = j; \\ \frac{\delta}{J} \text{ if } k \neq j \ . \end{cases}$$

the unique probability weights $\bar{Q}_i(k)$ for $i = 1, 2$ and $1 \le k \le J$ with $\sum_{k=1}^{J} \bar{Q}_i(k) = 1$ that re-weight the posteriors to regenerate each prior,

$$\sum_{k=1}^{J} \bar{\gamma}_k^{\delta} \bar{Q}_i(\bar{\gamma}_k^{\delta}) = \mu_i;$$

are all strictly positive probability, $\bar{Q}_i(\bar{\gamma}_k^{\delta}) > 0$ for $i = 1, 2$ and $1 \le k \le J$. Going forward we suppress the δ parameter and set $\bar{\gamma}_k^{\delta} = \bar{\gamma}_k$.

In the next step we adjust the weights $\bar{Q}_2(\bar{\gamma}_k)$ and define a distribution of posteriors $Q_2 \in \Delta\left(\Gamma^C(\mu_1)\right)$ with,

$$\Gamma(Q_2) = \eta \cup \{\bar{\gamma}_k | 1 \le k \le J\} \text{ and } \sum_{\gamma \in \Gamma(Q_2)} \gamma Q_2(\gamma) = \mu_2.$$

To accomplish this, we pick $\epsilon > 0$ small enough so that,

$$\max_{k=1..J} \frac{\epsilon \eta(k)}{1-\delta} < \min_{k=1..J} \{\bar{Q}_2(k)\},$$

define $Q_2(\eta) = \epsilon$ and then subtract the corresponding amount from the probability of $\bar{\gamma}_k$,

$$Q_2(\bar{\gamma}_k) = \bar{Q}_2(\bar{\gamma}_k) - \frac{\epsilon}{1-\delta}\left[\eta(k) - \frac{\delta}{J}\right],$$

so that,

$$\begin{aligned} Q_2(\eta) + \sum_{k=1}^{J} Q_2(\bar{\gamma}_k) &= \epsilon + 1 - \frac{\epsilon}{1-\delta}\left[\sum_{k=1}^{J} \eta(k) - \delta\right] \\ &= \epsilon + 1 - \epsilon = \epsilon, \end{aligned}$$

so that this is a probability distribution over posteriors. Note also that,

$$\sum_{\gamma \in \Gamma(Q_2)} \gamma(j) Q_2(\gamma) = \sum_{k=1}^{J} \bar{\gamma}_k(j) \left(\bar{Q}_2(\bar{\gamma}_k) - \left[\frac{\epsilon}{1-\delta}\left(\eta(k) - \frac{\delta}{J}\right)\right]\right) + \eta(j)\epsilon.$$

Note that $\sum_{k=1}^{J} \bar{\gamma}_k(j) \bar{Q}_2(\bar{\gamma}_k) = \mu_2(j)$ so that $\sum_{\gamma \in \Gamma(Q_2)} \gamma(j) Q_2(\gamma) = \mu_2(j)$ if and only if,

$$\sum_{k=1}^{J} \bar{\gamma}_k(j) \left(\left[\frac{1}{1-\delta}\left(\eta(k) - \frac{\delta}{J}\right)\right]\right) = \eta(j). \tag{54}$$

Directly,

$$\begin{aligned}\sum_{k=1}^{J}\bar{\gamma}_k(j)\left(\left[\frac{1}{1-\delta}\left(\eta(k)-\frac{\delta}{J}\right)\right]\right) &= \sum_{k=1}^{J}\frac{\delta}{J}\left[\frac{1}{1-\delta}\left(\eta(k)-\frac{\delta}{J}\right)\right]+(1-\delta)\left[\frac{1}{1-\delta}\left(\eta(j)-\frac{\delta}{J}\right)\right] \\ &= \frac{\delta}{J(1-\delta)}\sum_{k=1}^{J}\left(\eta(k)-\frac{\delta}{J}\right)+(1-\delta)\left[\frac{1}{1-\delta}\left(\eta(j)-\frac{\delta}{J}\right)\right] \\ &= \frac{\delta}{J(1-\delta)}[1-\delta]+\eta(j)-\frac{\delta}{J}=\eta(j),\end{aligned}$$

confirming (54).

Note by construction that $\Gamma(Q_2) \subset \Gamma^C(\mu_1)$. Given that $C \in \mathcal{C}^R$ the fact that $\sum_{\gamma\in\Gamma(Q_2)}\gamma Q_2(\gamma) = \mu_2$ implies that $\Gamma(Q_2) \subset \Gamma^C(\mu_1)$, hence that $\eta \in \Gamma^C(\mu_2)$, so that $\Gamma^C(\mu_1) \subset \Gamma^C(\mu_2)$. Note that the converse argument is identical since the state spaces are identical, so that $\Gamma^C(\mu_1) = \Gamma^C(\mu_2)$, completing the proof. ■

A4.2: Theorem 4

Theorem 4: Necessity If data set $C \in \mathcal{C}^R$ has a UPS representation it satisfies A2 through A9.

Proof. Given that $C \in \mathcal{C}^R$ has a UPS representation, it has a PS representation $K \in \mathcal{K}^{PS}$, and A2-A8 are satisfied. To establish A9, we pick $(\mu, A) \in \mathcal{D}$, $P \in C(\mu, A)$, and probabilities $\rho(a) > 0$ on $A' \subset \mathcal{A}(P)$ with $\sum_{a\in A'}\rho(a) = 1$. Since this is a PS representation and $P \in C(\mu, A)$, the corresponding revealed strategy is optimal by Lemma 2.15,

$$\boldsymbol{\lambda}(P) = (\mathbf{Q}_P, \mathbf{q}_P) \in \hat{\Lambda}(\mu, A|K).$$

By definition, $\boldsymbol{\lambda}(P) = (\mathbf{Q}_P, \mathbf{q}_P)$ is defined by $\Gamma(\mathbf{Q}_P) = \cup_{a\in\mathcal{A}(P)}\bar{\gamma}_P^a$ and:

$$\begin{aligned}\mathbf{Q}_P(\gamma) &= \sum_{\{a\in\mathcal{A}(P)|\bar{\gamma}_P^a=\gamma\}} P(a) \\ \mathbf{q}_P(a|\gamma) &= \begin{cases}\frac{P(a)}{\mathbf{Q}_P(\gamma)} & \text{if } \bar{\gamma}_P^a = \gamma; \\ 0 & \text{if } \bar{\gamma}_P^a \neq \gamma;\end{cases}\end{aligned}$$

on $\gamma \in \Gamma(\mathbf{Q}_P)$ and $a \in A$. Since $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K)$, it is definitional that $\Gamma(\mathbf{Q}_P) \subset \Gamma^C(\mu)$.

We now define $P' \in \mathcal{P}$ as in the LIP definition by$\mathcal{A}(P') = A'$, $\mathbf{Q}_{P'}(\gamma) = \sum_{\{a\in A'|\bar{\gamma}_P^a=\gamma\}}\rho(a)$;

$$\mathbf{q}_{P'}(a|\gamma) = \begin{cases}\frac{\rho(a)}{\mathbf{Q}_P(\gamma)} \text{ if } \bar{\gamma}_P^a = \gamma'; \\ \mathbf{q}_{P'}(a|\gamma) = 0 \text{ else.}\end{cases}$$

To show that $P' \in C(\mu', A')$, where,

$$\mu' = \sum_{a\in A'}\rho(a)\bar{\gamma}_P^a,$$

we consider the revealed attention strategy associated with $P' \in C(\mu', A')$, $\boldsymbol{\lambda}(P') = (\mathbf{Q}_{P'}, \mathbf{q}_{P'}) \in \Lambda(\mu, A)$:

$$\mathbf{Q}_{P'}(\gamma) = \begin{cases} \sum_{\{a \in A' | \bar{\gamma}_P^a = \gamma\}} \rho(a) \text{ if } \gamma \in \Gamma(\mathbf{Q}_P); \\ 0 \text{ else.} \end{cases}$$

$$\mathbf{q}_{P'}(a|\gamma) = \begin{cases} \frac{\rho(a)}{Q_\eta(\gamma)} \text{ if } \gamma = \bar{\gamma}_P^a; \\ 0 \text{ else.} \end{cases}$$

Note that this strategy is derived from $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K)$ precisely as prescribed in Lemma 4.2, since $\bar{\gamma}_P^a = \gamma_{\boldsymbol{\lambda}(P)}^a \in \Gamma(\mu)$ is indeed the unique posterior with $\mathbf{q}_{P'}(a|\gamma) > 0$ by Lemma 2.12. Hence, provided $\Gamma(\mathbf{Q}_{P'}) \subset \hat{\Gamma}(\mu', K)$ we can conclude from Lemma 4.2 that $\boldsymbol{\lambda}(P') \in \hat{\Lambda}(\mu', A'|K)$. To establish this, note that $\mathbf{Q}_{P'} \in \Delta(\Gamma(\mu))$ satisfies $\Gamma(\mathbf{Q}_{P'}) \subset \Gamma(\mathbf{Q}_P) \subset \Gamma^C(\mu)$, and,

$$\sum_{\gamma \in \Gamma(\mathbf{Q}_{P'})} \gamma \mathbf{Q}_{P'}(\gamma) = \sum_{a \in A'} \rho(a) \bar{\gamma}_P^a = \mu'.$$

Since $C \in \mathcal{C}^R$, we conclude that indeed $\Gamma(\mathbf{Q}_{P'}) \subset \Gamma^C(\mu')$, so that Lemma 4.2 does apply to ensure that $\boldsymbol{\lambda}(P') \in \hat{\Lambda}(\mu', A'|K)$. Since this is a PS representation, we know further that,

$$\mathbf{P}_{\boldsymbol{\lambda}(P')} \in C(\mu', A').$$

Note finally that by Lemma 2.13,

$$\mathbf{P}_{\boldsymbol{\lambda}(P')} = P',$$

completing the necessity proof. ■

Theorem 4: Sufficiency If data set $C \in \mathcal{C}$ satisfies A2 through A9, it has a UPS representation.

Proof. A2-A8 guarantee existence of a PS representation. To establish existence of a UPS representation, we know that we can identify corresponding functions $K_\mu \in \mathcal{K}^{PS}$ any $\mu \in \Gamma$ and a corresponding strictly convex functions $T_\mu : \Gamma(\mu) \to \bar{\mathbb{R}}$ that is real valued on $\tilde{\Gamma}(\mu)$. We use these functions to define our candidate real-valued function $T(\gamma) \in \mathbb{R}$ on $\gamma \in \Gamma$. Specifically, we define the corresponding uniform prior $\bar{\mu}(\gamma)$ that assigns probability $\frac{1}{|\Omega(\gamma)|}$ to each state in $\Omega(\gamma)$ and specify this as $T(\gamma)$:

$$T(\gamma) \equiv T_{\bar{\mu}(\gamma)}(\gamma), \tag{55}$$

As noted, this is real-valued by definition of a PS representation since $\gamma \in \tilde{\Gamma}(\bar{\mu}(\gamma))$. Note by Lemma 4.3 that $T(\gamma)$ is not unique: but the affine transforms are irrelevant as we will see. We establish now that this definition ensures that the defining property of the UPS representation holds: given $\mu \in \Gamma$,

$$K(\mu, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma) T(\gamma) - T(\mu), \tag{56}$$

all $(\mu, Q) \in \hat{\mathcal{F}}(\mu|K)$.

We establish this result in two stages according to the support. We prove first that (56) would follow provided it held true for the special class of priors that are uniform over a finite set of states.

We then establish that indeed (56) does hold for uniform priors.

With regard to the first step, we let μ_1 be a uniform prior over an arbitrary state space, and pick any distinct non-uniform prior with the same state space, so that $\Omega(\mu_1) = \Omega(\mu_2)$. Given that a PS representation exists, we know that we can identify corresponding functions $K_i \equiv K_{\mu_i} \in \mathcal{K}^{PS}$ for $i = 1, 2$. We fix corresponding strictly convex functions $T_i : \Gamma(\mu_i) \to \mathbb{R}$ which have the PS property: for feasible strategies $\lambda = (Q_\lambda, q_\lambda) \in \Lambda(\mu_i)$, using for T_1 the version used in constructing $T(\gamma)$. Correspondingly for $a \in \mathcal{A}$ we define,

$$N_i^a(\gamma) = \sum_{j=1}^{J} u(a, j)\gamma(j) - T_i(\gamma),$$

on $\gamma \in \Gamma(\mu_i)$.

With Lemma 4.5 we know that,

$$\Gamma^C(\mu_1) = \Gamma^C(\mu_2) \equiv \Gamma^C.$$

In light of Lemma 4.3, our goal in this part of the proof is to show that $T_2 : \Gamma(\mu_2) \to \bar{\mathbb{R}}$ is an affine transform of $T_1 : \Gamma(\mu_2) \to \bar{\mathbb{R}}$, since then the two functions can be reduced to equality.

We first focus on prior μ_1. By FIO, there is a 1-1 function $f_1 : \Gamma^C \to \mathcal{A}$ such that,

$$N_1^{f_1(\gamma)}(\phi) = \sum_{j=1}^{J} u(f_1(\gamma), j)\phi(j) - T_1(\phi) \le \sum_{j=1}^{J} u(f_1(\gamma), j)\gamma(j) - T_1(\gamma) = N_1^{f_1(\gamma)}(\gamma) \equiv 0, \tag{57}$$

all $\phi, \gamma \in \Gamma$. We find the particular actions associated with the spanning vectors introduced in Lemma 4.5,

$$\bar{a}_k = f_1(\bar{\gamma}_k) \in \mathcal{A},$$

for $1 \le k \le J$, where

$$\bar{\gamma}_k(j) = \begin{cases} \frac{\delta}{J} + (1 - \bar{\delta}) \text{ if } k = j; \\ \frac{\bar{\delta}}{J} \text{ if } k \neq j \text{ ;} \end{cases} \tag{58}$$

with $\bar{\delta} \in (0, 1)$ fixed to ensure strict positivity of all weights $\bar{Q}_i(k) > 0$ for $i = 1, 2$ and $1 \le k \le J$ with $\sum_{k=1}^{J} \bar{Q}_i(k) = 1$ that re-weight the posteriors to regenerate each prior,

$$\sum_{k=1}^{J} \bar{\gamma}_k \bar{Q}_i(\bar{\gamma}_k) = \mu_i.$$

We define the corresponding action set $\bar{A} = \cup_{k=1}^{J} \bar{a}_k$.

By FIO, we can identify an optimal strategy $\lambda(1) = (Q_1, q_1) \in \hat{\Lambda}(\mu_1, \bar{A}|K_1)$ having posteriors $\Gamma(Q_1) = \cup_{k=1}^{J} \bar{\gamma}_k$, placing probability weights on them according to $\bar{Q}_1$, and involving deterministic choice at each possible posterior of the corresponding action,

$$\begin{aligned} Q_1(\bar{\gamma}_k) &= \bar{Q}_1(k); \\ q_1(\bar{a}_k|\bar{\gamma}_k) &= 1. \end{aligned}$$

We also define strategy $\lambda(2) = (Q_2, q_2)$ as having the same possible posteriors, the same deterministic choice at each possible posterior, yet placing probability weights on them according to Q_2,

$$Q_2(\bar{\gamma}_k) = \bar{Q}_2(k).$$

The key observation is that with A9, $\lambda(2) \in \hat{\Lambda}(\mu_2, \bar{A}|K_2)$. To see this, note first that since this is a CIR $\lambda(1) = (Q_1, q_1) \in \hat{\Lambda}(\mu_1, \bar{A}|K_1)$, the corresponding SDSC data satisfies $\mathbf{P}_{\lambda_1} \in C(\mu_1, \bar{A})$. We now define data set P_2 by,

$$P_2(\bar{a}_k|j) = \frac{\bar{\gamma}_k(j)\bar{Q}_2(k)}{\mu_2(j)},$$

By construction note that this has unconditional action probabilities,

$$P_2(\bar{a}_k) = \sum_{j=1}^{J} \frac{\mu_2(j)\bar{\gamma}_k(j)\bar{Q}_2(k)}{\mu_2(j)} = \sum_{j=1}^{J} \bar{\gamma}_k(j)\bar{Q}_2(k) = \bar{Q}_2(k),$$

and revealed posteriors,

$$\bar{\gamma}_2^k(j) \equiv \bar{\gamma}_{P_2}^{\bar{a}_k}(j) = \frac{\mu_2(j)P_2(\bar{a}_k|j)}{P_2(\bar{a}_k)} = \frac{\mu_2(j)P_2(\bar{a}_k|\omega j)}{\bar{Q}_2(k)} = \bar{\gamma}_k(j).$$

Hence $P_2 \in P(\mu_2, \bar{A})$.

At this point we can apply LIP (A9) to conclude that $P_2 \in C(\mu_2, \bar{A})$ and, by Lemma 2.15, that the related revealed attention strategy $\boldsymbol{\lambda}_{P_2} = \lambda_2$ is optimal, $\lambda_2 \in \hat{\Lambda}(\mu_2, \bar{A}|K_2)$. The Lagrangian Lemma then ensures there are multipliers $\theta \in \mathbb{R}^{J-1}$ s.t., for $1 \leq k, l \leq J$ such that,

$$N_2^{\bar{a}_k}(\bar{\gamma}_k) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_k(j) = N_2^{\bar{a}_l}(\bar{\gamma}_l) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_l(j);$$

or,

$$\sum_{j=1}^{J} u(\bar{a}_k, j)\bar{\gamma}_k(j) - T_2(\bar{\gamma}_k) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_k(j) = \sum_{j=1}^{J} u(\bar{a}_l, j)\bar{\gamma}_l(j) - T_2(\bar{\gamma}_l) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_l(j). \tag{59}$$

By equation (57), we know also that,

$$\begin{aligned} T_1(\bar{\gamma}_k) &= \sum_{j=1}^{J} u(\bar{a}_k, j)\bar{\gamma}_k(j); \text{ and,} \\ T_1(\bar{\gamma}_l) &= \sum_{j=1}^{J} u(\bar{a}_l, j)\bar{\gamma}_l(j). \end{aligned}$$

Substitution in (59) yields,

$$T_1(\bar{\gamma}_k) - T_2(\bar{\gamma}_k) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_k(j) = T_1(\bar{\gamma}_l) - T_2(\bar{\gamma}_l) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_l(j).$$

Hence, for all $1 \leq k, l \leq J$,

$$\sum_{j=1}^{J-1} \theta(j) \left[\bar{\gamma}_k(j) - \bar{\gamma}_l(j)\right] = T_1(\bar{\gamma}_k) - T_2(\bar{\gamma}_k) - T_1(\bar{\gamma}_l) + T_2(\bar{\gamma}_l). \tag{60}$$

The next key claim is that, with FIO, the equation above applies not only to the spanning posteriors but to all pairs of posteriors, $\gamma, \gamma' \in \Gamma$. To see this, set $\gamma = \bar{\gamma}_{J+1}$ and $\gamma' = \bar{\gamma}_{J+2}$ and repeat the above argument to a larger set of posteriors $\cup_{k=1}^{J+2}\bar{\gamma}_k$ and the corresponding actions defined by $f_1 : \Gamma^C \to \mathcal{A}$ as defined by FIO and defined above:

$$\bar{B} = \bar{A} \cup f_1(\gamma) \cup f_1(\gamma').$$

We also find strictly positive probability weights $Q_i'(k)$ for $i = 1, 2$ on $1 \leq k \leq J + 2$ that re-weight the posteriors to regenerate each prior,

$$\sum_{k=1}^{J+2} \bar{\gamma}_k Q_i'(k) = \mu_i.$$

This is possible because the vectors $\bar{\gamma}_k$ span Γ_{12}, so that there are weights $\alpha(k)$ and $\alpha'(k)$ on them that average back to each of γ,γ':

$$\begin{aligned} \sum_{k=1}^{J} \alpha(k)\bar{\gamma}_k &= \gamma; \\ \sum_{k=1}^{J} \alpha'(k)\bar{\gamma}_k &= \gamma' \end{aligned}$$

Note also these weights must sum to 1, as

$$1 = \sum_{\omega \in \Omega} \gamma(\omega) = \sum_{\omega \in \Omega} \sum_{k=1}^{J} \alpha(k)\bar{\gamma}_k(\omega) = \sum_{k=1}^{J} \alpha(k) \sum_{\omega \in \Omega} \bar{\gamma}_k(\omega) = \sum_{k=1}^{J} \alpha(k).$$

Moreover, for all $\epsilon > 0$ and for $i = 1, 2$,

$$\epsilon \left(\gamma + \gamma'\right) + \sum_{k=1}^{J} \gamma_k \left[\bar{Q}_i(k) - \epsilon \left[\alpha(k) + \alpha'(k)\right]\right] = \mu_i.$$

Given that $\bar{Q}_i(k) > 0$ all k, we can select ϵ small enough to keep all terms

$$\bar{Q}_i(k) - \epsilon \left[\alpha(k) + \alpha'(k)\right],$$

strictly positive, as required. Thus, we define new weights by setting $Q_i'(k)$ equal to the above expression for $1 \leq k \leq J$, and equal to ϵ for $J + 1$ and $J + 2$. Repeating the entire remainder of the argument, we apply the Lagrangian Lemma to ensure the existence of multipliers defined by $\eta \in \mathbb{R}^{J-1}$ that produce the corresponding equality for all revealed posteriors, hence in particular for γ and γ':

$$T_1(\gamma) - T_2(\gamma) - \sum_{j=1}^{J-1} \eta(j)\gamma(j) = T_1(\gamma') - T_2(\gamma') - \sum_{j=1}^{J-1} \eta(j)\gamma'(j).$$

A key observation is that the multipliers on the larger set are identical to those on the smaller set, $\eta = \theta$. To see this, note that the equality conditions defining $\theta(j)$ also characterize $\eta(j)$ and have a unique solutions. Specifically, setting $k = j$ and $l = J$, note that posteriors $\bar{\gamma}_j$ and $\bar{\gamma}_J$ differ by $\delta > 0$ in coordinates j and J and are otherwise the same. Hence,

$$\sum_{k=1}^{J-1} \theta(k) \left[\bar{\gamma}_j(k) - \bar{\gamma}_J(k)\right] = \delta\theta(j).$$

This allows us to precisely pin down $\theta(j)$ in terms of the given functions $T_1(\bar{\gamma})$ and $T_2(\bar{\gamma})$ and $\bar{\delta}$ as defined in (58),

$$\theta(j) = \frac{T_1(\bar{\gamma}_j) - T_2(\bar{\gamma}_j) - T_1(\bar{\gamma}_J) + T_2(\bar{\gamma}_J)}{\bar{\delta}},$$

with the corresponding being true for η. This implies that indeed,

$$\begin{aligned} T_2(\gamma) &= T_1(\gamma) - \left[T_1(\gamma') - T_2(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j)\right] - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \\ &= T_1(\gamma) + H_{12}(\gamma') - \theta.\gamma, \end{aligned}$$

where $H_{12}(\gamma') = -\left[T_1(\gamma') - T_2(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j)\right] \in \mathbb{R}$ is independent of γ. This establishes $T_2 : \Gamma(\mu_2) \to \bar{\mathbb{R}}$ is an affine transform of $T_1 : \Gamma(\mu_2) \to \bar{\mathbb{R}}$ so that by Lemma 4.3 we can define,

$$T_2'(\gamma) = T_2(\gamma) + H_{12}(\gamma') - \theta.\gamma = T_1(\gamma);$$

without changing the cost of any attention strategies.

What we have now established is that provided the cost function obeys,

$$K(\bar{\mu}, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)T(\gamma) - T(\bar{\mu}),$$

for all $\bar{\mu} \in \Gamma$ and $(\bar{\mu}, Q) \in \hat{\mathcal{F}}(\bar{\mu}|K)$ that are uniform over some state space, then it holds for all $\mu \in \Gamma$ and $(\mu, Q) \in \hat{\mathcal{F}}(\bar{\mu}|K)$. The subtlety here is that the definition of $T(\gamma)$ adjusts with the cardinality of the support of γ. Hence the question is whether or not one can use the function associated with the lower dimensional prior in characterizing the corresponding cost of $\gamma \in \hat{\Gamma}(\bar{\mu}|K)$ with $|\Omega(\gamma)| < |\Omega(\bar{\mu})|$. This is what we now establish.

Our method of proof ignores the particulars of the state space and involves priors $\mu_1, \mu_2 \in \Omega$ such that $\Omega_2 = \Omega(\mu_2) \subset \Omega(\mu_1)$ with $\Omega(\mu_2) \neq \Omega(\mu_1)$. We define $\Gamma_1 = \Omega(\mu_1)$ and show that we can replace $T_1 : \Gamma_1 \to \bar{\mathbb{R}}$ (real-valued on Γ_1^C) by function $\bar{T}_1 : \Gamma_1 \to \bar{\mathbb{R}}$ that not only retains the PS property,

$$K_1(Q_\lambda) = \sum_{\gamma \in \Gamma(Q_\lambda)} Q_\lambda(\gamma)\bar{T}_1(\gamma) - \bar{T}_1(\mu_1), \tag{61}$$

but is also equal to T_2 on posteriors also on $\gamma \in \Gamma_1^C \cap \tilde{\Gamma}(\mu_2)$. We will use Γ_{12}^C to refer to the set of such posteriors,

$$\Gamma_{12}^C = \Gamma_1^C \cap \tilde{\Gamma}(\mu_2).$$

Note first that $\Gamma_{12}^C \subset \Gamma_2^C$, so that,

$$\Gamma_{12}^C \subset \Gamma_1^C \cap \Gamma_2^C.$$

This comes directly from the fact that all $\gamma \in \Gamma_{12}^C$ are interior to Ω_2 by construction, and so, by Completeness (A4) are used in some problem.

It will be convenient to index the states by first assigning indices $1, ..., J_2 - 1$ to states from Ω_2. We next assign indices $J_2 ... J_1 - 1$ to states from $\Omega(\mu_1)/\Omega(\mu_2)$. Finally, we assign to index J_1 the remaining state from Ω_2. This ensures that the "excluded state" from the Lagrangian statements belongs to Ω_2.

In establishing the existence of $\bar{T}_1$ such that (61) holds, we start as in the equal state space case with the function $f_1 : \Gamma_1^C \to \mathcal{A}$ such that,

$$\sum_{j=1}^{J} u(f_1(\gamma), j)\gamma(j) \equiv T_1(\gamma),$$

all $\gamma \in \Gamma_1$, with $\sum_{j=1}^{J} u(f_1(\gamma), j)\phi(j) \leq T_1(\phi)$ all $\phi \in \Gamma_1$. We also retain the spanning posteriors $\bar{\gamma}_k$ and strictly positive probabilities $\bar{Q}_1(k) > 0$ for $1 \leq k \leq J_1$ that regenerate the prior,

$$\sum_{k=1}^{J_1} \bar{\gamma}_k \bar{Q}_1(k) = \mu_1;$$

and the particular actions $\bar{a}_k = f_1(\bar{\gamma}_k) \in \mathcal{A}$ associated with the spanning vectors.

The key change in the proof is the selection of additional posteriors that sit in Γ_{12}^C and a corresponding set of new strictly positive probability weights. Specifically, we pick a basis for the set Γ_{12}^C and label the finite set of such posteriors as $\Gamma^B \subset \Gamma_{12}^C$. We also associate with these the corresponding average prior, $\bar{\mu}^B$,

$$\bar{\mu}^B(j) = \frac{\sum_{\gamma \in \Gamma^B} \gamma(j)}{|\Gamma^B|},$$

for $1 \leq j \leq J_{12}$. Since $\Omega_2 = \Omega(\bar{\mu}_B)$ we know from the first part of the proof that we can assume that the cost function $K(\bar{\mu}_B, .)$, associated $T_{\bar{\mu}_B}$ and $\Gamma^C(\bar{\mu}_B)$ are identical to $K(\mu_2, .)$, T_2 and $\Gamma^C(\mu_2)$. Below, we will therefore substitute the latter for the former. We identify also the actions $\bar{a}(\gamma) = f_1(\gamma)$ on $\gamma \in \Gamma^B$ and define the larger set of actions,

$$\bar{A}' = \cup_{k=1}^{J_1} \bar{a}_k \cup \{\bar{a}(\gamma) | \gamma \in \Gamma^B\}.$$

By construction, note that $\Omega(\bar{\mu}^B) = \Omega_2$ since a positive posterior probability of a state cannot be generated if all of the basis priors assign it zero probability. The new strictly positive probability weights that we work with place weight $Q_1'(k) > 0$ on all posteriors $\bar{\gamma}_k$ as well as a constant strictly positive weight $\bar{Q}_1'(\gamma) = \epsilon > 0$ on all $\gamma \in \bar{\mu}^B$, while still averaging out to μ_1:

$$\sum_{k=1}^{J_1} \bar{\gamma}_k \bar{Q}_1'(k) + \sum_{\gamma \in \Gamma^B} \epsilon\gamma = \mu_1.$$

The easiest way to do this is to assign a small probability $\delta > 0$ to the mean prior $\bar{\mu}^B$, compensating through appropriate reductions in $\bar{Q}_1(k) > 0$ for $1 \leq k \leq J_1$, as in the construction in Lemma 4.4, while retaining all strictly positive, and thereupon defining $\epsilon = \frac{\delta}{|\Gamma^B|}$.

We now note as before that existence of a PS representation and application of FIO enables us to characterize an optimal strategy $\lambda_1' = (Q_1', q_1') \in \hat{\Lambda}(\mu_1, \bar{A}'|K_1)$ having posteriors $\Gamma(Q_1') = \cup_{k=1}^{J} \bar{\gamma}_k \cup \Gamma^B$, placing probability weights on them according to $\bar{Q}_1'$, and involving deterministic choice at each possible posterior of the corresponding action,

$$\begin{aligned} Q_1'(\bar{\gamma}_k) &= \bar{Q}_1'(k); \\ q_1'(\bar{a}(\gamma)|\gamma) &= 1. \end{aligned}$$

We also define strategy $\lambda_2' = (Q_2', q_2')$ as restricting the posterior set to Γ^B, with equal probability weights, $Q_2'(\gamma) = \frac{1}{|\Gamma^B|}$, with the same deterministic choice at each possible posterior. By construction, $\lambda_2' \in \Lambda(\bar{\mu}^B, \bar{A}')$. In fact, with exactly the same logic as before, LIP (A9) implies that it is optimal, $\lambda_2' \in \hat{\Lambda}(\bar{\mu}^B, \bar{A}'|K_2)$. Hence we can repeat the application of the UPS Lagrangian Lemma to identify $\theta \in \mathbb{R}^{J_2-1}$ s.t., for $\bar{\gamma}, \bar{\gamma}' \in \Gamma^B$,

$$T_1(\bar{\gamma}) - T_2(\bar{\gamma}) - \sum_{j=1}^{J_2-1} \theta(j)\bar{\gamma}(j) = T_1(\bar{\gamma}') - T_2(\bar{\gamma}') - \sum_{j=1}^{J_2-1} \theta(j)\bar{\gamma}'(j)$$

The next key claim is that the equation above applies not only to the spanning posteriors $\bar{\gamma}, \bar{\gamma}' \in \Gamma^B$ but to all pairs of posteriors $\gamma, \gamma' \in \Gamma_{12}^C$. To see this, we repeat the above argument on a larger set of actions,

$$\bar{B}' = \bar{A}' \cup f_1(\gamma) \cup f_1(\gamma'),$$

using precisely the same procedure as before. Repeating the entire remainder of the argument, we apply the Lagrangian Lemma to ensure the existence of multipliers defined by $\eta \in \mathbb{R}^{J-1}$ that produce the corresponding equality for all revealed posteriors, hence in particular for γ and γ':

$$T_1(\gamma) - T_2(\gamma) - \sum_{j=1}^{J_2-1} \eta(j)\gamma(j) = T_1(\gamma') - T_2(\gamma') - \sum_{j=1}^{J_2-1} \eta(j)\gamma'(j).$$

For later purposes it is convenient to rewrite this as,

$$T_1(\gamma) - T_2(\gamma) - \left[T_1(\gamma') - T_2(\gamma')\right] = \sum_{j=1}^{J_2-1} \eta(j) \left[\gamma(j) - \gamma'(j)\right].$$

A key observation is that we can set $\eta = \theta$, so that,

$$T_1(\gamma) - T_2(\gamma) - \left[T_1(\gamma') - T_2(\gamma')\right] = \sum_{j=1}^{J_2-1} \theta(j)\left[\gamma(j) - \gamma'(j)\right]. \tag{62}$$

This follows from the fact that both η and θ work for the set of basis posteriors. For $\bar{\gamma}, \bar{\gamma}' \in \Gamma^B$,

$$\begin{aligned}
T_1(\bar{\gamma}) - T_2(\bar{\gamma}) - \sum_{j=1}^{J_2-1} \theta(j)\bar{\gamma}(j) &= T_1(\bar{\gamma}') - T_2(\bar{\gamma}') - \sum_{j=1}^{J_2-1} \theta(j)\bar{\gamma}'(j); \\
T_1(\bar{\gamma}) - T_2(\bar{\gamma}) - \sum_{j=1}^{J_2-1} \eta(j)\bar{\gamma}(j) &= T_1(\bar{\gamma}') - T_2(\bar{\gamma}') - \sum_{j=1}^{J_2-1} \eta(j)\bar{\gamma}'(j).
\end{aligned}$$

Subtraction yields,

$$\sum_{j=1}^{J_2-1} [\eta(j) - \theta(j)]\left[\bar{\gamma}(j) - \bar{\gamma}'(j)\right] = 0. \tag{63}$$

Since the set Γ^B spans Γ_{12}^C, we know that, given $\gamma, \gamma' \in \Gamma_{12}^C$ there exists weights $\rho(\bar{\gamma})$ and $\rho'(\bar{\gamma}) \in \mathbb{R}$ on $\bar{\gamma} \in \Gamma^B$ with $\sum_{\bar{\gamma}\in\Gamma^B} \rho(\bar{\gamma}) = \sum_{\bar{\gamma}\in\Gamma^B} \rho'(\bar{\gamma}) = 1$ and,

$$\gamma = \sum_{\bar{\gamma}\in\Gamma^B} \rho(\bar{\gamma})\bar{\gamma} \text{ and } \gamma' = \sum_{\bar{\gamma}\in\Gamma^B} \rho'(\bar{\gamma})\bar{\gamma}.$$

Hence,

$$\begin{aligned}
\sum_{j=1}^{J_2-1} [\eta(j) - \theta(j)]\left[\gamma(j) - \gamma'(j)\right] &= \sum_{j=1}^{J_2-1} [\eta(j) - \theta(j)]\left[\sum_{\bar{\gamma}\in\Gamma^B} \left(\rho(\bar{\gamma}) \ - \rho'(\bar{\gamma})\right)\bar{\gamma}\right] \\
&= \sum_{\bar{\gamma}\in\Gamma^B} \left(\rho(\bar{\gamma}) \ - \rho'(\bar{\gamma})\right) \sum_{j=1}^{J_2-1} [\eta(j) - \theta(j)]\,\bar{\gamma} = 0.
\end{aligned}$$

The final line above follows because all terms $\sum_{j=1}^{J_2-1} [\eta(j) - \theta(j)]\,\bar{\gamma}$ on the RHS are equal across $\bar{\gamma} \in \Gamma^B$ (by equation 63) and

$$\sum_{\bar{\gamma}\in\Gamma^B} \left(\rho(\bar{\gamma}) \ - \rho'(\bar{\gamma})\right) = 0.$$

Hence,

$$\sum_{j=1}^{J_2-1} \eta(j)\left[\gamma(j) - \gamma'(j)\right] = \sum_{j=1}^{J_2-1} \theta(j)\left[\gamma(j) - \gamma'(j)\right].$$

Substitution yields,

$$T_1(\gamma) - T_2(\gamma) = T_1(\gamma') - T_2(\gamma') + \sum_{j=1}^{J_2-1} \theta(j)\left[\gamma(j) - \gamma'(j)\right],$$

hence,

$$T_1(\gamma) - T_2(\gamma) - \sum_{j=1}^{J_2-1} \theta(j)\gamma(j) = T_1(\gamma') - T_2(\gamma') - \sum_{j=1}^{J_2-1} \theta(j)\gamma'(j),$$

verifying equation (62).

As before, we can define $\bar{T}_1 : \Gamma^C(\mu_1) \to \mathbb{R}$ as

$$\bar{T}_1(\gamma) = T_1(\gamma) + H_{12}(\gamma') + \theta.\gamma$$

where we define $H_{12}(\gamma')$ as the number $T_2(\gamma') - T_1(\gamma') + \sum_{j=1}^{J_2-1} \theta(j)\gamma'(j)$, and

$$\theta.\gamma = \sum_{j=1}^{J_1-1} \theta(j)\gamma'(j)$$

with $\theta(j) = 0$ for $j > J_2 - 1$. Note that, for $\gamma \in \Gamma_{12}^C$ we have $T_2(\gamma) = \bar{T}_1(\gamma)$, as required. Finally, note that for $i \neq i'$

$$\begin{aligned} K'(\mu_1, Q) &= \sum_{\gamma \in \Gamma(Q)} Q(\gamma)\bar{T}_1(\gamma) - T_1'(\mu_1) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)\left[T_1(\gamma) + H_{12}(\gamma') - \theta.\gamma\right] - \left[T_1(\mu_1) + H_{12}(\gamma') - \theta.\mu_1\right] \\ &= \sum_{\gamma \in \Gamma(Q)} Q(\gamma)T_1(\gamma) - T_1(\mu_1) - \theta.\left[\sum_{\gamma \in \Gamma(Q)} \gamma Q(\gamma) - \mu_1\right] = K(\mu_1, Q), \end{aligned}$$

as required. This completes the proof that

$$K(\bar{\mu}, Q) = \sum_{\gamma \in \Gamma(Q)} Q(\gamma)T(\gamma) - T(\bar{\mu}),$$

for all $\bar{\mu} \in \Gamma$ and $(\bar{\mu}, Q) \in \hat{\mathcal{F}}(\bar{\mu}|K)$ that are uniform over some state space, and with it the sufficiency proof. ■

0.1 UPS and Non-Regular Data

We illustrate a non-regular data set that suggests a path forward to generalizing the necessity aspect of the UPS theorem to cover all data sets, even those that are not regular. Consider the cost function $K(\mu, Q)$ defined as follows. It is the Tsallis cost function (defined below) with $\sigma = 2$ in all cases except when the prior specifies only two states are possible, and the strategy involves a posterior that rules out one of these states. In such cases the cost is infinite.

$$K(\mu, Q) = \begin{cases} \infty & \text{if } |\Omega(\mu)| = 2 \text{ and } \exists \gamma \in \Gamma(Q) \text{ with } |\Omega(\gamma)| = 1 \\ K_2^{TS}(\mu, Q) & \text{otherwise} \end{cases}$$

Note first that the simpler cost function $K_2^{TS}(\mu, Q)$ allows the use of all posteriors from any prior, $\Gamma^C(\mu) = \Gamma(\mu)$, and is UPS. The amended cost function is therefore also UPS, as the only change is that now, for priors such that $|\Omega(\mu)| = 2$, the DM will never choose to become fully informed, thus for such priors, strategies that involve degenerate posteriors do not appear in $\mathcal{F}^C(\mu)$.

However, with the amendment to infinite cost, LIP (Axiom A9) is no longer satisfied: the corresponding behavioral data associated with optimal choices violates the axiom. For priors such that $|\Omega(\mu)| > 2$, one can find an action set such that it is optimal to be fully informed, so that all posteriors are the unit vectors, for a prior that makes (say) three states equally likely. LIP (Axiom A9) requires that one see chosen the corresponding certain posteriors in the associated problem with the prior adjusted so that only two states ex ante possible. However this is not consistent with the infinite cost specified for fully informed posteriors when used from such priors.

Thus, this is an example of a UPS model which does not satisfy LIP (Axiom A9). Note, however, that the data produced by this model also violates regularity. As described above, degenerate posteriors which are used when there are three states in the prior and feasible when there are only two states are not used. Regularity rules out exactly this type of problem, and as a result means that LIP (Axiom A9) is implied by the UPS model.

One might at first think that this cost function produces a non-regular data set that satisfies UPS but not LIP, hence ruling out the possibility that the UPS theorem can be generalized. However this is false. In fact, the data set does not permit of a costly information representation at all. There are decision problems for which optimal choices do not exist. Specifically this is the case when the marginal utility of identifying the true state in a two state problem makes it ever more profitable to arrive at certainty, while the discontinuity in cost makes the limit strategy of discretely lower value, giving rise only to ϵ-optimal strategies.

We conjecture that this is a general phenomenon: that any effort to construct a regular data set based on a UPS model that does not satisfy LIP gives rise instead to a model in which there are action sets such that optimal strategies do not exist. We conjecture also that closedness of the convex function (in the sense of Rockafellar) is necessary and sufficient for this.

Appendix 5: Theorem 1

Theorem 1: *Data set $C \in \mathcal{C}$ with a UPS representation has a Shannon representation if and only if it satisfies IUC.*

In this appendix we go beyond the UPS case and characterize the Shannon function. We take Theorem 4 as established and show that addition of IUC (A1) is equivalent to the UPS representation being of Shannon form. We first show that if C has a Shannon representation, it satisfies IUC: this is a straight forward implication of existing characterizations of optimal strategies. The sufficiency proof is far more involved. Given their centrality we re-state the defining features of basic forms of a decision problem.

Definition 3 *Given $(\mu, A) \in \mathcal{D}$, a decision problem is **basic,** $(\mu, A) \in \mathcal{B}$, if, given $\omega \neq \omega' \in \Omega(\mu)$, there exists $a \in A$ such that $u(a,\omega) \neq u(a,\omega')$. We associate $(\mu, A) \in \mathcal{D}$ with a set of **basic forms** $(\bar{\mu}, A) \in \mathcal{B}(\mu, A) \subset \mathcal{B}$ by:*

1. *Partitioning $\Omega(\mu)$ into L **basic sets** $\left\{\Omega^l(\mu)\right\}_{1\leq l\leq L}$ comprising payoff equivalent states, so*

that, given $\omega \in \Omega^l(\mu)$ and $\omega' \in \Omega^m(\mu)$,

$$l = m \text{ iff } u(a,\omega) = u(a,\omega') \text{ all } a \in A.$$

2. *Labeling all possible states both by equivalence class and in order within each equivalence class:*

$$\Omega(\mu) = \{\omega_i^l \| 1 \le i \le I(l) = |\Omega^l(\mu)| \text{ and } 1 \le l \le L\}.$$

3. *Selecting $\bar{\imath}(l) \in \{1,..,I(l)\}$ all l and defining $\bar{\Omega}(\mu) = \cup_{l=1}^L \omega_{\bar{\imath}(l)}^l$.*

4. *Defining $\bar{\mu} \in \Gamma(\mu)$ by setting:*

$$\bar{\mu}(\omega_i^l) = \begin{cases} \sum_{j=1}^{I(l)} \mu(\omega_j^l) \text{ if } i = \bar{\imath}(l); \\ 0 \text{ if } i \neq \bar{\imath}(l). \end{cases}.$$

A5.1: Necessity

Note that the necessity aspect of theorem 1 can be simplified to the statement that a data set with a Shannon representation must satisfy IUC: one need not condition on a UPS representation, since a Shannon representation is a special form of UPS representation.

Theorem 1: Necessity If data set $C \in \mathcal{C}$ has a Shannon representation, it satisfies IUC (A1).

Proof. Consider data set $C \in \mathcal{C}$ that has a Shannon representation K_κ^S, where $\kappa > 0$ is the Shannon multiplicative parameter. Now consider $(\mu, A) \in \mathcal{D}$ and $(\bar{\mu}, A) \in \mathcal{B}(\mu, A)$ for $\bar{\imath}(l)$, $1 \le l \le L$,

$$C(\mu, A) = \{P \in \mathcal{P}(\mu, A) | \exists \bar{P} \in C(\bar{\mu}, A) \text{ s.t. } P(a|\omega_i^l) = \bar{P}\left[a|\omega_{\bar{\imath}(l)}^l\right] \text{ all } 1 \le i \le I(l),\ 1 \le l \le L\}. \tag{64}$$

To establish that IUC holds, we show that the LHS and RHS sets in (64) are mutual subsets. Note the defining feature, which is that utilities to all actions within a given equivalence class are identical in each equivalence class: for each l and for any $1 \le i, j \le I(l)$,

$$u(a, \omega_i^l) = u(a, \omega_j^l) \equiv u(a, l).$$

To establish (64) we apply known necessary and sufficient conditions for optimality. Matejka and McKay [2015] (their Corollary 1) show that transformed utilities play a key role,

$$z(a, l) = z(a, \omega_i^l) = \exp^{\frac{u(a,\omega_{i\_}^l)}{\kappa}}.$$

The key observation of Matejka and McKay [2015] is that a feasible policy $\lambda \in \Lambda(\mu, A)$ satisfies $\lambda \in \hat{\Lambda}(\mu, A|K_\kappa^S)$ if and only if $\mathbf{P}_\lambda = P$ is a maximizer on $P \in \mathcal{P}(\mu, A)$ of,

$$\sum_{l=1}^{L}\sum_{i=1}^{I(l)} \mu(\omega_i^l)\left(\sum_{a\in A} P(a|\omega_i^l) u(a,\omega_i^l)\right) - \kappa\left[\sum_{l=1}^{L}\sum_{i=1}^{I(l)} \mu(\omega_i^l)\left(\sum_{a\in A} P(a|\omega_i^l)\ln P(a|\omega_i^l)\right) - \sum_{a\in A} P(a)\ln P(a)\right],$$

and,

$$P(a) = \sum_{l=1}^{L} \sum_{i=1}^{I(l)} \mu(\omega_i^l) P(a|\omega_i^l).$$

The necessary (Matejka and McKay [2015]) and sufficient (Caplin *et al.* [2016]) conditions for this are:

$$\sum_{l=1}^{L} \sum_{i=1}^{I(l)} \frac{z(a,\omega_i^l)\mu(\omega_i^l)}{\sum_{b \in A} P(b) z(b,l)} \leq 1, \text{ all } a \in A;$$

with equality for $a \in A$ such that $P(a) > 0$, and,

$$P(a|\omega_i^l) = \frac{P(a)z(a,l)}{\sum_{b \in A} P(b) z(b,l)}.$$

By definition of a Shannon representation,

$$C(\mu, A) = \{P \in \mathcal{P}(\mu, A) | \exists \lambda \in \hat{\Lambda}(\mu, A | K_\kappa^S) \text{ with } P = \mathbf{P}_\lambda\}.$$

To show the set inclusion,

$$C(\mu, A) \subset \{P \in \mathcal{P}(\mu, A) | \exists \bar{P} \in C(\bar{\mu}, A) \text{ s.t. } P(a|\omega_i^l) = \bar{P}(a|\omega_{\bar{\imath}(l)}^l) \text{ all } 1 \leq i \leq I(l),\ 1 \leq l \leq L\},$$

we consider $P \in C(\mu, A)$. Since the data has a Shannon representation, there exists an optimal policy $\lambda \in \hat{\Lambda}(\mu, A|K_\kappa^S)$ with $\mathbf{P}_\lambda = P$ satisfying the optimality conditions. We now define $\bar{P} \in \mathcal{P}(\bar{\mu}, A)$ as above to satisfy the stated condition

$$\bar{P}(a|\omega_{\bar{\imath}(l)}^l) = P(a|\omega_i^l).$$

all $1 \leq i \leq I(l)$, $1 \leq l \leq L$.

Given $a \in A$, we know that

$$\begin{aligned} \bar{P}(a) &= \sum_{l=1}^{L} \bar{\mu}(\omega_{\bar{\imath}(l)}^l) \bar{P}(a|\omega_{\bar{\imath}(l)}^l) = \sum_{l=1}^{L} \sum_{i=1}^{I(l)} \mu(\omega_i^l) P(a|\omega_{\bar{\imath}(l)}^l) \\ &= \sum_{l=1}^{L} \sum_{i=1}^{I(l)} \mu(\omega_i^l) P(a|\omega_i^l) = P(a). \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{l=1}^{L} \frac{z(a,l)\bar{\mu}(\omega_{\bar{\imath}(l)}^l)}{\sum_{b \in A} \bar{P}(b) z(b,l)} &= \sum_{l=1}^{L} \frac{z(a,l) \sum_{i=1}^{I(l)} \mu(\omega_i^l)}{\sum_{b \in A} P(b) z(b,l)} \\ &= \sum_{l=1}^{L} \sum_{i=1}^{I(l)} \frac{z(a,l)\mu(\omega_i^l)}{\sum_{b \in A} P(b) z(b,l)}. \end{aligned}$$

This implies that both of the conditions defining this as data of the form $\mathbf{P}_{\bar{\lambda}} = \bar{P}$ for an optimal strategy $\bar{\lambda} \in \hat{\Lambda}(\bar{\mu}, A|K^S_\kappa)$ are met:

$$\sum_{l=1}^{L} \frac{z(a,l)\bar{\mu}(\omega^l_{\bar{\imath}(l)})}{\sum_{b\in A} P(b)z(b,l)} = \sum_{l=1}^{L}\sum_{i=1}^{I(l)} \frac{z(a,l)\mu(\omega^l_i)}{\sum_{b\in A} P(b)z(b,l)} = 1;$$

with the corresponding weak inequality applying to all actions. In fact we can identify an optimal strategy that produces this data as $\boldsymbol{\lambda}(\bar{P}) \in \hat{\Lambda}(\bar{\mu}, A|K^S_\kappa)$. By Lemma 2.13,

$$\mathbf{P}_{\boldsymbol{\lambda}(\bar{P})} = \bar{P}.$$

Since the data has a Shannon representation, this implies that $\bar{P} \in C(\bar{\mu}, A)$.

Analogous reasoning works in the converse direction. We consider $\bar{P} \in C(\bar{\mu}, A)$ and define $P \in \mathcal{P}(\mu, A)$. Since the data has a Shannon representation, there exists an optimal policy $\bar{\lambda} \in \hat{\Lambda}(\mu, A|K^S_\kappa)$ with $\mathbf{P}_{\bar{\lambda}} = \bar{P}$ satisfying the optimality conditions. We now define $P \in \mathcal{P}(\bar{\mu}, A)$ as above to satisfy the stated condition,

$$P(a|\omega^l_i) = \bar{P}(a|\omega^l_{\bar{\imath}(l)}).$$

all $1 \leq i \leq I(l)$, $1 \leq l \leq L$. Given $a \in A$, we run the above equations in reverse to confirm that unconditional probabilities are not affected,

$$P(a) = \bar{P}(a).$$

By precisely the reverse string of equations we find,

$$\sum_{l=1}^{L}\sum_{i=1}^{I(l)} \frac{z(a,l)\mu(\omega^l_i)}{\sum_{b\in A} P(b)z(b,l)} = \sum_{l=1}^{L} \frac{z(a,l)\bar{\mu}(\omega^l_{\bar{\imath}(l)})}{\sum_{b\in A} \bar{P}(b)z(b,l)},$$

all $a \in A$. Again this implies that both the equality and the inequality conditions defining this as data of the form $\mathbf{P}_\lambda = P$ for an optimal strategy $\lambda \in \hat{\Lambda}(\mu, A|K^S_\kappa)$ an optimal strategy are met. Again we can identify the optimal strategy that produces this data as that it induces in the data, $\boldsymbol{\lambda}(P) \in \hat{\Lambda}(\mu, A|K^S_\kappa)$, so that $P \in C(\bar{\mu}, A)$ as required. This completes the proof of (64), and with it establishes that any data set $C \in \mathcal{C}$ that has a Shannon representation satisfies IUC (A1). ■

A5.2: Lemmas for Sufficiency

We now develop the machinery required to prove sufficiency: if a data set $C \in C$ with a UPS representation satisfies IUC, it has a Shannon representation. There are many distinct aspects to this proof. In what follows, we will let $K \in \mathcal{K}^{UPS}$ be the UPS representation that is known to exist, and let $T : \Gamma \to \mathbb{R}$ be a strictly convex function such that,

$$K(\mu, Q) = \sum_{\gamma\in\Gamma(Q)} Q(\gamma)T(\gamma) - T(\mu)..$$

all $(\mu, Q) \in \mathcal{F}$ such that $Q \in \hat{\mathcal{Q}}(\mu|K)$.

In the first set of lemmas we establish implied symmetry properties. In the second we analyze differentiability and additive separability. We then establish a PDE that characterizes the representation interior to state spaces of dimension 4 or higher. This sets up the proof itself, which analyzes this PDE and considers links between problems of different dimensions. The results display the many different aspects and great power of IUC.

A5.2.1: Symmetry: Definitions and Results

In this subsection we introduce and demonstrate the powerful symmetry implications of IUC (A1). We begin by defining symmetry.

Definition 4 *Beliefs* $\gamma_1, \gamma_2 \in \Gamma$ *are* ***symmetric,***

$$\gamma_1 \sim_\Gamma \gamma_2,$$

if there exists a bijection $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ *such that, for all* $\omega \in \Omega(\gamma_1)$*,*

$$\gamma_1(\omega) = \gamma_2(\sigma(\omega)).$$

Two decision problems are ***symmetric,***

$$(\mu_1, A_1) \sim_{\mathcal{D}} (\mu_2, A_2),$$

if $\mu_1 \sim_\Gamma \mu_2$ *based on bijection* $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ *and there exists a bijection* $\phi : A_1 \to A_2$ *such that,*

$$u(a, \omega) = u(\phi(a), \sigma(\omega)),$$

all $\omega \in \Omega(\gamma_1)$*. Two decision problems with the same prior* $\mu \in \Gamma$ *are* ***equivalent,*** $(\mu, A_1) \equiv_{\mathcal{D}} (\mu, A_2)$*, if there exists a bijection* $\phi : A_1 \to A_2$ *such that,*

$$u(a, \omega) = u(\phi(a), \omega),$$

all $\omega \in \Omega(\mu)$*.*

All three binary relations are symmetric and transitive.

Lemma 5.1: $\sim_\Gamma$, $\sim_{\mathcal{D}}$, and $\equiv_{\mathcal{D}}$ are symmetric and transitive binary relations.

Proof. If $\gamma_1 \sim_\Gamma \gamma_2$, then $|\Omega(\gamma_1)| = |\Omega(\gamma_2)| = J$ and there exists a bijection $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ such that $\gamma_1(\omega) = \gamma_2(\sigma(\omega))$ all $\omega \in \Omega(\gamma_1)$. Given the bijective nature of σ, its inverse σ^{-1} is also bijective, so that,

$$\gamma_1(\sigma^{-1}(\omega')) = \gamma_2(\omega')$$

all $\omega' \in \Omega(\gamma_2)$, This establishes $\gamma_2 \sim_\Gamma \gamma_1$, hence symmetry of $\sim_\Gamma$. With regard to transitivity, note first that $\gamma_1 \sim_\Gamma \gamma_2$ and $\gamma_2 \sim_\Gamma \gamma_3$ imply that $|\Omega(\gamma_1)| = |\Omega(\gamma_3)| = J$ and produce bijections $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ and $\sigma' : \Omega(\gamma_2) \to \Omega(\gamma_3)$. Their bijective nature implies that the composite mapping,

$$\sigma''(\omega) = \sigma' \left[\sigma(\omega)\right],$$

defined on $\omega \in \Omega(\gamma_1)$, is also bijective, establishing $\gamma_1 \sim_\Gamma \gamma_3$ and with it transitivity of $\sim_\Gamma$.

To establish symmetry of $\sim_{\mathcal{D}}$, note that $(\mu_1, A_1) \sim_{\mathcal{D}} (\mu_2, A_2)$ implies existence of bijections $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ and $\phi : A_1 \to A_2$ such that, for all $\omega \in \Omega(\gamma_1)$ and $a \in A_1$:

$$\begin{aligned} \gamma_1(\omega) &= \gamma_2(\sigma(\omega)); \\ u(a, \omega) &= u(\phi(a), \sigma(\omega)). \end{aligned}$$

Since both σ and ϕ are bijective, they have bijective inverses σ^{-1} and ϕ^{-1}. These immediately satisfy the requirements for $(\mu_2, A_2) \sim_{\mathcal{D}} (\mu_1, A_1)$,

$$\begin{aligned} \gamma_1(\sigma^{-1}(\omega')) &= \gamma_2(\omega') \\ u(\phi^{-1}(a'), \sigma^{-1}(\omega')) &= u(a', \omega'), \end{aligned}$$

all $\omega' \in \Omega(\gamma_2)$ and $a' \in A_2$. Hence $(\mu_2, A_2) \sim_{\mathcal{D}} (\mu_1, A_1)$ establishing symmetry of $\sim_{\mathcal{D}}$.

To establish transitivity of $\sim_{\mathcal{D}}$, note that $(\mu_1, A_1) \sim_{\mathcal{D}} (\mu_2, A_2)$ and $(\mu_2, A_2) \sim_{\mathcal{D}} (\mu_3, A_3)$ implies existence of bijections $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2), \sigma' : \Omega(\gamma_2) \to \Omega(\gamma_3)$ and $\phi : A_1 \to A_2$, $\phi' : A_2 \to A_3$, such that,

$$\begin{aligned} \gamma_1(\omega) &= \gamma_2(\sigma(\omega)) \text{ and } \gamma_2(\omega) = \gamma_3(\sigma'(\omega)); \\ u(a, \omega) &= u(\phi(a), \sigma(\omega)) \text{ and } u(a, \omega) = u(\phi'(a), \sigma'(\omega)). \end{aligned}$$

Again their bijective nature implies that the composite mappings,

$$\begin{aligned} \sigma''(\omega) &= \sigma'[\sigma(\omega)] \text{ on } \omega \in \Omega(\gamma_1) : \\ \phi''(a) &= \phi'[\phi(a)] \text{ on } a \in A_1; \end{aligned}$$

are also bijective. Hence $(\mu_1, A_1) \sim_{\mathcal{D}} (\mu_3, A_3)$ establishing transitivity of $\sim_{\mathcal{D}}$.

Note finally that symmetry of $\equiv_D$ follows directly from the bijective nature of the mapping ϕ while transitivity follows likewise from that of ϕ'', completing the proof. ■

Note that, since payoffs are the same in all possible states, equivalent decision problems differ only in that the actions may have distinct payoffs in impossible states (recall that in our formulation an action specifies payoffs in all states, not just the states possible according to the prior). It is a direct implication of existence of a CIR that equivalent choice data is observed for equivalent decision problems.

Lemma 5.2: If C has a CIR with $K \in \mathcal{K}$ and $(\mu, A_1) \equiv_{\mathcal{D}} (\mu, A_2)$ based on bijection $\phi : A_1 \to A_2$ and $P_1(a|\omega) = P_2(\phi(a), \omega)$ all $a \in A_1$ and $\omega \in \Omega(\mu)$, then,

$$P_1 \in C(\mu, A_1) \Longleftrightarrow P_2 \in C(\mu, A_2), \tag{65}$$

Proof. Since $(\mu, A_1) \equiv_{\mathcal{D}} (\mu, A_2)$ based on bijection $\phi : A_1 \to A_2$, there is a one-one and onto mapping $H : \Lambda(\mu, A_1) \to \Lambda(\mu, A_2)$ defined by mapping $\lambda = (\bar{Q}, q_1) \in \Lambda(\mu, A_1)$ to $H(\lambda) = (\bar{Q}, q_2) \in \Lambda(\mu, A_2)$ with,

$$q_2(\phi(a)|\gamma) = q_1(a|\gamma).$$

Given that the distribution of posteriors is identical, so are costs according to $K \in \mathcal{K}$ at $K(\mu, \bar{Q})$. Moreover we know also that $u(a, \omega) = u(\phi(a), \omega)$, so that this mapping preserves expected utility and hence also value,

$$U(\lambda) = U(H(\lambda)).$$

Hence,

$$V(\mu, \lambda|K) = U(\lambda) - K(\mu, \bar{Q}) = U(H(\lambda)) - K(\mu, \bar{Q}) = V(\mu, H(\lambda)|K)$$

The value functions are also therefore the same

$$\hat{V}(\mu, A_1|K) = \hat{V}(\mu, A_2|K).$$

Finally, this implies that optimal strategies are equivalent,

$$\lambda \in \hat{\Lambda}(\mu, A_1|K) \Longleftrightarrow H(\lambda) \in \hat{\Lambda}(\mu, A_2|K). \tag{66}$$

To confirm (65), we need to show that, $P_1 \in C(\mu, A_1)$ if and only if $P_2 \in C(\mu, A_2)$, where $P_2(\phi(a), \omega) = P_1(a|\omega)$ all $a \in A_1$ and $\omega \in \Omega(\mu)$. Note first that, since C has a CIR $K \in \mathcal{K}$, we know $P_1 \in C(\mu, A_1)$ if and only if there exists $\lambda = (\bar{Q}, q_1) \in \hat{\Lambda}(\mu, A_1|K)$ such that $P_1 = \mathbf{P}_\lambda$.By construction, note that,

$$\begin{aligned} P_1(a|\omega) &= \mathbf{P}_\lambda(a|\omega) = \frac{\sum_{\gamma \in \Gamma(\bar{Q})} \bar{Q}(\gamma) q_1(a|\gamma)\gamma(\omega)}{\mu(\omega)} \\ &= \frac{\sum_{\gamma \in \Gamma(\bar{Q})} \bar{Q}(\gamma) q_2(\phi(a)|\gamma)\gamma(\omega)}{\mu(\omega)} = \mathbf{P}_{H(\lambda)}(\phi(a), \omega) = P_2(\phi(a)|\omega). \end{aligned}$$

To complete the proof, note by (66) that $H(\lambda) = (\bar{Q}, q_2) \in \hat{\Lambda}(\mu, A_2|K)$, whereupon since this is a CIR we know that $\mathbf{P}_{H(\lambda)} \in C(\mu, A_2)$. That this is if and only if follows from the bijective nature of ϕ, which allows the argument to be repeated reversing the labels 1 and 2 on action sets and decision problems. ■

While the equivalence of the data from equivalent decision problems is entirely general, the same is not true for symmetric decision problems. The distinction is that these generally involve learning about distinct states, and there is nothing in Axioms A2 through A9 that imposes symmetry. The fact that Compression does imply symmetry requires more insight. First, we note that symmetry of beliefs survives under taking of particular convex combinations.

Lemma 5.3: Consider $\gamma_1, \gamma_2 \in \Gamma$ satisfying $\gamma_1 \sim_\Gamma \gamma_2$ and $\bar{\gamma}_1, \bar{\gamma}_2 \in \Gamma$ with $\Omega(\bar{\gamma}_i) = \Omega(\gamma_i)$ and $\bar{\gamma}_1 \sim_\Gamma \bar{\gamma}_2$ both based on $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$. Given $\alpha \in (0, 1)$, define,

$$\mu_i = \alpha\gamma_i + (1 - \alpha)\bar{\gamma}_i,$$

for $i = 1, 2$. Then $\mu_1 \sim_\Gamma \mu_2$.

Proof. By definition $\gamma_1 \sim_\Gamma \gamma_2$ based on $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$ implies $|\Omega(\gamma_1)| = |\Omega(\gamma_2)| = J$ and,

$$\gamma_1(\omega) = \gamma_2(\sigma(\omega)),$$

all $\omega \in \Omega(\gamma_1)$. Since $\Omega(\bar{\gamma}_i) = \Omega(\gamma_i)$ and $\bar{\gamma}_1 \sim_\Gamma \bar{\gamma}_2$ based on σ, the equivalent holds:

$$\bar{\gamma}_1(\omega) = \bar{\gamma}_2(\sigma(\omega)),$$

all $\omega \in \Omega(\bar{\gamma}_1) = \Omega(\gamma_1)$. Hence,

$$\begin{aligned} \mu_1(\omega) &= \alpha\gamma_1(\omega) + (1-\alpha)\bar{\gamma}_1(\omega) \\ &= \alpha\gamma_2(\sigma(\omega)) + (1-\alpha)\bar{\gamma}_2(\sigma(\omega)) = \mu_2(\sigma(\omega)), \end{aligned}$$

establishing that indeed $\mu_1 \sim_\Gamma \mu_2$ based on σ. ■

Another important observation is that, if two decision problems are symmetric, $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$, their basic versions are symmetric. It is helpful first to show that symmetric problems can have their actions and states aligned in a natural and useful manner.

Lemma 5.4: Consider $(\mu_1, A_1) \in \mathcal{D}$ with $|A_1| = M$ and $|\Omega(\mu_1)| = J$. Then $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$ if and only if, for $i = 1, 2$, one can index all states $\omega_i(j) \in \Omega(\mu_i)$ and all actions $a_i(m) \in A_i$ so that,

$$\begin{aligned} \mu_1(\omega_1(j)) &= \mu_2(\omega_2(j)) \equiv \mu(j); \text{ and} \qquad (67) \\ u(a_1(m), \omega_1(j)) &= u(a_2(m), \omega_2(j)) \equiv u(m, j). \qquad (68) \end{aligned}$$

Proof. If such an indexing exists, then $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$ follows directly from the 1-1 mappings defined by the indices. To establish that $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$ implies existence of such an indexing follows directly by arbitrarily indexing the M actions in A_1 and J states in $\Omega(\mu_1)$ and defining $\mu(j) = \mu_1(\omega_1(j))$ and $u(m, j) = u(a_1(m), \omega_1(j))$. One then uses the bijections ϕ and σ to identify the corresponding elements of A_2 and $\Omega(\mu_2)$:

$$\begin{aligned} \omega_2(j) &= \sigma[\omega_1(j)] \in \Omega(\mu_2) \text{ for } 1 \leq j \leq J; \\ a_2(m) &= \phi[a_1(m)] \in A_2 \text{ for } 1 \leq m \leq M. \end{aligned}$$

It follows directly from the definition of the binary relations that probabilities and utilities are equalized:

$$\begin{aligned} \mu_2(\omega_2(j)) &= \mu_2(\sigma[\omega_1(j)]) = \mu_1(\omega_1(j)) = \mu(j); \\ u(a_2(m), \omega_2(j)) &= u(\phi[a_1(m)], \sigma[\omega_1(j))]) = u(a_1(m), \omega_1(j)) = u(m, j); \end{aligned}$$

establishing the Lemma. ■

Lemma 5.5: If $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$, $(\bar{\mu}_1, A_1) \in \mathcal{B}(\mu_1, A_1)$. and $(\bar{\mu}_2, A_2) \in \mathcal{B}(\mu_2, A_2)$, then,

$$(\bar{\mu}_1, A_1) \sim_\mathcal{D} (\bar{\mu}_2, A_2).$$

Proof. The first step is to apply Lemma 5.4 to index states in $\Omega(\mu_1)$ by j and actions in A_1 by m and to correspondingly index A_2 and $\Omega(\mu_2)$ so that (67) and (68) hold. We then partition $\Omega(\mu_1)$ into its L basic sets $\left\{\Omega^l(\mu_1)\right\}_{1 \leq l \leq L}$ comprising payoff equivalent states, and map initial state labels $1 \leq j \leq J$ to their specific basic set $1 \leq l(j) \leq L$ and then in order within each basic set as $1 \leq i(j) \leq I(l) = |\Omega^l(\mu_1)|$. We refer to the state using both labels as $\omega_1(l(j), i(j))$. Given that we have aligned payoffs and utilities according to the index j, note that the corresponding partitioning and labelling applies also to $\Omega(\mu_2)$ and that the equality of probabilities is preserved,

$$\mu_1(\omega_1(l(j), i(j))) = \mu_2(\omega_2(l(j), i(j))) \equiv \mu(j). \tag{69}$$

With regard to utility, not only is equality preserved,

$$u(a_1(m), \omega_1(l(j), i(j))) = u(a_2(m), \omega_2(l(j), i(j)))$$

but by definition of the equivalence class, this utility is common across $1 \leq i \leq I(l)$, so that,

$$u(a_1(m), \omega_1(l(j), i(j))) = u(a_2(m), \omega_2(l(j), i(j))) \equiv u(m, l(j)). \tag{70}$$

Now define an arbitrary basic version $(\bar{\mu}_1, A_1)$ of (μ_1, A_1) by selecting $\bar{\imath}_1(l)$ for $1 \leq l \leq L$, and an arbitrary basic version $(\bar{\mu}_2, A_2)$ of (μ_2, A_2) by selecting $\bar{\imath}_2(l)$ for $1 \leq l \leq L$. Given $1 \leq l \leq L$ we now compute the corresponding prior probability for each state:

$$\bar{\omega}_1(l) \equiv \omega_1(l, \bar{\imath}_1(l)) \in \Omega(\bar{\mu}_1),$$

as

$$\bar{\mu}_1(\bar{\omega}_1(l)) = \sum_{\{j|l(j)=l\}} \mu_1(\omega_1(l, i(j)).$$

Correspondingly for $\bar{\omega}_2(l) \equiv \omega_2(l, \bar{\imath}_2(l)) \in \Omega(\bar{\mu}_2)$,

$$\bar{\mu}_2(\bar{\omega}_2(l)) = \sum_{\{j|l(j)=l\}} \mu_2(\omega_2(l, i(j)).$$

These are equal by (69),

$$\bar{\mu}_1(\bar{\omega}_1(l)) = \sum_{\{j|l(j)=l\}} \mu_1(\omega_1(l, i(j)) = \sum_{\{j|l(j)=l\}} \mu(j) = \sum_{\{j|l(j)=l\}} \mu_2(\omega_2(l, i(j)) = \bar{\mu}_2(\bar{\omega}_2(l)).$$

Hence the identity mapping on labels l is the bijection that establishes $\bar{\mu}_1 \sim_\Gamma \bar{\mu}_2$.

To complete the proof that $(\bar{\mu}_1, A_1) \sim_\mathcal{D} (\bar{\mu}_2, A_2)$ note that, when we align states by label l and the actions by m equivalence of utilities follows directly from (70),

$$u(a_1(m), \bar{\omega}_1(l)) = u(m, l) = u(a_2(m), \bar{\omega}_2(l))$$

all $1 \leq l \leq L$, proving the Lemma. ■

A key result is that IUC, Axiom A1, implies that all states are equally difficult to learn about, which translates into equivalence of the observed data. This is established in the next lemma.

Lemma 5.6: (Symmetric Data) If $C \in \mathcal{C}$ has a CIR and satisfies IUC, $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$ for bijections $\sigma : \Omega_1 \to \Omega_2$ and $\phi : A_1 \to A_2$, $P_1 \in C(\mu_1, A_1)$, and $P_2(\phi(a)|\sigma(\omega)) = P_1(a, \omega)$ for all $a \in A$ and $\omega \in \Omega(\mu_1)$, then $P_2 \in C(\mu_2, A_2)$.

Proof. Given $(\mu_1, A_1) \sim_\mathcal{D} (\mu_2, A_2)$, we start by repeating the full procedure in Lemma 5.5. We index states in $\Omega(\mu_1)$ by $j \in \{1, \ldots J\}$ and actions in A_1 by $m \in \{1, \ldots M\}$ and correspondingly index A_2 and $\Omega(\mu_2)$ so that (67) and (68) hold. We then partition $\Omega(\mu_1)$ into its L basic sets $\{\Omega^l(\mu_1)\}_{1\leq l\leq L}$ and map initial state labels $1 \leq j \leq J$ to their specific basic set $1 \leq l(j) \leq L$ and then in order within each basic set as $1 \leq i(j) \leq I(l) = |\Omega^l(\mu_1)|$ so that:

$$\begin{aligned} \mu_1(\omega_1(l(j), i(j))) &= \mu_2(\omega_2(l(j), i(j))) \equiv \mu(j); \\ u(a_1(m), \omega_1(l(j), i(j))) &= u(a_2(m), \omega_2(l(j), i(j))) = u(m, l(j)). \end{aligned}$$

For $k = 1, 2$ we then define basic versions $(\bar{\mu}_k, A_k) \in \mathcal{B}(\mu_k, A_k)$ by selecting $\bar{\imath}(l)$ with $1 \leq \bar{\imath}(l) \leq I(l)$, defining $\bar{\omega}_k(l) \equiv \omega_k(l, \bar{\imath}(l)) \in \Omega(\bar{\mu}_k)$, and,

$$\bar{\mu}_k(\bar{\omega}_k(l)) = \sum_{\{j|l(j)=l\}} \mu_k(\omega_k(l, i(j)),$$

all $1 \leq l \leq L$.

One refinement is that we ensure that the labels have particular structure for notational simplicity in what follows. Specifically we let $\bar{L} \leq L$ be the cardinality $\Omega(\bar{\mu}_1)/\Omega(\bar{\mu}_2)$, which by symmetry is equal to that $\Omega(\bar{\mu}_2)/\Omega(\bar{\mu}_1)$. In $\Omega(\bar{\mu}_1)$ we use the first $\bar{L}$ of the L state labels for states in $\Omega(\bar{\mu}_1)/\Omega(\bar{\mu}_2)$, so that their matched states in $\Omega(\bar{\mu}_2)/\Omega(\bar{\mu}_1)$ are also the first $\bar{L}$ in $\Omega(\bar{\mu}_2)$. From Lemma 5.5 we know also that, for all l and m:

$$\begin{aligned} \bar{\mu}_1(\bar{\omega}_1(l)) &= \bar{\mu}_2(\bar{\omega}_2(l)); \\ u(a_1(m), \bar{\omega}_1(l)) &= u(m, l) = u(a_2(m), \bar{\omega}_2(l)); \end{aligned}$$

and hence that $(\bar{\mu}_1, A_1) \sim_{\mathcal{D}} (\bar{\mu}_2, A_1)$.

We now introduce a new problem (μ_3, B) with $\Omega(\mu_3) = \Omega(\bar{\mu}_1) \cup \Omega(\bar{\mu}_2)$ by setting $\mu_3 = \frac{1}{2}(\bar{\mu}_1 + \bar{\mu}_2)$ so that, given $\omega \in \Omega(\mu_3)$:

$$\mu_3(\omega) = \begin{cases} \frac{\bar{\mu}_1(\bar{\omega}_1(l))}{2} \text{ for } \omega = \bar{\omega}_1(l) \text{ for } 1 \leq l \leq \bar{L} \\ \frac{\bar{\mu}_2(\bar{\omega}_2(l))}{2} \text{ for } \omega = \bar{\omega}_2(l) \text{ for } 1 \leq l \leq \bar{L}; \\ \bar{\mu}_1(\bar{\omega}_1(l)) = \bar{\mu}_2(\bar{\omega}_2(l)) \text{ for } L+1 \leq l \leq L. \end{cases}$$

Now define actions $b(m)$ for $1 \leq m \leq M$ by setting their payoffs in states $\omega \in \Omega(\mu_3)$ as:

$$u(b(m), \omega) = \begin{cases} u(a_1(m), \omega) \text{ for } \omega \in \Omega(\bar{\mu}_1)/\Omega(\bar{\mu}_2); \\ u(a_2(m), \omega) \text{ for } \omega \in \Omega(\bar{\mu}_2)/\Omega(\bar{\mu}_1); \\ u(a_1(m), \omega) = u(a_2(m), \omega) \text{ for } \omega \in \Omega(\bar{\mu}_2) \cap \Omega(\bar{\mu}_1). \end{cases}$$

WLOG, set payoffs at zero, $u(b(m), \omega) = 0$, in all other states $\omega \notin \Omega(\mu_3)$. Finally define the choice set of interest.

$$B = \cup_{m=1}^{M} b(m).$$

The idea behind the construction is that $(\bar{\mu}_1, B)$ and $(\bar{\mu}_2, B)$ are both basic versions of (μ_3, B). To see this, we conveniently index the states in $\Omega(\mu_3)$ by s for $1 \leq s \leq L + \bar{L}$. Specifically,

$$\omega_3(s) = \begin{cases} \bar{\omega}_1(s) \text{ for } 1 \leq s \leq L; \\ \bar{\omega}_2(s - L) \text{ for } L+1 \leq s \leq L + \bar{L}. \end{cases}$$

This allows us to partition $\Omega(\mu_3)$ into L basic sets of two forms. There are $\bar{L}$ such sets with 2 elements:

$$\{\omega_3(s), \omega_3(L+s)\} = \{\bar{\omega}_1(s), \bar{\omega}_2(s)\} \text{ for } 1 \leq s \leq \bar{L}.$$

The remaining sets are singletons containing each of the shared elements

$$\omega_3(s) = \bar{\omega}_1(s) = \bar{\omega}_2(s) \text{ for } \bar{L}+1 \leq s \leq L.$$

We can now construct a basic version of (μ_3, B) by selecting $\bar{\omega}_1(s)$ whenever there are two elements. This selection which we denote $i_1^*(l)$ reproduces $(\bar{\mu}_1, B)$ since

$$\frac{\bar{\mu}_1(\bar{\omega}_1(s))}{2} + \frac{\bar{\mu}_2(\bar{\omega}_2(s))}{2} = \bar{\mu}_1(\bar{\omega}_1(s)).$$

Similarly we could select the second element (which we denote $i_1^*(l)$) reproducing $(\bar{\mu}_2, B)$. Therefore both $(\bar{\mu}_1, B)$ and $(\bar{\mu}_2, B)$ belong to $\mathcal{B}(\mu_3, B)$. For future reference, note that $(\bar{\mu}_1, B)$ has the same payoffs in all possible states as $(\bar{\mu}_1, A_1)$, and likewise $(\bar{\mu}_2, B)$ and $(\bar{\mu}_1, A_2)$, which will allow us to link the associated data using Lemma 2.19.

To use the above to establish the lemma, we need to show that, if $P_1 \in C(\mu_1, A_1)$ and $P_2(\phi(a)|\sigma(\omega)) = P_1(a|\omega)$ all $a \in A_1$ and $\omega \in \Omega_1$, then $P_2 \in C(\mu_2, A_2)$. The proof is conceptually simple, but involves a large number of steps as we use symmetry to move from (μ_1, A_1) to $(\bar{\mu}_1, A_1)$ to $(\bar{\mu}_1, B)$ and then to (μ_3, B) and then back from (μ_3, B) to $(\bar{\mu}_2, B)$ to $(\bar{\mu}_2, A_1)$ and finally to (μ_2, A_1). To keep the notation straight, we use the notation from above: for $k = 1, 2$, the $\omega_k(l, i)$ index the states in the original problems $\Omega(\mu_k)$, the $\bar{\omega}_k(l)$ index the states in the basic problems $\Omega(\bar{\mu}_k)$, and finally $\omega_3(s)$ indexes the states in $\Omega(\mu_3)$.

1. IUC links the data in $P_1 \in C(\mu_1, A_1)$ to $\bar{P}_1 \in C(\bar{\mu}_1, A_1)$. Specifically, define $\bar{P}_1 \in \mathcal{P}(\bar{\mu}_1, A_1)$ by
$$\bar{P}_1(a_1(m)|\bar{\omega}_1(l)) = P_1(a_1(m)|\omega_1(l, \bar{\imath}(l))$$
for all $1 \leq l \leq L$, and $1 \leq m \leq M$. IUC implies $\bar{P}_1 \in C(\bar{\mu}_1, A_1)$.

2. We use the fact noted at the end of the last paragraph, that $(\bar{\mu}_1, A_1)$ and $(\bar{\mu}_1, B)$ are constructed to be essentially equivalent in that all actions $b(m) \in B$ have the same payoffs in all possible states $\omega \in \Omega(\bar{\mu}_1)$ as $a_1(m) \in A_1$. Hence Lemma 2.19 applies to ensure that, since $\bar{P}_1 \in C(\bar{\mu}_1, A_1)$ and C has a CIR, the corresponding data is observed as $\bar{P}_1^B \in C(\bar{\mu}_1, B)$,
$$\bar{P}_1^B(b(m)|\bar{\omega}_1(l)) = \bar{P}_1(a_1(m)|\bar{\omega}_1(l)),$$
all $1 \leq l \leq L$ and $1 \leq m \leq M$.

3. Since (μ_1, B) is a basic version of (μ_3, B) based on $i_1^*(l)$, we can apply IUC to identify the corresponding data in $\bar{P}_3^B \in C(\bar{\mu}_3, B)$ satisfying,
$$\bar{P}_3^B(b(m)|\omega_3(s)) = \begin{array}{c} \bar{P}_1^B(b(m)|\bar{\omega}_1(s)) \text{ for } 1 \leq s \leq L; \\ \bar{P}_1^B(b(m)|\bar{\omega}_1(s-L) \text{ for } L+1 \leq s \leq L+\bar{L}. \end{array}$$
for all $1 \leq m \leq M$.

4. Since (μ_2, B) is also a basic version of (μ_3, B). We can then use i_2^* to define
$$\bar{P}_2^B(b(m)|\bar{\omega}_2(l)) = \bar{P}_3^B(b(m)|\omega_3(s)),$$

for all $1 \leq l \leq L$, and $1 \leq m \leq M$. IUC implies $\bar{P}_2^B \in C(\bar{\mu}_2, B)$.

5. We reapply Lemma 2.19 to relate $(\bar{\mu}_2, B)$ to $(\bar{\mu}_2, A_2)$. Let

$$\bar{P}_2(a_2(m)|\bar{\omega}_2(l)) = \bar{P}_2^B(b(m)|\bar{\omega}_2(l)),$$

$1 \leq l \leq L$, and $1 \leq m \leq M$. $\bar{P}_2 \in C(\bar{\mu}_2, A_2)$

6. Finally, IUC relates $(\bar{\mu}_2, A_2)$ to (μ_2, A_2). Let

$$P_2(a_2(m)|\omega_2(l,i)) = \bar{P}_2(a_2(m)|\bar{\omega}_2(l))$$

for all $1 \leq i \leq I(l), 1 \leq l \leq L$, and $1 \leq m \leq M$. IUC implies $P_2 \in C(\mu_2, A_2)$.

Stringing these 6 steps together establishes

$$P_2(a_2(m)|\omega_1(l,i)) = P_1(a_1(m)|\omega_1(l,i))$$

and thereby completes the proof of the Lemma. ■

The final symmetry result we establish is that if $C \in \mathcal{C}$ with a UPS representation satisfies IUC, the underlying strictly convex function $T : \Gamma \longrightarrow \mathbb{R}$ is symmetric in the natural sense.

Definition 5 *Strictly convex function $T : \Gamma \longrightarrow \mathbb{R}$ is **symmetric** if it is equal on symmetric beliefs,*

$$\gamma_1 \sim_\Gamma \gamma_2 \Longrightarrow T(\gamma_1) = T(\gamma_2).$$

Lemma 5.7: (Symmetric Costs) Given $C \in \mathcal{C}$ with a UPS representation satisfying IUC, any function $T : \Gamma \longrightarrow \mathbb{R}$ in a UPS representation $K(Q) = \sum_{\Gamma(Q)} Q(\gamma) T(\gamma)$ must be symmetric.

Proof. Consider $\gamma_1, \gamma_2 \in \Gamma$ satisfying $\gamma_1 \sim_\Gamma \gamma_2$ based on $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_2)$. If $\gamma_1 = \gamma_2$, $T(\gamma_1) = T(\gamma_2)$ trivially. Therefore consider $\gamma_1 \neq \gamma_2$. Consider now a distinct pair $\bar{\gamma}_1 \neq \bar{\gamma}_2 \in \Gamma$ with $\Omega(\bar{\gamma}_1) = \Omega(\gamma_1)$, $\Omega(\bar{\gamma}_2) = \Omega(\gamma_2)$, and $\bar{\gamma}_1 \sim_\Gamma \bar{\gamma}_2$ based on σ. Define two distinct weighted averages,

$$\begin{aligned} \mu_1 &= \frac{3\gamma_1 + \bar{\gamma}_1}{4} \text{ and } \mu_2 = \frac{3\gamma_2 + \bar{\gamma}_2}{4}; \\ \bar{\mu}_1 &= \frac{\gamma_1 + 3\bar{\gamma}_1}{4} \text{ and } \bar{\mu}_2 = \frac{\gamma_2 + 3\bar{\gamma}_2}{4}. \end{aligned}$$

By Lemma 5.3, $\mu_1 \sim_\Gamma \mu_2$ and $\bar{\mu}_1 \sim_\Gamma \bar{\mu}_2$.

By UPS Feasibility Implies Optimality, there exists $(\mu_1, A_1) \in \mathcal{D}$ such that there exists $\lambda_{A1}, \lambda_{B1} \in \hat{\Lambda}(\mu_1, A_1|K)$ with:

$$\begin{aligned} Q_{\lambda_{A1}}(\gamma_1) &= \frac{3}{4} \text{ and } Q_{\lambda_{A1}}(\bar{\gamma}_1) = \frac{1}{4}; \\ Q_{\lambda_{B1}}(\mu_1) &= 1. \end{aligned}$$

Since $C \in \mathcal{C}$ has a UPS representation, we know that the corresponding data is seen, with $\mathbf{P}_{\lambda_{A1}}, \mathbf{P}_{\lambda_{B1}} \in C(\mu_1, A_1)$. We know by Lemma 2.14 that $\mathbf{Q}_{\mathbf{P}_{\lambda_{A1}}} = Q_{\lambda_{A1}}$ and $\mathbf{Q}_{\mathbf{P}_{\lambda_{B1}}} = Q_{\lambda_{B1}}$.

Create the set of actions A_2, as follows. For each $a \in A_1$, create a corresponding $\phi(a) \in A_2$ such that

$$u(a, \omega) = u(\phi(a), \sigma(\omega))$$

for all $\omega \in \Omega(\gamma_1)$. Since $\sigma : \Omega(\gamma_1) \to \Omega(\gamma_1)$ is a bijection by assumption, $\phi : A_1 \to A_2$ is a bijection by construction, $\Omega(\gamma_1) = \Omega(\mu_1)$ and $\Omega(\gamma_2) = \Omega(\mu_2)$ by construction, and $\mu_1 \sim_\Gamma \mu_2$ from above, we have $(\mu_1, A_1) \sim_{\mathcal{D}} (\mu_2, A_2)$.

By Lemma 5.2, defining $P_2(a|\omega) = P_1(\phi^{-1}(a)|\sigma^{-1}(\omega))$ for all $a \in A_1$ and $\omega \in \Omega(\gamma_2)$,

$$P_1 \in C(\mu_1, A_1) \Longleftrightarrow P_2 \in C(\mu_2, A_2).$$

Hence in particular we can find $P_{A2}, P_{B2} \in C(\mu_2, A_2)$ satisfying:

$$\begin{aligned}\mathbf{Q}_{P_{A2}}(\gamma_2) &= \frac{3}{4} \text{ and } \mathbf{Q}_{P_{A2}}(\bar{\gamma}_2) = \frac{1}{4};\\ \mathbf{Q}_{P_{B2}}(\mu_2) &= 1.\end{aligned}$$

Hence by Lemma 2.15, $\exists \lambda_{A2}, \lambda_{B2} \in \hat{\Lambda}(\mu_2, A_2|K)$ with the corresponding properties,

$$\begin{aligned}Q_{\lambda_{A2}}(\gamma_2) &= \frac{3}{4} \text{ and } Q_{\lambda_{A2}}(\bar{\gamma}_2) = \frac{1}{4};\\ Q_{\lambda_{B2}}(\mu_2) &= 1.\end{aligned}$$

We repeat the entire structure of the argument to find $\bar{\lambda}_{Ai}, \bar{\lambda}_{Bi} \in \hat{\Lambda}(\bar{\mu}_i, \bar{A}_i|K)$ for $i = 1, 2$ with the reversed probabilities:

$$\begin{aligned}Q_{\bar{\lambda}_{A1}}(\bar{\gamma}_1) &= Q_{\bar{\lambda}_{B1}}(\bar{\gamma}_2) = \frac{3}{4};\\ Q_{\bar{\lambda}_{A1}}(\gamma_1) &= Q_{\bar{\lambda}_{B1}}(\gamma_2) = \frac{1}{4};\\ \text{and } Q_{\lambda_{A2}}(\bar{\mu}_1) &= Q_{\lambda_{B2}}(\bar{\mu}_2) = 1.\end{aligned}$$

Now consider any UPS representation $K \in \mathcal{K}^{UPS}$ with $K(Q) = \sum_{\Gamma(Q)} Q(\gamma) T(\gamma)$. To establish symmetry of T, we use simultaneous optimality of an inattentive strategy and an attentive strategy allows us to pin down the difference in costs as based on the difference in expected utility. Taking first $\lambda_{A1}, \lambda_{B1} \in \hat{\Lambda}(\mu_1, A_1|K)$, we note that this implies equality of the corresponding net utilities,

$$\frac{3}{4}[\hat{u}(\gamma_1, A_1) - T(\gamma_1)] + \frac{1}{4}[\hat{u}(\bar{\gamma}_1, A_1) - T(\bar{\gamma}_1)] = \hat{u}(\mu_1, A_1). \tag{71}$$

where recall that $\hat{u}(\gamma, A) \equiv \max_{a \in A} \bar{u}(\gamma, a)$.

Taking now $\lambda_{A2}, \lambda_{B2} \in \hat{\Lambda}(\mu_2, A_2|K)$ equality of the corresponding net utilities reduces to,

$$\frac{3}{4}[\hat{u}(\gamma_2, A_2) - T(\gamma_2)] + \frac{1}{4}[\hat{u}(\bar{\gamma}_2, A_2) - T(\bar{\gamma}_2)] = \hat{u}(\mu_2, A_2). \tag{72}$$

Given all the symmetries, note that expected utilities are equivalent:

$$\begin{aligned}\hat{u}(\gamma_1, A_1) &= \hat{u}(\gamma_2, A_2);\\ \hat{u}(\bar{\gamma}_1, A_1) &= \hat{u}(\bar{\gamma}_2, A_2);\\ \hat{u}(\mu_1, A_1) &= \hat{u}(\mu_2, A_2).\end{aligned}$$

Substitution in (71) and (72) and using the equality between them, we derive,

$$\frac{3}{4}T(\gamma_1) + \frac{1}{4}T(\bar{\gamma}_1) = \frac{3}{4}T(\gamma_2) + \frac{1}{4}T(\bar{\gamma}_2). \tag{73}$$

We follow precisely the same logic with respect to the strategies $\bar{\lambda}_{Ai}, \bar{\lambda}_{Bi} \in \hat{\Lambda}(\bar{\mu}_i, \bar{A}_i|K)$ for $i = 1, 2$ to conclude that the corresponding equation holds with reversed weights

$$\frac{1}{4}T(\gamma_1) + \frac{3}{4}T(\bar{\gamma}_1) = \frac{1}{4}T(\gamma_2) + \frac{3}{4}T(\bar{\gamma}_2). \tag{74}$$

Adding up equations (73) and (74) yields equality of sums, while subtracting them yields equality of differences,

$$\begin{aligned}T(\gamma_1) + T(\bar{\gamma}_1) &= T(\gamma_2) + T(\bar{\gamma}_2);\\ T(\gamma_1) - T(\bar{\gamma}_1) &= T(\gamma_2) - T(\bar{\gamma}_2).\end{aligned}$$

Adding these together we finally conclude that $T(\gamma_1) = T(\gamma_2)$, completing the proof. ■

A5.2.2: Directional Derivatives: Basic Results

For the next several sections of the proof, we fix an arbitrary strictly convex function $T : \Gamma \to \mathbb{R}$ for $C \in \mathcal{C}$ with a UPS representation satisfying IUC. In light of Lemma 5.7, it is symmetric, with costs invariant under $\sim$, the equivalence relation on beliefs. We further restrict attention to interior posteriors that place strictly positive probability on a fixed set of underlying states of cardinality 4 or higher. We are particularly interested in those interior posteriors at which this domain-restricted cost function is differentiable.

Definition 6 *We fix a strictly convex and symmetric function $T : \Gamma \to \mathbb{R}$ in a UPS representation. We fix also a set of states $\tilde{\Omega} \subset \Omega$ of cardinality $J \geq 4$, with the states indexed by $1 \leq j \leq J$. We define $\tilde{\Gamma}$ to the the set of posteriors with $\Omega(\gamma) = \tilde{\Omega}$. We define $\tilde{T}$ to be the restriction of the underlying symmetric and strictly convex function on Γ to $\tilde{\Gamma}$,*

$$\tilde{T} : \tilde{\Gamma} \to \mathbb{R},$$

We let $\tilde{\Gamma}' \subset \tilde{\Gamma}$ be posteriors at which $\tilde{T} : \tilde{\Gamma} \to \mathbb{R}$ is differentiable. We let $\tilde{K}(\mu, \lambda) = \sum Q_\lambda(\gamma)\tilde{T}(\gamma)$ be the attention cost function on this limited domain.

The fundamental objects of interest in what follows are certain derivatives of the function $\tilde{T}$ on $\tilde{\Gamma}$. Note that $\tilde{\Gamma}$ does not allow for independent variation in any single state-specific posterior $\gamma(j)$ due to the adding up constraint on probabilities. Hence we use the directional derivatives in what follows. Given convex function $\tilde{T} : \tilde{\Gamma} \to \mathbb{R}$ and $\gamma \in \tilde{\Gamma}$, the directional derivative at $\gamma \in \tilde{\Gamma}$ in direction $y \in \mathbb{R}^J$ is defined as,

$$\tilde{T}'(\gamma|y) = \lim_{\epsilon \downarrow 0} \frac{\tilde{T}(\gamma + \epsilon y) - \tilde{T}(\gamma)}{\epsilon}, \tag{75}$$

if it exists. We use special notation for the directional derivatives of interest.

Definition 7 *Given $\gamma \in \tilde{\Gamma}$ and any pair of states $1 \leq i \neq j \leq J$, we define the* ***one-sided derivative in direction*** *ji, $\tilde{T}_{\overrightarrow{ji}}(\gamma)$, as the directional derivative associated with increasing the ith coordinate and equally reducing the jth:*

$$\tilde{T}_{\overrightarrow{ji}}(\gamma) = \tilde{T}'(\gamma|e_i - e_j) = \lim_{\epsilon \downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon}; \tag{76}$$

where $e_k \in \mathbb{R}^J$ is the vector with its only non-zero element being 1 in the k^{th} coordinate. Where it exists, we define the ***two-sided derivative in direction*** *ji, $\tilde{T}_{(ji)}$, by:*

$$\tilde{T}_{(ji)}(\gamma) = \lim_{\epsilon \to 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon}. \tag{77}$$

In what follows we will use standard results of convex analysis, almost all gathered in Rockafellar's comprehensive treatise. The first such standard result that we translate to our setting establishes existence of one-sided directional derivatives, as well as an inequality concerning one-sided directional derivatives in opposite directions. For completeness, we note also the standard results that a real-valued convex function is continuous on its relative interior, in this case $\tilde{\Gamma}$.

Lemma 5.8: $\tilde{T}$ is continuous on $\gamma \in \tilde{\Gamma}$, and, given $1 \leq i \neq j \leq J$, $\tilde{T}_{\overrightarrow{ji}}(\gamma)$ exists. Moreover,

$$-\tilde{T}_{\overrightarrow{ij}}(\gamma) \leq \tilde{T}_{\overrightarrow{ji}}(\gamma). \tag{78}$$

Proof. Continuity of $\tilde{T}$ on its relative interior is theorem 10.1 in Rockafellar. By (75), given $\gamma \in \tilde{\Gamma}$,

$$\tilde{T}_{\overrightarrow{ji}}(\gamma) = \lim_{\epsilon \downarrow 0} \frac{\tilde{T}(\gamma + \epsilon y) - \tilde{T}(\gamma)}{\epsilon};$$

where $y = e_i - e_j$. Rockafellar theorem 23.1 establishes that, since $\tilde{T} : \mathbb{R}^J \to \bar{\mathbb{R}}$ is convex and $\tilde{T}(\gamma)$ is finite at $\gamma \in \tilde{\Gamma}$, for any $y \in \mathbb{R}^J$, the RHS of (75) is a non-decreasing function of $\epsilon > 0$. Hence $\tilde{T}_{\overrightarrow{ji}}(\gamma)$ exists. With regard to (78), note directly from the definition that,

$$\tilde{T}_{\overrightarrow{ij}}(\gamma) = \lim_{\epsilon \downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_j - e_i)) - \tilde{T}(\gamma)}{\epsilon} = \tilde{T}'(\gamma| - (e_i - e_j)).$$

Theorem 23.1 in Rockafellar establishes with full generality that,

$$-\tilde{T}'(\gamma| - y) \leq \tilde{T}'(\gamma|y).$$

Applying this to $y = e_i - e_j$ completes the proof of the Lemma,

$$-\tilde{T}_{\overrightarrow{ij}}(\gamma) = -\tilde{T}'(\gamma|e_j - e_i) \le \tilde{T}'(\gamma|e_i - e_j) = \tilde{T}_{\overrightarrow{ji}}(\gamma).$$

■

We are particularly interested in posteriors $\gamma \in \tilde{\Gamma}$ at which the inequality (78) is replaced with an equality. The next result shows this to be equivalent to existence of the two-sided derivative. We add also the standard result that differentiability of T implies existence of all 2-sided directional derivatives.

Lemma 5.9: $\tilde{T}_{(ji)}(\gamma)$ exists if and only if (78) holds with equality,

$$-\tilde{T}_{\overrightarrow{ij}}(\gamma) = \tilde{T}_{\overrightarrow{ji}}(\gamma), \tag{79}$$

in which case

$$\tilde{T}_{(ji)}(\gamma) = \tilde{T}_{\overrightarrow{ji}}(\gamma) = -\tilde{T}_{\overrightarrow{ij}}(\gamma) = -\tilde{T}_{(ij)}(\gamma). \tag{80}$$

Moreover, given $\gamma \in \tilde{\Gamma}'$, $\tilde{T}_{(ji)}(\gamma)$ exists all $1 \le i \ne j \le J$.

Proof. Note first that if (79) holds so that $\tilde{T}_{\overrightarrow{ji}}(\gamma) = -\tilde{T}_{\overrightarrow{ij}}(\gamma)$, this corresponds to equality of the limits from the left and right

$$\begin{aligned}\lim_{\epsilon\downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon} &= \tilde{T}_{\overrightarrow{ji}}(\gamma) = -\tilde{T}_{\overrightarrow{ij}}(\gamma) = -\lim_{\epsilon\downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_j - e_i)) - \tilde{T}(\gamma)}{\epsilon} \\ &= -\lim_{\delta=-\epsilon\uparrow 0} \frac{\tilde{T}(\gamma - \delta(e_j - e_i)) - \tilde{T}(\gamma)}{-\delta} = \lim_{\delta\uparrow 0} \frac{\tilde{T}(\gamma + \delta(e_i - e_j)) - \tilde{T}(\gamma)}{\delta}.\end{aligned}$$

It is standard that this implies that the equal left and right limits define the limit itself,

$$\tilde{T}_{(ji)}(\gamma) = \lim_{\epsilon\downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon},$$

establishing equivalence of (79) and existence of $\tilde{T}_{(ji)}(\gamma)$.

Conversely, note that if $\tilde{T}_{(ji)}(\gamma)$ exists,

$$\tilde{T}_{(ji)}(\gamma) = \lim_{\epsilon\longrightarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon} = \lim_{\epsilon\uparrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon} = \tilde{T}_{\overrightarrow{ji}}(\gamma);$$

and,

$$\begin{aligned}\tilde{T}_{(ji)}(\gamma) &= \lim_{\epsilon\longrightarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon} = \lim_{\epsilon\downarrow 0} \frac{\tilde{T}(\gamma + \epsilon(e_i - e_j)) - \tilde{T}(\gamma)}{\epsilon} \\ &= -\lim_{\delta=-\epsilon\uparrow 0} \frac{\tilde{T}(\gamma + \delta(e_j - e_i)) - \tilde{T}(\gamma)}{-\delta} = \lim_{\delta\uparrow 0} \frac{\tilde{T}(\gamma + \delta(e_j - e_i)) - \tilde{T}(\gamma)}{\delta} = -\tilde{T}_{\overrightarrow{ij}}(\gamma).\end{aligned}$$

These equations together verify that (79) holds and also that,

$$\tilde{T}_{(ji)}(\gamma) = \tilde{T}_{\overrightarrow{ji}}(\gamma) = -\tilde{T}_{\overrightarrow{ij}}(\gamma). \tag{81}$$

To complete the proof that (80) holds, note that since $\tilde{T}_{\overrightarrow{ji}}(\gamma) = -\tilde{T}_{\overrightarrow{ij}}(\gamma)$, we know from (79) that $\tilde{T}_{(ij)}(\gamma)$ exists, and therefore that it satisfies the corresponding equality,

$$\tilde{T}_{(ij)}(\gamma) = \tilde{T}_{\overrightarrow{ij}}(\gamma) = -\tilde{T}_{\overrightarrow{ji}}(\gamma). \tag{82}$$

In combination, (82) and (81) imply (80).

With regard to the final clause of the Lemma, that $\tilde{T}_{(ji)}(\gamma)$ exists all $1 \leq i \neq j \leq J$ if $\gamma \in \tilde{\Gamma}'$, Rockafellar theorem 25.2 shows that $\gamma \in \tilde{\Gamma}'$ implies all directional derivatives $\tilde{T}'(\gamma|y)$ exist and are linear in $y = (y(1), ..., y(J))$. Hence they can be written in terms of partial derivatives $\tilde{T}_j(\gamma)$ as,

$$\tilde{T}'(\gamma|y) = \sum_{j=1}^{J} y(j)\tilde{T}_j(\gamma).$$

Hence, given $1 \leq i \neq j \leq J$,

$$-\tilde{T}_{\overrightarrow{ij}}(\gamma) = -\left[\tilde{T}_j(\gamma) - \tilde{T}_i(\gamma)\right] = \tilde{T}_i(\gamma) - \tilde{T}_j(\gamma) = \tilde{T}'(\gamma|e_i - e_j)) = \tilde{T}_{\overrightarrow{ji}}(\gamma),$$

verifying (79) and thereby establishing existence of all 2-sided directional derivatives. ■

Our next preliminary result shows that symmetry of the cost function has implications for directional derivatives. The Lemma specifies the inherited symmetry property precisely.

Lemma 5.10: If $\tilde{T}_{(ji)}(\gamma)$ exists, then for any bijection $\sigma : \{1, .., J\} \to \{1, .., J\}$,

$$\tilde{T}_{(ji)}(\gamma) = \tilde{T}_{(\sigma(j)\sigma(i))}(\gamma^{\sigma}),$$

where,

$$\gamma^{\sigma}(j) = \gamma(\sigma^{-1}(j)).$$

Proof. Suppose that $\tilde{T}_{(ji)}(\gamma)$ exists and note by symmetry of $\tilde{T}$ (Lemma 5.7) and the bijective nature of σ,

$$\tilde{T}(\gamma) = \tilde{T}(\gamma^{\sigma}).$$

Now consider the posterior $\gamma^{\sigma} + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})$ and note that,

$$\tilde{\gamma}(k) = \begin{cases} \gamma^{\sigma}(k) + \epsilon = \gamma(\sigma^{-1}(\sigma(i)) + \epsilon = \gamma(i) + \epsilon \text{ if } k = \sigma(i); \\ \gamma^{\sigma}(k) - \epsilon = \gamma(\sigma^{-1}(\sigma(j)) - \epsilon = \gamma(j) - \epsilon \text{ if } k = \sigma(j); \\ \gamma^{\sigma}(k) \text{ else.} \end{cases}$$

Hence,

$$\gamma^{\sigma} + \epsilon(e_{\sigma(i)} - e_{\sigma(j)}) \sim_{\Gamma} \gamma^{\sigma} + \epsilon(e_i - e_j),$$

so that by the symmetry of $\tilde{T}$,

$$\tilde{T}\left[\gamma^{\sigma} + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})\right] = \tilde{T}(\gamma + \epsilon(e_i - e_j)).$$

Hence,

$$\tilde{T}_{(ji)}(\gamma^\sigma) = \lim_{\epsilon\to 0} \frac{\tilde{T}\left[\gamma^\sigma + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})\right] - \tilde{T}(\gamma^\sigma)}{\epsilon} = \lim_{\epsilon\to 0} \frac{\tilde{T}\left[(\gamma + \epsilon(e_i - e_j))^\sigma\right] - \tilde{T}(\gamma^\sigma)}{\epsilon} = \tilde{T}_{(ji)}(\gamma),$$

establishing the Lemma. ■

A5.2.3: Lagrangians and Directional Derivatives

The following result explains to some extent the relevance of directional derivatives to our approach. It follows from the Lagrangian Lemma.

Lemma 5.11: (Optimality and Directional Derivatives) Suppose $C \in \mathcal{C}$ has a UPS representation with $K \in \mathcal{K}$, and consider $(\mu, A) \in \mathcal{D}$ and $P \in C(\mu, A)$ with $a, b \in \mathcal{A}(P)$ with $\{\bar{\gamma}_P^a, \bar{\gamma}_P^b\} \subset \tilde{\Gamma}$. Then for all pairs of states $1 \le i \ne j \le J$:

1. For $c \in \{a, b\}$, if $\tilde{T}_{(ji)}(\bar{\gamma}_P^c)$ does not exist, $\bar{\gamma}_P^c \in \tilde{\Gamma}\backslash\tilde{\Gamma}'$,

$$-\tilde{T}_{\overrightarrow{ij}}(\bar{\gamma}_P^c) \le u(c,i) - u(c,j) - [\theta(i) - \theta(j)] \le \tilde{T}_{\overrightarrow{ji}}(\bar{\gamma}_P^c); \tag{83}$$

   with at least one inequality strict.

2. For $c \in \{a, b\}$, if $\tilde{T}_{(ji)}(\bar{\gamma}_P^c)$ exists, $\bar{\gamma}_P^c \in \tilde{\Gamma}'$,

$$\tilde{T}_{(ji)}(\bar{\gamma}_P^c) = u(c,i) - u(c,j) - [\theta(i) - \theta(j)]; \tag{84}$$

3. If $a, b \in \mathcal{A}(P)$ are such that $\{\bar{\gamma}_P^a, \bar{\gamma}_P^b\} \subset \tilde{\Gamma}'$,

$$\tilde{N}^a_{(ji)}(\bar{\gamma}_P^a) = \tilde{N}^b_{(ji)}(\bar{\gamma}_P^b). \tag{85}$$

Proof. Since $C \in \mathcal{C}$ has a UPS representation and $\bar{\gamma}_P^a, \bar{\gamma}_P^b \in \Gamma(P)$, we know from Lemma 2.15 that there exists an optimal policy $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K)$ with $\bar{\gamma}_P^a, \bar{\gamma}_P^b \in \Gamma(Q_\lambda)$. Given that $J \ge 3$, we apply Lemma 5.10 to ensure that directional derivatives are invariant to re-indexing states if needed to make that state J is neither i nor j. We now apply the UPS Lagrangian Lemma to the decision problem $(\mu, A) \in \mathcal{D}$ to identify corresponding multipliers $\theta(j)$. Introduce the function $F^c(\gamma)$ on $c \in A$ and $\gamma \in \Gamma(\mu)$ and its supremal value:

$$F^c(\gamma) \equiv \tilde{N}^c(\gamma) - \sum_{k=1}^{J-1} \theta(k)\gamma(k); \tag{86}$$

where,

$$\tilde{N}^c(\gamma) = \sum_{k=1}^{J} u(c,k)\gamma(k) - \tilde{T}(\gamma),$$

Note that for $\gamma \in \tilde{\Gamma}'$, net utility $\tilde{N}^c_{(ji)}(\gamma)$ is well-defined since $\tilde{T}_{(ji)}(\gamma)$ is well-defined, and, since the limit operation changes only posteriors i and j,

$$\lim_{\epsilon\to 0} \frac{u(c,i)[\gamma(i) + \epsilon - \gamma(i)] + u(c,j)[\gamma(j) - \epsilon - \gamma(j)]}{\epsilon} = u(c,i) - u(c,j).$$

With $\tilde{T}_{(ji)}(\gamma)$ well-defined, the same holds for $F^c_{(ij)}(\gamma)$, which we can analogously compute from (86) as,

$$F^c_{(ij)}(\gamma) = \tilde{N}^c_{(ij)}(\gamma) - \theta(i) + \theta(j). \tag{87}$$

The Lagrangian Lemma implies that $\hat{F}$, the supremal value of $F^c(\gamma)$,

$$\hat{F} = sup_{c \in A, \gamma \in \Gamma(\mu)} \left[F^c(\gamma) \right]$$

is achieved by the posteriors associated with any optimal policy. By Lemma 2.14 this means that it is achieved both by setting $(c, \gamma) = (a, \bar{\gamma}^a_P)$ and $(c, \gamma) = \left(b, \bar{\gamma}^b_P\right)$. By Lemma 5.9, if $\tilde{T}_{(ji)}(\nu)$ does not exist, we know that the derivative from the left must be non-negative and from the right non-positive and that they cannot be equal. This corresponds precisely to

$$-\tilde{T}_{\overrightarrow{ij}}(\bar{\gamma}^c_P) + u(c,i) - u(c,j) - [\theta(i) - \theta(j)] \leq 0 \leq \tilde{T}_{\overrightarrow{ji}}((\bar{\gamma}^c_P) + u(c,i) - u(c,j) - [\theta(i) - \theta(j)],$$

confirming (83). Conversely, if $\tilde{T}_{(ji)}(\bar{\gamma}^c_P)$ exists for $c \in \{a, b\}$, this maximization implies that the corresponding derivative $\tilde{N}^c(\bar{\gamma}^c_P)$ must equal zero,

$$\tilde{N}^c_{(ji)}(\bar{\gamma}^c_P) = -\tilde{T}_{(ji)}(\bar{\gamma}^c_P) + u(c,i) - u(c,j) - [\theta(i) - \theta(j)] = 0,$$

confirming (84). To complete the proof, note that when $\{\bar{\gamma}^a_P, \bar{\gamma}^b_P\} \subset \tilde{\Gamma}'$, (84) applies at both posteriors. Hence by (87),

$$\tilde{N}^a_{(ji)}(\bar{\gamma}^a_P) = \theta(i) - \theta(j) = \tilde{N}^b_{(ji)}(\bar{\gamma}^b_P),$$

confirming (85) and establishing the Lemma. ■

A5.2.4: Ratio Sets and Linearity of Posteriors

The next key observation is that we can design decision problems in which the derivatives of net utility are profoundly informative about the cost function. To do this we use decision problems for which IUC places strong restrictions on how posteriors change as the prior changes. These are decision problems with two equivalent states and corresponding posteriors satisfying a ratio condition. To show what IUC implies for these problems, we provide a key extension to the Feasibility implies Optimality Lemma. We show that IUC places strong restrictions on the optimal strategies. In particular, it enables us to move the prior between equivalent states without altering the observed state dependent stochastic choice data. Since that the data does not change, Bayes' rule alone determines how changes in the prior impact the observed posteriors. We apply this in the context of a set of parametrized decision problems (μ_t, A) in which the parameter $t \in [0, 1]$ adjusts the weight between the payoff equivalent states $k \neq l \in \Omega(\gamma)$.

Definition 8 *Given $\alpha \in (0, \infty)$ and two states $1 \leq k \neq l \leq J$, we define the corresponding **ratio set** $\Gamma_{kl}(\alpha) \subset \tilde{\Gamma}$ as the set of posteriors in which α is the ratio between $\gamma(k)$ and $\gamma(l)$:*

$$\Gamma_{kl}(\alpha) = \left\{ \gamma \in \tilde{\Gamma} \,\middle|\, \frac{\gamma(k)}{\gamma(l)} = \alpha \right\}.$$

$\Gamma_{kl}(\alpha)$ is the intersection of $\tilde{\Gamma}$ and a $J-2$ dimensional linear subspace of $\mathbb{R}^J$. It is therefore convex and has dimension $J-2$. Figure 10 depicts $\tilde{\Gamma}$ for $J = 4$. The blue triangle represents $\Gamma_{12}(\alpha)$

for $\alpha = 2$ so that $\gamma \in \tilde{\Gamma}$ if and only if $\gamma(1) = 2\gamma(2)$. In the Figure $\Gamma_{12}(2)$ connects all points with $\gamma(1) = \gamma(2) = 0$ (i.e. the line segment connecting $\gamma(3) = 1$ to $\gamma(4) = 1$) to the point with $\gamma(1) = \frac{2}{3}$ and $\gamma(2) = \frac{1}{3}$.

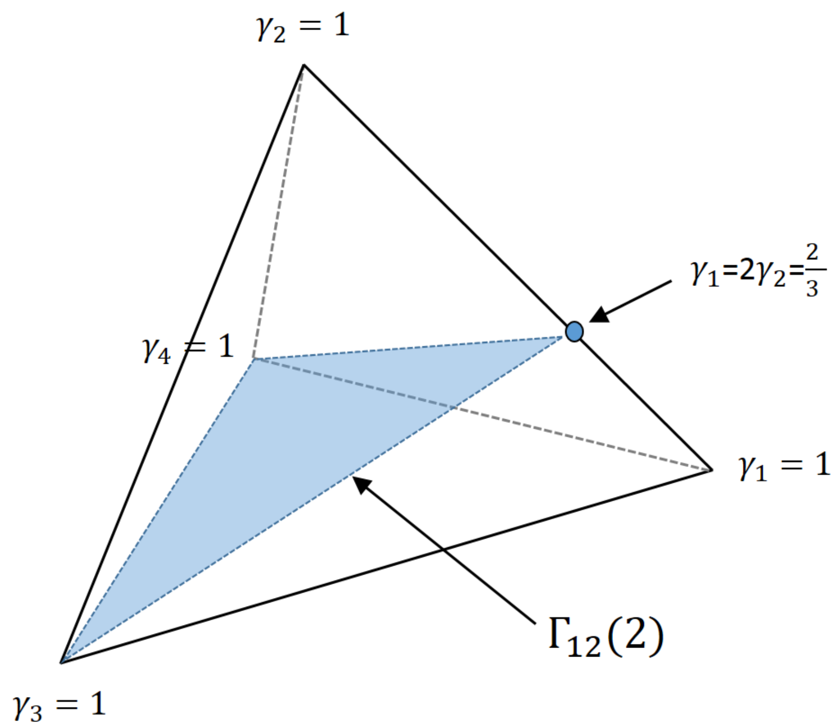

Figure 10

Lemma 5.12: Suppose $C \in \mathcal{C}$ has a UPS representation K and satisfies IUC (A1). Consider $\eta \neq \nu \in \Gamma$ with,

$$\Omega(\eta) = \Omega(\nu) = \{j | 1 \leq j \leq J\},$$

for $J \geq 3$, and $1 \leq k \neq l \leq J$ such that $\eta, \nu \in \Gamma_{kl}(\alpha)$ some $\alpha \in (0, \infty)$,

$$\frac{\eta(k)}{\eta(l)} = \frac{\nu(k)}{\nu(l)} = \alpha. \tag{88}$$

Define the mean belief,

$$\bar{\mu} = \frac{\eta + \nu}{2},$$

and for $t \in [0,1]$ define μ_t, η_t, and ν_t by:

$$\mu_t(j) = \begin{cases} t[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = k; \\ (1-t)[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = l; \\ \bar{\mu}(j) & \text{otherwise;} \end{cases} \tag{89}$$

$$\zeta_t(j) = \left[\frac{\zeta(j)}{\bar{\mu}(j)}\right] \mu_t(j) \text{ for } 1 \leq j \leq J; \tag{90}$$

for $\zeta = \eta, \nu$. Then there exists $a, b \in \mathcal{A}$ with $u(a,k) = u(a,l)$ and $u(b,k) = u(b,l)$ such that,

$$C(\mu_t, \{a,b\}) = \{P_t\};\ \bar{\gamma}^a_{P_t} = \eta_t,\ \text{and}\ \bar{\gamma}^b_P = \nu_t. \tag{91}$$

Specifically, for $\bar{t} = \frac{\bar{\mu}(k)}{\bar{\mu}(k)+\bar{\mu}(l)} \in (0,1)$, $\mu_{\bar{t}} = \bar{\mu}$ and,

$$\zeta_{\bar{t}}(j) = \left[\frac{\zeta(j)}{\bar{\mu}(j)}\right] \bar{\mu}(j) = \zeta(j)$$

for $\zeta = \eta, \nu$.

Proof. Fix $\eta \neq \nu \in \Gamma$ with $\Omega(\eta) = \Omega(\nu) = \{j | 1 \leq j \leq J\}$ for $J \geq 3$ and $k \neq l \in \Omega(\gamma)$ for which (88) holds. Note that this is only possible if $|\Omega(\eta)| \geq 3$ since otherwise (88) implies $\gamma = \eta$. By construction, note from adding the last two equations that,

$$\frac{\eta_t(j) + \nu_t(j)}{2} = \frac{1}{2}\left[\frac{\eta(j) + \nu(j)}{\bar{\mu}(j)}\right] \bar{\mu}_t(j) = \bar{\mu}_t(j).$$

Now consider the case $t = 1$ so that $\mu_1(l) = \eta_1(1) = \eta_1(l) = 0$, $\mu_1(j) = \bar{\mu}(j)$ for $j \neq k, l$, and:

$$\begin{aligned}
\mu_1(k) &= \bar{\mu}(k) + \bar{\mu}(l); \\
\zeta_1(k) &= \zeta(k)\left[\frac{\mu_1(k)}{\bar{\mu}(k)}\right] = \zeta(k)\left[\frac{\bar{\mu}(k) + \bar{\mu}(l)}{\bar{\mu}(k)}\right] = \zeta(k)\left[1 + \frac{\bar{\mu}(l)}{\bar{\mu}(k)}\right] \\
&= \zeta(k) + \zeta(l);
\end{aligned}$$

for $\zeta = \eta, \nu$, where the last line follows (88) and the definition of the mean belief,

$$\frac{\bar{\mu}(l)}{\bar{\mu}(k)} = \frac{\eta(l) + \nu(l)}{\eta(k) + \nu(k)} = \frac{\eta(l)}{\eta(k)} = \frac{\nu(l)}{\nu(k)}.$$

Note that since $\gamma \neq \eta$ and (88) holds, we know that there exists $j \in \Omega(\gamma) \backslash \{k, l\}$ with $\eta(j) \neq \nu(j)$, so that $\eta_1 \neq \nu_1$. Hence we can apply Feasibility Implies Optimality to find action set $\{a, b\}$ such that there is an optimal strategy for the corresponding mean belief,

$$\lambda(1) = (Q_1, q_1) \in \hat{\Lambda}(\mu_1, \{a,b\}|K)$$

in which the only chosen posteriors are η_1 and ν_1, so that $Q_1(\eta_1) = Q_1(\nu_1) = 0.5$. Feasibility Implies Optimality implies also that the deterministic strategy involving each action being chosen deterministically at its corresponding posterior is optimal. In fact. given that the two posteriors are distinct, we know that they are linearly independent, so that the optimal strategy is unique by Lemma 2.4. WLOG,

$$q_1(a|\eta_1) = q_1(b|\nu_1) = 1.$$

We can readily characterize the corresponding revealed posteriors. Since $C \in \mathcal{C}$ has a UPS representation K and the data corresponding to the optimal strategy is observed, we know by Lemmas 2.13 that $C(\mu_1, \{a,b\}) = \{P_1\}$ has the given revealed posteriors,

$$\bar{\gamma}^a_{P_1} = \eta_1 \text{ and } \bar{\gamma}^b_{P_1} = \nu_1.$$

By Lemma 2.19 note that since $\mu_1(l) = 0$, we can set $u(a,l) = u(a,k)$ and $u(b,l) = u(b,k)$ without affecting the observed pattern of SDSC data, P_1.

We now consider decision problem $(\mu_t, \{a,b\}) \in \mathcal{D}$ noting that $(\mu_1, \{a,b\}) \in \mathcal{B}(\mu_t, \{a,b\})$ all $t \in (0,1)$. Since $C(\mu_1, \{a,b\}) = \{P_1\}$ and C satisfies IUC (A1) , we conclude that $P_t \in C(\mu_t, \{a,b\})$ if and only if it satisfies,

$$P_t(c|j) = \begin{cases} P_1(c|j) & \text{if } j \in \Omega(\eta)\backslash\{k,l\}; \\ P_1(c|k) & \text{if } j \in \{k,l\}; \end{cases} \tag{92}$$

for $c \in \{a,b\}$. We now show that $\bar{\gamma}^a_{P_t} = \eta_t$ and that $\bar{\gamma}^b_{P_t} = \nu_t$ all $t \in (0,1)$.

Note first that Bayes' rule combined with (92) implies that, for all $j \in \Omega(\gamma)\backslash\{k,l\}$, and for $c \in \{a,b\}$,

$$\bar{\gamma}^c_{P_t}(j) = \left[\frac{P_t(c|j)}{P_t(c)}\right]\mu_t(j) = \left[\frac{P_1(c)}{P_t(c)}\right]\left[\frac{P_1(c|j)}{P_1(c)}\right]\mu_1(j) = \left[\frac{P_1(c)}{P_t(c)}\right]\bar{\gamma}^c_{P_1}(j),$$

since $\mu_t(j) = \mu_1(j)$. To compute the unconditional choice probabilities $P_1(c)$, note that $\mu_t(j) = \mu_1(j)$ for $j \neq k,l$ and that, given $t \in (0,1)$,

$$\mu_t(k) + \mu_t(l) = \mu_1(k).$$

Hence,

$$\begin{aligned} P_t(c) &= \sum_{j\neq k,l} P(c|j)\mu_t(j) + P(c|k)\mu_t(k) + P(c|l)\mu_t(l) \\ &= \sum_{j\neq k,l} P_1(c|j)\mu_1(j) + P_1(c|k)\left(\mu_t(k) + \mu_t(l)\right) \\ &= \sum_{j\neq k,l} P_1(c|j)\mu_1(j) + P_1(c|k)\mu_1(k) = P_1(c). \end{aligned}$$

Applying this to actions $c = a,b$ separately we derive that for all $j \in \Omega(\gamma)\backslash\{k,l\}$,

$$\begin{aligned} \bar{\gamma}^a_{P_t}(j) &= \bar{\gamma}^a_{P_1}(j) = \eta_1(j) = \eta(j); \\ \bar{\gamma}^b_{P_t}(j) &= \bar{\gamma}^b_{P_1}(j) = \nu_1(j) = \nu(j) \end{aligned}$$

For $j = k$ we make the corresponding substitutions,

$$\begin{aligned} \bar{\gamma}^c_{P_t}(k) &= \left[\frac{P_t(c|k)}{P_t(c)}\right]\mu_t(k) = \left[\frac{P_1(c)}{P_t(c)}\right]\left[\frac{P_1(c|k)}{P_1(c)}\right] t\left[\bar{\mu}(k) + \bar{\mu}(l)\right] \\ &= \left[\frac{P_1(c|k)}{P_1(c)}\right] t\mu_1(k) = t\bar{\gamma}^c_{P_1}(k). \end{aligned}$$

Applying this to actions $c = a,b$ separately we derive,

$$\begin{aligned} \bar{\gamma}^a_{P_t}(k) &= t\bar{\gamma}^a_{P_1}(k) = t\eta_1(k) = \eta_t(k); \\ \bar{\gamma}^b_{P_t}(k) &= t\bar{\gamma}^b_{P_1}(k) = t\nu_1(k) = \nu_t(k); \end{aligned}$$

where the final equalities derive from,

$$\frac{\eta_t(k)}{\eta_1(k)} = \frac{\nu_t(k)}{\nu_1(k)} = \frac{\mu_t(k)}{\mu_1(k)} = t.$$

Finally, using the corresponding substitutions for $j = l$ we derive,

$$\begin{aligned}\bar{\gamma}^c_{P_t}(l) &= \left[\frac{P_t(c|l)}{P_t(c)}\right]\mu_t(l) = \left[\frac{P_1(c)}{P_t(c)}\right]\left[\frac{P_1(c|k)}{P_1(c)}\right](1-t)\left[\bar{\mu}(k)+\bar{\mu}(l)\right] \\ &= \left[\frac{P_1(c|k)}{P_1(c)}\right](1-t)\,\mu_1(k) = (1-t)\bar{\gamma}^c_{P_1}(k).\end{aligned}$$

Applying this to actions $c = a, b$ separately we derive,

$$\begin{aligned}\bar{\gamma}^a_{P_t}(l) &= (1-t)\bar{\gamma}^a_{P_1}(k) = (1-t)\eta_1(k) = \eta_t(l); \\ \bar{\gamma}^b_{P_t}(l) &= (1-t)\bar{\gamma}^b_{P_1}(k) = (1-t)\nu_1(k) = \nu_t(l);\end{aligned}$$

where the final equalities derive from,

$$\frac{\eta_t(l)}{\eta_1(k)} = \frac{\nu_t(l)}{\nu_1(k)} = \frac{\mu_t(l)}{\mu_1(k)} = (1-t).$$

The above concludes the proof that $\bar{\gamma}^a_{P_t} = \eta_t$ and $\bar{\gamma}^b_{P_t} = \nu_t$ all $t \in (0,1)$. To finish the proof of the main clause, note that the result directly holds for $t = 1$. But note that by symmetry that the state labels are irrelevant, so that it also holds for $t = 0$.

The final step in the proof involves direct substitution to show that,

$$\bar{t} = \frac{\bar{\mu}(k)}{\bar{\mu}(k)+\bar{\mu}(l)} \Longrightarrow \mu_{\bar{t}} = \bar{\mu},$$

and correspondingly that,

$$\zeta_{\bar{t}}(j) = \left[\frac{\zeta(j)}{\bar{\mu}(j)}\right]\bar{\mu}(j) = \zeta(j)$$

for $\zeta = \eta, \nu$. ■

A5.2.5: Condition D

Ratio sets are very useful in establishing separability properties, but we need to operate with four posteriors to apply rectangle conditions. The key step is to compare the values of $\tilde{T}_{\overrightarrow{ji}}(\gamma)$ at four different posteriors η^1, η_2, ν_1, $\nu_2 \in \tilde{\Gamma}$ that satisfy powerful regularity conditions.

Definition 9 *Given $\hat{\gamma} \in \tilde{\Gamma}$ and two states $k, l \in J$, we let $\Phi_{kl}(\hat{\gamma})$ denote the set of posteriors that agree with γ on all states $j \neq k, l$*

$$\Phi_{kl}(\hat{\gamma}) = \{\gamma | \gamma(j) = \hat{\gamma}(j), j \neq k, l\}$$

$\Phi_{kl}(\hat{\gamma})$ represents a line in $\tilde{\Gamma}$ through $\hat{\gamma}$ in the direction $e_k - e_l$ where e_j is the unit vector in $\mathbb{R}^J$

with a one in the jth coordinate.

Figure 11 reproduces Figure 10 and adds the point $\hat{\gamma} \in \Gamma_{12}(2)$. The red line in the Figure represents $\Phi_{12}(\hat{\gamma})$. It is the set of points for on which $\gamma(3) = \hat{\gamma}(3)$ and $\gamma(4) = \hat{\gamma}(4)$. This line segment is parallel to the line connecting $\gamma(1) = \gamma(2) = 1$.

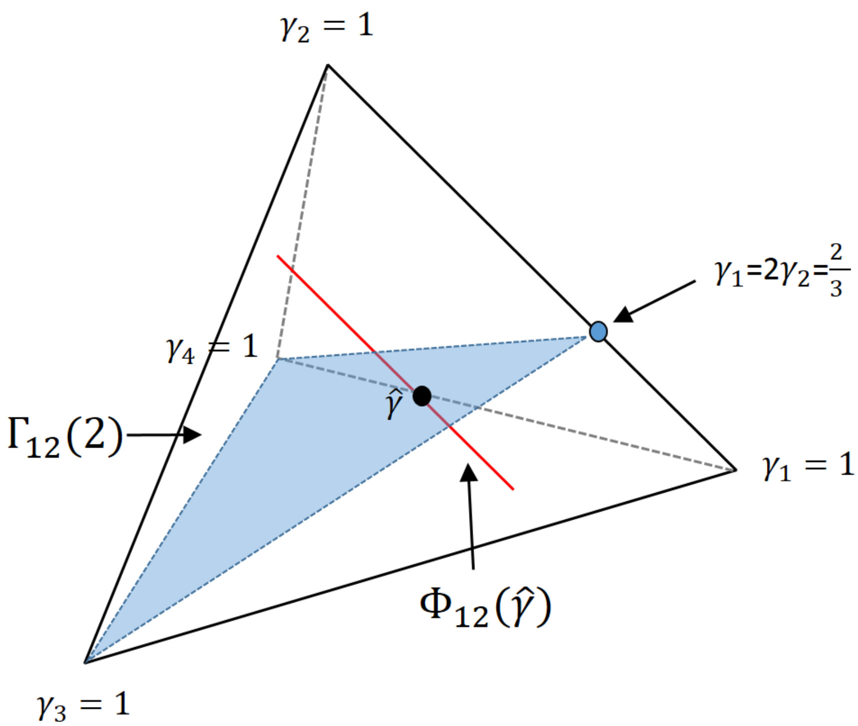

Figure 11

With these definitions we introduce the key sets of four posteriors.

Definition 10 *A set of four distinct posteriors $\eta_1, \eta_2, \nu_1, \nu_2 \in \tilde{\Gamma}$ satisfy* ***condition*** *D if there exist distinct $\alpha(1) \neq \alpha(2) > 0$ and $1 \leq k \neq l \leq J$ such that:*

1. η_1,$\nu_1 \in \Gamma_{kl}(\alpha_1)$,
$$\frac{\eta_1(k)}{\eta_1(l)} = \frac{\nu_1(k)}{\nu_1(l)} = \alpha_1.$$
2. η_2,$\nu_2 \in \Gamma_{kl}(\alpha_2)$,
$$\frac{\eta_2(k)}{\eta_2(l)} = \frac{\nu_2(k)}{\nu_2(l)} = \alpha_2.$$
3. $\eta_2 \in \Phi_{kl}(\eta_1)$, *so that* $\eta_2(j) = \eta_1(j)$ *for* $j \neq k, l$.
4. $\nu_2 \in \Phi_{kl}(\nu_1)$, *so that* $\nu_2(j) = \nu_1(j)$ *for* $j \neq k, l$.

Note that the condition $\alpha_1 \neq \alpha_2$ is imposed since with $\alpha_1 = \alpha_2$, the conditions would give rise to $\eta_1 = \eta_2$ and $\nu_1 = \nu_2$, contrary to the defining feature that these are distinct posteriors.

Figure 12 illustrates Condition D. η_1 and ν_1 both lie in $\Gamma_{12}(\alpha_1)$ and η_2 and ν_3 both lie in $\Gamma_{12}(\alpha_2)$. η_1 and η_2 both on the line segment $\Phi_{12}(\eta_1)$ and ν_1 and ν_2 both on the line segment $\Phi_{12}(\nu_1)$. The key observation is that, since the points in $\Phi_{12}(\eta_1)$ and $\Phi_{12}(\nu_1)$ only differ in their first and second coordinate, the two line segments are parallel, so that the points η_1, η_2, ν_1 and ν_2 form a trapezoid. Below we will use this observation to establish the additive separability of $\tilde{T}_{(ji)}(\eta_1)$.

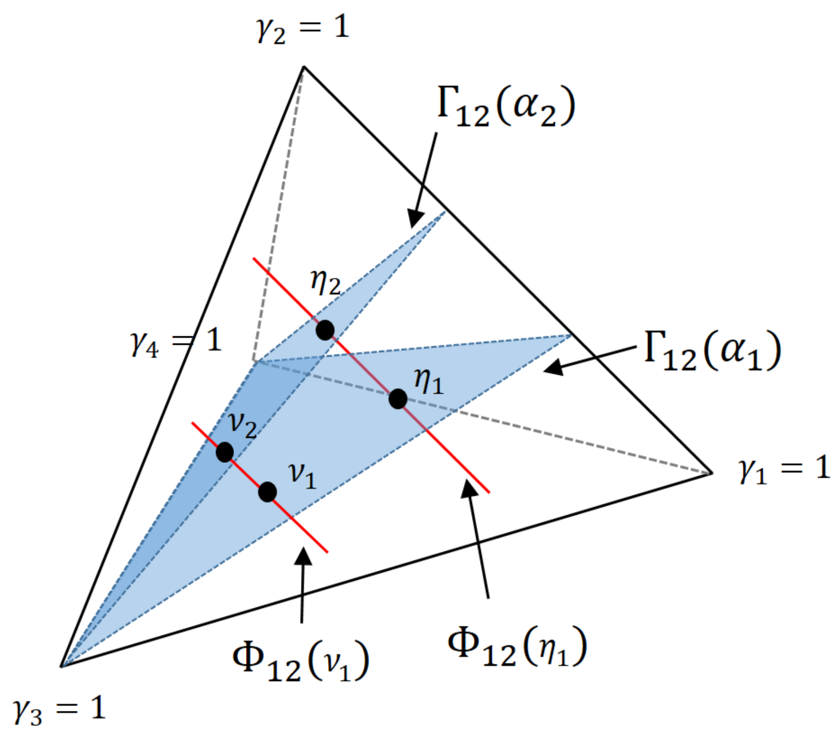

Figure 12

A5.2.6: Equalization of Differences For Two-Sided Directional Derivatives

Our next result relates Condition D, IUC, and the Lagrangian Lemma. If four posteriors η_1, η_2, ν_1, and ν_2 satisfy Condition D and if $\tilde{T}$ is differentiable at each of these points, then IUC and the Lagrangian Lemma relate the change in $\tilde{T}_{(ji)}$ between η_1 and η_2 to that between ν_1 and ν_2.

Lemma 5.13: Suppose $C \in \mathcal{C}$ has a UPS representation $\tilde{T}$ and satisfies IUC (A1). If $\eta_1, \eta_2, \nu_1, \nu_2 \in \tilde{\Gamma}'$ satisfy condition D for some pair of distinct states $1 \leq k \neq l \leq J$, then

$$\tilde{T}_{(ji)}(\eta_1) - \tilde{T}_{(ji)}(\eta_2) = \tilde{T}_{(ji)}(\nu_1) - \tilde{T}_{(ji)}(\nu_2) \tag{93}$$

for all pairs of distinct states $1 \leq i \neq j \leq J$

Proof. Consider η_1,ν_1,η_2 and ν_2 satisfying condition D such that $\tilde{T}_{(ji)}$ exists at all four points. By Lemma 5.12, defining,

$$\bar{\mu} = \frac{\eta_1 + \nu_1}{2},$$

there exists $a, b \in \mathcal{A}$ with $u(a,k) = u(a,l)$ and $u(b,k) = u(b,l)$ such that, for $t \in [0,1]$,

$$C(\mu(t), \{a,b\}) = \{P_t\};\ \bar{\gamma}^a_{P_t} = \eta(t),\ \text{and } \bar{\gamma}^b_{P_t} = \nu(t).$$

where,

$$\mu(t,j) = \begin{cases} t[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = k; \\ (1-t)[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = l; \\ \bar{\mu}(j) & \text{otherwise.} \end{cases}$$

and

$$\begin{aligned} \eta(t,j) &= \left[\frac{\eta_1(j)}{\bar{\mu}(j)}\right] \mu(t,j) \text{ for } 1 \le j \le J; \\ \nu(t,j) &= \left[\frac{\nu_1(j)}{\bar{\mu}(j)}\right] \mu(t,j) \text{ for } 1 \le j \le J; \end{aligned}$$

Where we have placed t as an argument in brackets and avoided the subscript so as to avoid confusion with η_1 and ν_1.

In particular, for $\bar{t}_1 = \frac{\bar{\mu}(k)}{\bar{\mu}(k)+\bar{\mu}(l)} \in (0,1)$,

$$\begin{aligned} \mu(\bar{t}_1) &= \bar{\mu}, \\ \eta(\bar{t}_1) &= \eta_1, \\ \nu(\bar{t}_1) &= \nu_1. \end{aligned}$$

Moreover, since $\tilde{T}$ is differentiable at both η_1 and ν_1, Lemma 5.11, the Optimality and Directional Derivatives Lemma, then implies,

$$\tilde{N}^a_{(ji)}(\eta_1) = \tilde{N}^b_{(ji)}(\nu_1), \tag{94}$$

for all $i, j \in J$.

Now note generally that:

$$\begin{aligned} \eta(t,k) &= \left[\frac{\eta_1(k)}{\bar{\mu}(k)}\right] \bar{\mu}(t,k) \\ &= t\left[\frac{\eta_1(k)}{\bar{\mu}(k)}\right] [\bar{\mu}(k) + \bar{\mu}(l)] \\ &= t\eta_1(k)\left[\frac{\eta_1(k) + \nu_1(k) + \eta_1(l) + \nu_1(l)}{\eta_1(k) + \nu_1(k)}\right] \\ &= t\eta_1(k)\left[\frac{\alpha(1)\eta_1(l) + \alpha(1)\nu_1(l) + \eta_1(l) + \nu_1(l)}{\alpha(1)\eta_1(l) + \alpha(1)\nu_1(l)}\right] \\ &= t\eta_1(k)\left[\frac{\alpha(1)+1}{\alpha(1)}\right]; \end{aligned}$$

where $\alpha_1 = \eta_1(k)/\eta_1(l)$.

Hence if we define $\bar{t}_2$ such that

$$\bar{t}_2 = \frac{\eta_2(k)\alpha_1}{\eta_1(k)(1+\alpha_1)},$$

then,

$$\eta(\bar{t}_2, k) = \eta_2(k).$$

and since $\eta_2(j) \in \Phi_{jk}(\eta_1)$ implies that $\eta(t, j) = \eta_2(j)$ for all $j \neq k, l$,

$$\eta(\bar{t}_2) = \eta_2.$$

Note that since $\eta_2(j) \in \Phi_{jk}(\eta_1)$, $\eta_2(j)$ lies on the line segment connecting $\eta(0)$ and $\eta(1)$, hence $\bar{t}_2 \in [0, 1]$

Similarly, we can show that

$$\nu(t, k) = t\nu_1(k)\left[\frac{\alpha_1 + 1}{\alpha_1}\right].$$

so that

$$\nu(\bar{t}_2, k) = \frac{\eta_2(k)}{\eta_1(k)}\nu_1(k) = \nu_2(k)$$

where the last equality follows from the following line of reasoning:

1. $\eta_2 \in \Phi_{jk}(\eta_1)$ implies $\eta_1(k) + \eta_1(l) = \eta_2(k) + \eta_2(l)$.
2. $\eta_1 \in \Gamma_{12}(\alpha_1)$ and $\eta_2 \in \Gamma_{12}(\alpha_2)$ imply further that $\eta_1(k)[1 + 1/\alpha_1] = \eta_2(k)[1 + 1/\alpha_2]$ for $\alpha_1 = \eta_1(k)/\eta_1(l)$ and $\alpha_2 = \eta_2(k)/\eta_2(l)$.
3. Similarly, $\nu_1(k)[1 + 1/\alpha_1] = \nu_2(k)[1 + 1/\alpha_2]$.
4. So that $\eta_1(k)/\nu_1(k) = \eta_2(k)/\nu_2(k)$ as required.

Therefore given the problem $(\mu(\bar{t}_2), A)$, η_2 is the revealed posterior associated with a and ν_2 is the revealed posterior associated with b. Since $\tilde{T}$ is differentiable at both η_2 and ν_2, Lemma 5.11 then implies that,

$$\tilde{N}^a_{(ji)}(\eta_2) = \tilde{N}^b_{(ji)}(\nu_2) \tag{95}$$

for all $i, j \in J$. Since $\tilde{N}^a(\gamma) = \sum_j u(a, j)\gamma(j) - \tilde{T}(\gamma)$, equations (94) and (95) imply

$$\tilde{T}_{(ji)}(\eta_1) = u(b, i) - u(b, j) - u(a, i) + u(a, j) + \tilde{T}_{(ji)}(\nu_1)$$

and

$$\tilde{T}_{(ji)}(\eta_2) = u(b, i) - u(b, j) - u(a, i) + u(a, j) + \tilde{T}_{(ji)}(\nu_2)$$

Subtracting these two equations yields the desired result. ■

A5.2.7: Monotonicity and Limits

Condition (93) only holds at points at which the two-sided directional derivatives exist. Our next goal is to generalize to one-sided directional derivatives which always exist, thereby extending the result to all sets of posteriors in $\tilde{\Gamma}$ that satisfy Condition D. Our strategy will be to take limits of well-chosen sequences at which the two-sided derivatives exist. Two standard features of the subdifferential map (associating with each posterior $\gamma \in \tilde{\Gamma}$ the full set of corresponding subderivatives of T) allow us to select appropriate sequences. The first is that the sub-differential maps of convex functions are monotone. The second is that they satisfy a form of lower hemi-continuity. The next two lemmas translate these standard results to our setting, starting with the monotonicity lemma.

Lemma 5.14: Given $\gamma \in \tilde{\Gamma}$ and $\epsilon > 0$ such that $\gamma + \epsilon(e_i - e_j) \in \tilde{\Gamma}$,

$$\tilde{T}_{\overrightarrow{ji}}(\gamma + \epsilon(e_i - e_j)) \geq -\tilde{T}_{\overrightarrow{ij}}(\gamma + \epsilon(e_i - e_j)) \geq \tilde{T}_{\overrightarrow{ji}}(\gamma) \tag{96}$$

Proof. The result follows directly from monotonicity properties of the subdifferential maps of convex functions (Rockafellar p. 240). This is particularly simple for one dimensional functions as the general version of the statement that differentiable convex functions have non-decreasing first derivatives. To use in this simple setting, let $\delta > \epsilon > 0$ such that,

$$Y \equiv (\gamma - \delta(e_i - e_j), \gamma + \delta(e_i - e_j)) \subset \tilde{\Gamma},$$

which is possible since $\tilde{\Gamma}$ is relatively open on the line. We then define convex function $G : \mathbb{R} \to \mathbb{R}$ to have value

$$G(\alpha) = \tilde{T}(\gamma + \alpha(e_i - e_j)),$$

on $\alpha \in (-\delta, \delta)$, with its value being infinite elsewhere. It is direct from the definitions that $\tilde{T}_{\overrightarrow{ji}}(\gamma + \alpha(e_i - e_j))$ and the right derivatives of $G(\alpha)$ are equivalent at corresponding points,

$$\begin{aligned} \tilde{T}_{\overrightarrow{ji}}(\gamma + \alpha(e_i - e_j)) &= \lim_{\varsigma \downarrow 0} \frac{\tilde{T}(\gamma + \alpha(e_i - e_j) + \varsigma(e_i - e_j)) - \tilde{T}(\gamma + \alpha(e_i - e_j))}{\varsigma} \\ &= \lim_{\varsigma \downarrow 0} \frac{G(\alpha + \varsigma) - G(\alpha)}{\varsigma} \equiv G'_{+}(\alpha). \end{aligned}$$

Similarly, $-\tilde{T}_{\overrightarrow{ij}}$ and the left derivatives of $G'_{-}(\alpha)$ are equivalent where

$$G'_{-}(\alpha) = -\lim_{\varsigma \uparrow 0} \frac{G(\alpha + \varsigma) - G(\alpha)}{\varsigma}$$

Rockafellar theorem 24.1 establishes monotonicity properties for $G'_{+}(\alpha)$ and $G'_{-}(\alpha)$ when G is convex. In particular for $\epsilon > 0$,

$$G'_{+}(0) \leq G'_{-}(\epsilon) \leq G'_{+}(\epsilon).$$

which, given the equivalence between $\tilde{T}$ and G, establishes (96). ∎

Lemma 5.15: Given any sequence $\{\epsilon_n\}_{n=1}^{\infty}$ with $\epsilon_n > 0$ and $\lim_{n\to\infty} \epsilon_n = 0$,

$$\lim_{n\to\infty} \tilde{T}_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) = \tilde{T}_{\overrightarrow{ji}}(\gamma) \tag{97}$$

Proof. By the directional derivative monotonicity lemma 5.14, given $\{\epsilon_n\}_{n=1}^{\infty}$ with $\epsilon_n > 0$, and $\lim_{n\to\infty} \epsilon_n = 0$, such that $\gamma + \epsilon_n(e_i - e_j) \in \tilde{\Gamma}$, (96) implies that the limit exists and that the inequality survives,

$$\lim_{n\to\infty} \tilde{T}_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) \geq \tilde{T}_{\overrightarrow{ji}}(\gamma). \tag{98}$$

Conversely, Rockafellar 24.5 shows with full generality that given convex function $\tilde{T} : \mathbb{R}^J \to \bar{\mathbb{R}}$, any $\gamma \in \mathbb{R}^J$ at which $\tilde{T}(\gamma)$ is finite, and any sequence $\{\gamma^n\}_{n=1}^{\infty} \to \gamma$ and $y \in \mathbb{R}^J$,

$$\limsup \tilde{T}'(\gamma^n|y) \leq \tilde{T}'(\gamma|y). \tag{99}$$

Defining $\gamma^n = (\gamma + \epsilon_n(e_i - e_j)$ and $y = e_i - e_j$ we get,

$$\lim_{n\to\infty} \tilde{T}_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) \leq \tilde{T}_{\overrightarrow{ji}}(\gamma). \tag{100}$$

Combining (98) with (100) establishes (97), and completes the proof of the Lemma. ■

A5.2.8: Equal Difference Conditions for One-Sided Directional Differences

To complete the transition from equal difference conditions on two-sided to one-sided directional derivatives, we apply standard results on "almost everywhere" differentiability of convex functions on various important subdomains. These are all convex subdomains of $\tilde{\Gamma}$ whose affine hull is of lower dimension, several of which have appeared already in the proof. When thinking about $\tilde{T}$ even on its full domain $\tilde{\Gamma}$, there is a subtlety in the statement. Since $\tilde{\Gamma}$ respects the adding up constraint on probabilities, it has measure zero as a subset of $\mathbb{R}^J$. For that reason the full measure result applies "relative to" $\tilde{\Gamma}$. This is how we state the corresponding result for more general convex subdomains Y. We note also the preservation of one-sided and two-sided directional derivatives on subdomains. The precise formalism is standard.

Definition 11 *Given a non-empty convex set $Y \subset \mathbb{R}^J$, define $\tilde{\Gamma}(Y)$ to be the corresponding subdomain of $\tilde{\Gamma}$,*

$$\tilde{\Gamma}(Y) \equiv Y \cap \tilde{\Gamma};$$

and define $\tilde{T}^Y : \tilde{\Gamma}(Y) \to \mathbb{R}$ with $\tilde{T}^Y(\gamma) \equiv \tilde{T}(\gamma)$ to be the restriction of $\tilde{T}$ to this domain. We define $\tilde{\Gamma}'(Y) \subset \tilde{\Gamma}(Y)$ to be the set on which $\tilde{T}^Y$ is differentiable. Finally, given $\gamma \in \tilde{\Gamma}(Y)$ and $1 \leq i \neq j \leq J$ such that Y for which there exists $\delta > 0$ such that,

$$[\gamma - \delta(e_i - e_j), \gamma + \delta(e_i - e_j)] \subset \tilde{\Gamma}(Y), \tag{101}$$

we define one and two-sided directional derivative $\tilde{T}^Y_{\overrightarrow{ji}}$ and $\tilde{T}^Y_{(ji)}$ in the standard manner.

We now state the key result about convergence of one-sided directional derivatives for appropriately selected sequences of posteriors.

Lemma 5.16: For any non-empty convex set $Y \subset \mathbb{R}^J$, $\tilde{T}^Y$ is almost everywhere differentiable in the relative interior of $\tilde{\Gamma}(Y)$ and $\tilde{\Gamma}(Y) \backslash \Gamma'(Y)$ is of measure zero, and whenever $\tilde{T}^Y_{\overrightarrow{ji}}(\gamma)$ is well-defined,

$$\tilde{T}^Y_{\overrightarrow{ji}}(\gamma) = \tilde{T}_{\overrightarrow{ji}}(\gamma). \tag{102}$$

Proof. Given that Y is non-empty and convex set, $\tilde{T}^Y$ is a proper convex function. Rockafellar theorem 25.5 translates precisely to the fact that $\tilde{\Gamma}'(Y)$ is dense in $\tilde{\Gamma}(Y)$ and its complement $\tilde{\Gamma}(Y)\backslash\tilde{\Gamma}'(Y)$ is of measure zero in the relative interior of $\tilde{\Gamma}(Y)$. The equality $\tilde{T}^Y_{\overrightarrow{ji}} = \tilde{T}_{\overrightarrow{ji}}$ is definitional given the existence of appropriate convergent sequences in the shared domain and equality of the underlying function. ■

Lemma 5.17: Suppose $C \in \mathcal{C}$ has a UPS representation $\tilde{T}$ and satisfies IUC (A1). Given $\{\eta_1, \eta_2, \nu_1, \nu_2\} \subset \tilde{\Gamma}$ satisfying condition D for some pair of distinct states $1 \leq k \neq l \leq J$, then

$$\tilde{T}_{\overrightarrow{ji}}(\eta_1) - \tilde{T}_{\overrightarrow{ji}}(\nu_1) = \tilde{T}_{\overrightarrow{ji}}(\eta_2) - \tilde{T}_{\overrightarrow{ji}}(\nu_2), \tag{103}$$

for all pairs of distinct states $i, j \in \{1, \dots J\}\backslash\{k, l\}$ that are distinct from k and l.

Proof. Consider an arbitrary set of four posteriors $\{\xi_m\} \subset \tilde{\Gamma}$ for $\xi = \eta, \nu$ and $m = 1, 2$ satisfying condition D. We construct four corresponding sequences of posteriors $\{\xi_m^n\}_{n=1}^{\infty}$ that converge to ξ_m as $n \to \infty$. We ensure that at each n (93) holds and that (93) converges to (103). In light of Lemma 5.14 and 4.15, we ensure that the sequence is picked in a special manner ensuring proper convergence. Specifically, let $\Theta(\xi_m)$ be the set containing posteriors that lie within $\frac{1}{n}$ of ξ_m in the direction $\overrightarrow{ji}$:

$$\Theta(\xi_m) = \{\gamma \in \tilde{\Gamma} | \gamma = \xi_m + \lambda(e_i - e_j) \text{ and } \lambda \in (0, \frac{1}{n})\}.$$

for all $\gamma \in \Theta(\xi_m)$.

Note that $\Theta(\xi_m)$ is a convex subset of $\tilde{\Gamma}$, hence satisfies the conditions of the Lemma 5.16, so that $\tilde{T}^{\Theta(\xi_m))}_{(ji)} = \tilde{T}_{(ji)}$ exists for almost all $\lambda \in (0, \frac{1}{n})$. Let $\Lambda_n(\xi_m)$ denote the set of $\lambda \in (0, \frac{1}{n})$ at which the two-sided directional derivative $\tilde{T}_{(ji)}$ exists,

$$\Lambda_n(\xi_m) = \{\lambda \in (0, \frac{1}{n}) | \tilde{T}_{(ij)}(\gamma) \text{ exists at } \gamma = \xi_m + \lambda(e_i - e_j)\}.$$

It follows that $\Lambda_n(\xi_m)$ has measure $\frac{1}{n}$, as does the corresponding intersection,

$$\Lambda(n) \equiv \cap_{\xi=\eta,\nu} \cap_{m=1,2} \Lambda_n(\xi_m).$$

Select $\bar{\lambda}(n) \in \Lambda(n)$ and correspondingly define,

$$\xi_m^n = \xi_m + \bar{\lambda}(n)(e_i - e_j),$$

for $\xi = \eta, \nu$ and $m = 1, 2$.

By construction, for each n, $\xi_m^n \in \tilde{\Gamma}$ for $\xi = \eta, \nu$ and $m = 1, 2$ satisfy condition D. To confirm, note that for $\xi = \eta, \nu$ and $m = 1, 2$, $\xi_m^n(k)$ and $\xi_m(k)$ differ only in coordinates i and j. Hence we know that $\xi_m^n(k) = \xi_m(k)$ and $\xi_m^n(l) = \xi_m(l)$. Given that the ξ_m satisfy condition D, it follows that the ξ_m^n satisfy condition D as well. Since $\tilde{T}_{(ji)}$ exists at each of the ξ_m^n, and since the ξ_m^n satisfy condition D, Lemma 5.13 states

$$\tilde{T}_{(ji)}(\eta_1^n) - \tilde{T}_{(ji)}(\nu_1^n) = \tilde{T}_{(ji)}(\eta_2^n) - \tilde{T}_{(ji)}(\nu_2^n) \tag{104}$$

Now consider each element $\tilde{T}_{(ji)}(\xi_m^n)$ and note that, since $\bar{\lambda}(n) > 0$ and

$$\xi_m^n = \xi_m + \bar{\lambda}(n)(e_i - e_j),$$

we can apply Lemma 5.14 directly to conclude that (96) holds,

$$\tilde{T}_{\overrightarrow{ji}}(\xi_m^n) \geq \tilde{T}_{\overrightarrow{ji}}(\xi_m) \tag{105}$$

Since $\bar{\lambda}(n) \to 0$ we know in addition that $\lim_{n \longrightarrow \infty} \xi_m^n = \xi_m$, so that Lemma 5.15 applies to show that,

$$\lim_{n \to \infty} \tilde{T}_{\overrightarrow{ji}}(\xi_m^n) = \lim_{n \to \infty} \tilde{T}_{(ji)}(\xi_m^n) = \tilde{T}_{\overrightarrow{ji}}(\xi_m). \tag{106}$$

Substituting (106) in (104) establishes (103), and completes the proof. ■

A5.2.9: Propagating Existence of Directional Derivatives

A key result shows how to propagate existence of directional derivatives. If the two-sided directional derivative $\tilde{T}_{(ji)}(\eta)$ exists at a point η, then $\tilde{T}_{(ji)}(\nu)$ exists for all $\nu \in \Gamma_{ji}^{\alpha}$ where $\alpha = \frac{\eta(j)}{\eta(i)}$. The intuition for the result is that this set of v can be linked to η by a problem in which the states i and j are redundant. IUC then allows us to alter the prior on i and j and thereby smoothly shift the the resulting posteriors. These posteriors maintain the original ratio between states j and i: $\frac{\nu(j)}{\nu(i)} = \frac{\nu(j)}{\nu(i)}$. If $\tilde{T}(\eta)$ is smooth in the direction (ji), $\tilde{T}(\nu)$ must also be smooth, if the posteriors are to evolve proportionately.

Before proving the result we establish some additional continuity properties that we can import to our apparatus directly from Rockafellar.

Lemma 5.18: Given $\eta \in \tilde{\Gamma}$ and $1 \leq i \neq j \leq J$ such that $\tilde{T}_{(ji)}(\eta)$ exists, then $\tilde{T}_{(ji)}(\nu)$ exists for all $\nu \in \tilde{\Gamma}$ such that:

$$\frac{\nu(j)}{\nu(i)} = \frac{\eta(j)}{\eta(i)}. \tag{107}$$

Proof. The proof is by contradiction. Choose η such that $\tilde{T}_{(ji)}(\eta)$ exists and suppose that there exists ν satisfying (107) such that $\tilde{T}_{(ji)}(\nu)$ does not exist. By Lemma 5.9 above, this means that $-\tilde{T}_{\overrightarrow{jl}}(\nu) \neq \tilde{T}_{\overrightarrow{lj}}(\nu)$.

By Lemma 5.12 there exists $(\mu, A) \in \mathcal{D}$ with $A = \{a, b\}$ such that η is the revealed posterior related to a and ν is the revealed posterior related to b and the states i and j are redundant. The lemma then guarantees that given the parameterized set of problems (μ_t, A) where

$$\mu_t(k) = \begin{cases} t\,[\mu(i) + \mu(j)] & \text{for } k = i; \\ (1-t)\,[\mu(i) + \mu(j)] & \text{for } k = j; \\ \mu(k) & \text{otherwise;} \end{cases}$$

for $t \in (0,1)$, η_t is the revealed posterior for action a and ν_t is the revealed posterior for action b and

$$\eta_t(j) = \frac{\eta(j)}{\mu(j)}\mu_t(j) \qquad \text{and} \qquad \nu_t(j) = \frac{\nu(j)}{\mu(j)}\mu_t(j)$$

Let $\bar{t}$ be defined by $\mu_{\bar{t}} = \mu$.

By the Lagrangian Lemma, for each t, there exists $\theta_t \in \mathbb{R}^{J-1}$ such that

$$\tilde{N}^a(\eta_t) - \sum_{k=1}^{J-1} \theta_t(k)\eta_t(k) = \tilde{N}^b(\nu_t) - \sum_{k \neq j} \theta_t(k)\nu_t(k) \geq \tilde{N}^c(\gamma) - \sum_{k=1}^{J-1} \theta(k)\gamma(k)$$

for all $\gamma \in \Gamma$ and $c = \{a, b\}$.

Since $\tilde{T}_{(ji)}(\eta)$ exists, the Optimal Directional Derivative Lemma (4.11) tells us that

$$\tilde{T}_{(ji)}(\eta) = u(a,i) - u(a,j) - \theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) \tag{108}$$

and since $\tilde{T}_{(ji)}(\nu)$ does not exist, Lemma 5.11 implies that,

$$-\tilde{T}_{\overrightarrow{ij}}(\nu) - u(b,i) + u(b,J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) \leq 0 \leq \tilde{T}_{\overrightarrow{ji}}(\nu) - u(b,i) + u(b,J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j),$$

with one of these two inequalities strict. Without loss of generality suppose

$$\tilde{T}_{\overrightarrow{ji}}(\nu) - u(b,i) + u(b,J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) = \Delta > 0. \tag{109}$$

Define now $Y \subset \mathbb{R}^J$ as all vectors η_t,

$$Y = \{\eta_t \in \tilde{\Gamma} | t \in [0,1]\},$$

noting that, since $\tilde{\Gamma}(Y) = Y$ since $Y \subset \tilde{\Gamma}$. Lemma 5.16 implies that $\tilde{T}^Y$ is differentiable for almost all η_t, so that Lemma 5.9 implies that the two-sided directional derivative,

$$\tilde{T}^Y_{(ji)}(\eta_t) = \tilde{T}_{(ji)}(\eta_t),$$

also exists for almost all $\eta_t \in Y$.

Now consider a sequence $\eta_{t(n)} \to \eta$ such that $t(n) > \bar{t}$ and $\tilde{T}_{(ji)}(\eta_{t(n)})$ exists. Lemma 5.15 implies that

$$\lim_{t(n) \to \bar{t}} \tilde{T}_{(ji)}(\eta_{t(n)}) = \tilde{T}_{(ji)}(\eta).$$

Therefore there exists $t(m) \neq \bar{t}$ such that $\tilde{T}_{(ji)}(\eta_{t(m)}) \in (\tilde{T}_{(ji)}(\eta), \tilde{T}_{(ji)}(\eta) + \Delta)$. Given that $\tilde{T}_{(ji)}(\eta_{t(m)})$ exists,

$$\tilde{T}_{(ji)}(\eta_{t(m)}) = u(a,i) - u(a,j) - \theta_{t(m)}(i) - \theta_{t(m)}(j). \tag{110}$$

Hence, with $\tilde{T}_{(ji)}(\eta_{t(m)}) - \tilde{T}_{(ji)}(\eta) \in (0, \Delta)$, we can subtract the right-hand sides of (108) from (110) to conclude that,

$$\theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) \in (0, \Delta),$$

so that,

$$-\theta_{t(m)}(i) - \theta_{t(m)}(j) < -\theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) + \Delta. \tag{111}$$

Applying now the Optimal Directional Derivative Lemma (4.11) to $\nu_{t(m)}$ we conclude that,

$$-\tilde{T}_{\overrightarrow{ij}}(\nu_{t(m)}) \leq u(b,i) - u(b,j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) \leq \tilde{T}_{\overrightarrow{ji}}(\nu_{t(m)}).$$

Substitution of (111) thereupon yields,

$$-\tilde{T}_{\overrightarrow{ij}}(\nu_{t(m)}) \leq u(b,i) - u(b,j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) < u(b,i) - u(b,j) - \theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) + \Delta = \tilde{T}_{\overrightarrow{ji}}(\nu)$$

But $t(m) > \bar{t}$, so that the Lemma 5.14 implies directly that $-\tilde{T}_{\overrightarrow{ij}}(\nu_{t(m)}) > \tilde{T}_{\overrightarrow{ji}}(\nu)$. This contradiction establishes the result. ■

A5.2.10: Weak Form Additive Separability

In this section we establish a weak form of additive separability. The basic observation is that (93) is very close to the rectangle condition for this form of additive separability. The difference is that $\eta_1, \eta_2, \nu_1, \nu_2 \in \tilde{\Gamma}$ satisfying condition D form a trapezoid, not a rectangle. We rectify this problem by deforming $\tilde{\Gamma}$.

Definition 12 *Given $M \in \mathbb{N}$, define Z^M as the strictly positive vectors summing to strictly below 1,*

$$Z^M = \{(z_1, \dots z_M) \in R^M | z_m > 0 \text{ all } m \text{ and } \sum_m z_m < 1\},$$

and define $X = (0,1) \times Z^{J-2}$.

We now deform $\tilde{\Gamma}$ to create rectangles. To simplify the notation, we start with arbitrary states k and l but then re-order the states so that $k = 1$ and $l = J$. Given Lemma 5.7, this is without loss of generality. With this naming convention we will suppress the dependence of the function Ψ on k and l in the following definition.

Definition 13 *Define $\Psi : \tilde{\Gamma} \to X$ with $\Psi(\gamma) = x \in X$ where:*

$$x(j) = \begin{cases} \frac{\gamma(j)}{\gamma(j)+\gamma(J)} & \text{for } j = 1; \\ \gamma(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases}$$

The next Lemma points out that Ψ is bijective.

Lemma 5.19: $\Psi : \tilde{\Gamma} \to X$ is bijective.

Proof. The mapping Ψ can be constructed as the combination of two mappings each of which we show to be bijective. The first maps the $J-1$ dimensional simplex $\tilde{\Gamma}$ to Z^{J-1} by dropping the coordinate $\gamma(J)$. Given $\gamma \in \tilde{\Gamma}$ define $\Psi_1(\gamma) \in Z^{J-1}$ by:

$$\Psi_1(\gamma(j)) = \gamma(j) \text{ for } 1 \leq j \leq J-1.$$

The second function divides $\gamma(1)$ by,

$$\gamma(1) + \gamma(J) = 1 - \sum_{m=2}^{J-1} \gamma(m).$$

Technically, given $z \in Z^{J-1}$ define $\Psi_2(z) \in X$ by,

$$\Psi_2(z(j)) = \begin{cases} \frac{z(j)}{1-\sum_{m=2}^{J-1} z(m)} & \text{for } j = 1; \\ z(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases}$$

Clearly $\Psi_1 : \tilde{\Gamma} \to Z^{J-1}$ is bijective. With regard to Ψ_2, note that if that $z_1, z_2 \in Z^{J-1}$ both map to $x \in X$, it is immediate that $z_1 = z_2$. Hence $\Psi_2 : Z^{J-1} \to X$ is injective. To show that it is also surjective, given $x \in X$, define $h(x) \in Z^{J-1}$ by,

$$h(x(j)) = \begin{cases} \left[1 - \sum_{m=2}^{J-1} x(m)\right] x(1) & \text{if } j = 1 \\ x(j) & \text{for } 2 \leq j \leq J-1. \end{cases}$$

We now consider $\Psi_2(h(x)) \in X$. By construction, this satisfies:

$$\Psi_2(h(x(j))) = \begin{cases} \frac{h(x(j))}{1-\sum_{m=2}^{J-1} h(x(m))} & \text{for } j = 1; \\ x(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases}$$

where

$$\frac{h(x(j))}{1 - \sum_{m=2}^{J-1} h(x(m))} = \frac{\left[1 - \sum_{m=2}^{J-1} x(m)\right] x(1)}{\left[1 - \sum_{m=2}^{J-1} x(m)\right]} = x(1).$$

Hence $\Psi_2(h(x)) = x$ so that Ψ_2 is surjective. Given that it is also injective, it is bijective.

To complete the proof, we now show that $\Psi = \Psi_2 \circ \Psi_1$ is the composition of these mappings:

$$\begin{aligned} \Psi &= \Psi_2(\Psi_1(\gamma)) = \begin{cases} \frac{\Psi_1(\gamma(j))}{1-\sum_{m=2}^{J-1} \Psi_1(\gamma(m))} & \text{for } j = 1; \\ \Psi_1(\gamma(j)) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases} \\ &= \begin{cases} \frac{\gamma(j)}{1-\sum_{m=2}^{J-1} \gamma(m)} & \text{for } j = 1; \\ \gamma(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases} \end{aligned}$$

This completes the proof that Ψ is bijective. ■

The next lemma shows that, in this space, condition D transforms into a rectangle condition on X.

Lemma 5.20: Given $\eta_1, \eta_2, \nu_1, \nu_2 \in \tilde{\Gamma}$ that satisfy condition D for states $k = 1$ and $j = J$, the elements $x_1, x_2, y_1, y_2 \in X$ such that $x_m = \Psi(\eta_m)$ and $y_m = \Psi(\nu_m)$ for $m = 1, 2$, form a rectangle:

$$\begin{aligned} x_1(1) &= x_2(1) \text{ and } y_1(1) = y_2(1); \\ x_1(j) &= y_1(j) \text{ and } x_2(j) = y_2(j); \text{ for } 2 \leq j \leq J-1. \end{aligned}$$

Proof. Consider $x_1, x_2, y_1, y_2 \in X$ such that $x_m = \Psi(\eta_m)$ and $y_m = \Psi(\upsilon_m)$ for $m = 1, 2$. By condition D and the definition of Ψ, for $2 \leq j \leq J-1$ and $m = 1, 2$,

$$x_m(j) = \Psi(\eta_m(j)) = \eta_m(j) = \nu_m(j) = \Psi(\nu_m(j)) = \nu_m(j).$$

Note also that,

$$\begin{aligned} x_1(1) - x_2(1) &= \frac{\eta_1(1)}{\eta_1(1)+\eta_1(J)} - \frac{\eta_2(1)}{\eta_2(1)+\eta_2(J)} \\ &= \frac{\frac{\eta_1(1)}{\eta_1(J)}}{\frac{\eta_1(1)}{\eta_1(J)}+1} - \frac{\frac{\eta_2(1)}{\eta_2(J)}}{\frac{\eta_2(1)}{\eta_2(J)}+1} = \frac{\alpha}{\alpha+1} - \frac{\alpha}{\alpha+1} = 0 \end{aligned}$$

Similarly,

$$y_1(1) - y_2(1) = \frac{\frac{\nu_1(1)}{\nu_1(J)}}{\frac{\nu_1(1)}{\nu_1(J)}+1} - \frac{\frac{\nu_1(1)}{\nu_1(J)}}{\frac{\nu_1(1)}{\nu_1(J)}+1} = 0,$$

completing the proof. ■

With this we are in position to establish our first version of additive separability.

Lemma 5.21: Suppose $C \in \mathcal{C}$ has a UPS representation and satisfies IUC (A1). Then, given $2 \leq i \neq j \leq J-1$, $\tilde{T}_{\overrightarrow{ji}}(\gamma)$ is additively separable in $\left[\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right]$ and $\{\gamma(j) |\ 2 \leq j \leq J-1\}$ in that there exists $\boldsymbol{A} : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and $\boldsymbol{B} : \mathbb{R}^{J-2} \longrightarrow \mathbb{R}$ such that,

$$\tilde{T}_{\overrightarrow{ji}}(\gamma) = \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2), ..., \gamma(J-1)\right)$$

Proof. We consider any four posteriors $\eta_1, \eta_2, \nu_1, \nu_2 \in \tilde{\Gamma}$ that satisfy Condition D for states $k = 1$ and $l = J$. We now define $x_1, x_2, y_1, y_2 \in X$ by $x_m = \Psi(\eta_m)$ and $y_m = \Psi(\upsilon_m)$ for $m = 1, 2$. We transfer the directional derivatives to this space by defining the function $\mathcal{T} : X \to \mathbb{R}$ by,

$$\mathcal{T}(x) \equiv \tilde{T}_{\overrightarrow{ji}}(\Psi^{-1}(x)),$$

using the bijective function $\Psi : \tilde{\Gamma} \to X$ introduced above.

Note that the space X is of the cross-product form $X = X_A \times X_B$ with $X_A = (0, 1)$ and $X_B = Z^{J-2}$. A standard condition for such an arbitrary function $f : X \longrightarrow \mathbb{R}$ on such a space to be additively,

$$f(a, b) = f_1(a) + f_2(b)$$

is that the rectangle conditions are satisfied: given $a_1, a_2 \in X_A$ and $b_1, b_2 \in X_B$,

$$f(a_1, b_1) - f(a_2, b_1) = f(a_1, b_2) - f(a_2, b_2).$$

To confirm, pick arbitrary $\left(\bar{a}, \bar{b}\right) \in X_A \times X_B$ and note that for any $(a, b) \in X_A \times X_B$,

$$f(a, b) = f(a, \bar{b}) + f(\bar{a}, b) - f\left(\bar{a}, \bar{b}\right),$$

which is of the additively separable form for $f_1(a) = f(a, \bar{b}) - f\left(\bar{a}, \bar{b}\right)$ and $f_2(b) = f(\bar{a}, b)$.

Since $\eta_1, \eta_2, \upsilon_1$, and ν_2 satisfy Condition D, Lemma 5.13 states,

$$\tilde{T}_{\overrightarrow{ji}}(\eta_1) - \tilde{T}_{\overrightarrow{ji}}(\nu_1) = \tilde{T}_{\overrightarrow{ji}}(\eta_2) - \tilde{T}_{\overrightarrow{ji}}(\nu_2)$$

By the definition of $\mathcal{T}$ we have,

$$\mathcal{T}(x_1) - \mathcal{T}(y_1) = \mathcal{T}(x_2) - \mathcal{T}(y_2)$$

By Lemma 5.20, $x_m = \Psi(\eta_m)$ and $y_m = \Psi(\upsilon_m)$ for $m = 1, 2$, form a rectangle:

$$\begin{aligned} x_1(1) &= x_2(1) \equiv a_1 \text{ and } y_1(1) = y_2(1) \equiv a_2; \\ x_1(j) &= y_1(j) \text{ and } x_2(j) = y_2(j); \text{ for } 2 \leq j \leq J-1. \end{aligned}$$

Define $b_m \in Z^{J-2}$ for $m = 1, 2$ by,

$$b_m(j) = x_m(j+1) \text{ for } 2 \leq j \leq J-1,$$

substitution yields the rectangle condition,

$$\mathcal{T}(a_1, b_1) - \mathcal{T}(a_2, b_1) = \mathcal{T}(x_1) - \mathcal{T}(y_1) = \mathcal{T}(x_2) - \mathcal{T}(y_2) = \mathcal{T}(a_1, b_2) - \mathcal{T}(a_2, b_2).$$

It follows that $\mathcal{T}$ is additively separable between $a \in X_A = (0, 1)$ and $b \in X_B = Z^{J-2}$

$$\mathcal{T}(a, b) = \boldsymbol{A}(a) + \boldsymbol{B}(b)$$

In the final step, we use Ψ to move from $\mathcal{T}$ to $\tilde{T}_{\overrightarrow{ji}}$. Given $x = \Psi(\gamma)$,

$$\begin{aligned} \tilde{T}_{\overrightarrow{ji}}(\gamma) &= \mathcal{T}(\Psi^{-1}(\gamma)) = \mathcal{T}(x) = \boldsymbol{A}(x(1)) + \boldsymbol{B}(x(2), .., x(J-1)) \\ &= \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1) + \gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2), .., \gamma(J-1)\right), \end{aligned}$$

completing the proof. ■

A5.2.11: Strong Form Additive Separability

In this section we establish a stronger form of additive separability relying on already established symmetry and differentiability properties of the $\tilde{T}$ function.

Lemma 5.22: Suppose $C \in \mathcal{C}$ has a UPS representation and satisfies IUC (A1). If $\gamma \in \tilde{\Gamma}'$, then, given $\gamma \in \tilde{\Gamma}'$ and $1 < i \neq j < J$, there exists $\boldsymbol{B} : \mathbb{R}^{J-2} \longrightarrow \mathbb{R}$ such that

$$\tilde{T}_{(ji)}(\gamma) = \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right) \tag{112}$$

Proof. We arbitrarily order states, fix states 1 and J, and consider distinct states $2 \leq i \neq j \leq J-1$. By Lemma 5.9, $\gamma \in \tilde{\Gamma}'$ implies $\tilde{T}_{(ji)}(\gamma)$ exists. We set $i = 2$ and $j = 3$. Given Lemma 5.7, this is without loss of generality.

Applying Lemma 5.21

$$\tilde{T}_{(32)}(\gamma) = \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1) + \gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right).$$

By Lemma 5.9 we also know that,

$$\tilde{T}_{(23)}(\gamma) = -\tilde{T}_{(32)}(\gamma) = -\boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) - \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right). \tag{113}$$

Define the mapping $\sigma : \{1, .., J\} \longrightarrow \{1, .., J\}$ that permutes elements 2 and 3:

$$\sigma(k) = \begin{cases} 3 \text{ if } k = 2; \\ 2 \text{ if } k = 3; \\ k \text{ otherwise.} \end{cases}$$

Defining $\gamma^\sigma \in \tilde{\Gamma}$ as the correspondingly permuted posterior, $\gamma^\sigma(j) = \gamma(\sigma^{-1}(j))$. Lemma 5.10 then that, since $\tilde{T}_{(ji)}(\gamma)$ exists,

$$\tilde{T}_{(23)}(\gamma) = \tilde{T}_{(32)}(\gamma^\sigma)$$

Directly by Lemma 5.21,

$$\begin{aligned} \tilde{T}_{(32)}(\gamma^\sigma) &= \boldsymbol{A}\left(\frac{\gamma^\sigma(1)}{\gamma^\sigma(1)+\gamma^\sigma(J)}\right) + \boldsymbol{B}\left(\gamma^\sigma(2), \gamma^\sigma(3), ..., \gamma^\sigma(J-2), \gamma^\sigma(J-1)\right) \\ &= \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(3), \gamma(2), ..., \gamma(J-2), \gamma(J-1)\right). \end{aligned} \tag{114}$$

Since both equal $\tilde{T}_{(23)}(\gamma)$ we know that the right-hand sides of (113) and (114) are equal,

$$2\boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) = -\boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right) - \boldsymbol{B}\left(\gamma(3), \gamma(2), ..., \gamma(J-2), \gamma(J-1)\right) \tag{115}$$

By assumption $\tilde{T}_{(32)}(\gamma)$ exists so that by Lemma 5.18 it also exists for all η such that $\eta(2)/\eta(3) = \gamma(2)/\gamma(3)$, including all at which,

$$\rho \equiv \frac{\eta(1)}{\eta(1)+\eta(J)} > 0$$

takes arbitrary values while $\eta(k) = \gamma(k)$ for all $k \neq 1, J$, which by construction differ from i, j. Hence (115) must hold for all $\rho > 0$. Since the right-hand side of the equation is independent of ρ, $\boldsymbol{A}(\rho)$ is independent of ρ,

$$\boldsymbol{A}(\rho) = \bar{\boldsymbol{A}} \in \mathbb{R}.$$

Hence we can add $\bar{\boldsymbol{A}}$ to $\boldsymbol{B}$ and normalize to $\boldsymbol{A}(x) = 0$, completing the proof. ■

In the proceeding, the has been no guarantee that there is a single $\boldsymbol{B}$ that works for all pairs of states. In the next lemma we further restrict the functional dependence of the two-sided directional derivative, and in the process show that there exists a single function $\bar{\boldsymbol{B}}$ that characterizes this derivative.

Lemma 5.23: Suppose $C \in \mathcal{C}$ has a UPS representation and satisfies IUC (A1), then there exists $\bar{\boldsymbol{B}} : (0,1) \times (0,1) \to \mathbb{R}$ such that, given $\gamma \in \tilde{\Gamma}'$, and states $1 \leq i \neq j \leq J$,

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)). \tag{116}$$

Proof. Given arbitrarily fixed states 1 and J with $J \geq 4$, Lemma 5.22 establishes that if we consider distinct states $i = 2$ and $j = 3$, there exists $\boldsymbol{B} : \mathbb{R}^{J-2} \longrightarrow \mathbb{R}$ such that, given $\bar{\gamma} \in \tilde{\Gamma}'$,

$$\tilde{T}_{(32)}(\gamma) = \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right)$$

If $J = 4$, then $\boldsymbol{B}$ has only two arguments and is of the desired form,

$$\tilde{T}_{(ji)}(\bar{\gamma}) = \boldsymbol{B}\left(\bar{\gamma}(i), \bar{\gamma}(j)\right) \equiv \boldsymbol{B}\left(\bar{\gamma}(2), \bar{\gamma}(3)\right) \equiv \bar{\boldsymbol{B}}\left(\bar{\gamma}(2), \bar{\gamma}(3)\right).$$

By the symmetry Lemma 5.10, this same function applies regardless of how we label states, completing the proof for $J = 4$.

$$\tilde{T}_{(ji)}(\bar{\gamma}) = \bar{\boldsymbol{B}}(\bar{\gamma}(i), \bar{\gamma}(j))$$

If $J > 4$ we again arbitrarily fixed states 1 and J, and consider state $s \neq i, j$ with $2 \leq s \leq J-1$. Hence by Lemma 5.22 and Lemma 5.10, we can transpose posteriors 1 and s without changing the form of the function, so that,

$$\tilde{T}_{(ji)}(\gamma) = \boldsymbol{B}(\gamma(2), .., \gamma(s-1), \gamma(1), \gamma(s+1), ..., \gamma(J-2), \gamma(J-1)). \tag{117}$$

Raising $\gamma(s)$ and reducing $\gamma(J)$ has no effect on the right hand side of (117), hence no effect on the RHS of (112) so that $\boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right)$ is independent of $\gamma(s)$. Proceeding in this matter for all $s \neq \{i, j\}$, we have

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)),$$

where $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ is the common value. To complete the proof, note again that by the symmetry lemma, the same function applies regardless of how we label the states, completing the proof.

Note that the function $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ is pinned down only for $\gamma \in \tilde{\Gamma}'$ and not the full domain $(0,1) \times (0,1)$. However we know that it is pinned down on a dense subset of this space, so that it is natural to think of using a limit operation to fill out the function. The next Lemma establishes that this can be done in an unambiguous manner, and characterizes the one-sided directional derivative. ■

Lemma 5.24: There exists $\bar{\boldsymbol{B}} : (0,1) \times (0,1) \to \mathbb{R}$ such that, given $\gamma \in \tilde{\Gamma}$,

$$\tilde{T}_{\overrightarrow{ji}}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)). \tag{118}$$

Proof. Where $\tilde{T}_{(ji)}(\gamma)$, exists, Lemma 5.9 shows that it is equal to $\tilde{T}_{\overrightarrow{ji}}(\gamma)$. Hence the function defined in (116) is of the appropriate form for $\gamma \in \tilde{\Gamma}'$. What is left is to establish that we can define $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ that equals $\tilde{T}_{\overrightarrow{ji}}(\gamma)$ on $\gamma \in \tilde{\Gamma} \backslash \tilde{\Gamma}'$.

Consider $\gamma \in \tilde{\Gamma} \backslash \tilde{\Gamma}'$, and consider any sequence $\{\gamma_n\}_{n=1}^{\infty}$ with $\gamma_n = \gamma + \epsilon_n(e_i - e_j)$ such that $\tilde{T}_{(ji)}(\gamma_n)$ exists for all n and $\epsilon_n \downarrow 0$. To see that such a sequence must exist, let $Y(\gamma, i, j) = \{x \in \mathbb{R}^J | x(k) = \gamma(k) \text{ for all } k \neq i, j\}$. $Y(\gamma, i, j)$ is a convex set, and $\tilde{T}^{Y(\gamma,i,j)} : \tilde{\Gamma}(Y(\gamma, i, j)) \to \mathbb{R}$ is the restriction of $\tilde{T}$ to $Y(\gamma, i, j) \cap \tilde{\Gamma}$. Lemma 5.16 states that $\tilde{T}^{Y(\gamma,i,j)}$ is almost everywhere differentiable in the relative interior of $\tilde{\Gamma}(Y(\gamma, i, j))$, and that $\tilde{T}^{Y(\gamma,i,j)}_{(ij)} = \tilde{T}_{(ji)}$. We can therefore select the sequence $\{\gamma_n\}_{n=1}^{\infty}$ from $\tilde{\Gamma}(Y(\gamma, i, j))$.

As $\tilde{T}_{(ji)}(\gamma_n)$ exists, Lemma 5.23 implies,

$$\tilde{T}_{(ji)}(\gamma_n) = \bar{\boldsymbol{B}}(\gamma_n(i), \gamma_n(j)).$$

Lemma 5.15 then ensures that,

$$\lim_{n\to\infty} \tilde{T}_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) = \tilde{T}_{\overrightarrow{ji}}(\gamma).$$

We therefore define $\bar{\boldsymbol{B}}(\gamma)$ on $\gamma \in \tilde{\Gamma}\backslash\tilde{\Gamma}'$ as,

$$\bar{\boldsymbol{B}}(\gamma) \equiv \lim_{n\to\infty} \bar{\boldsymbol{B}}(\gamma_n(i), \gamma_n(j)) = \tilde{T}_{\overrightarrow{ji}}(\gamma), \tag{119}$$

By construction we know that $\tilde{T}_{\overrightarrow{ji}}(\gamma) = \bar{\boldsymbol{B}}(\gamma)$ on the full domain, and that it is of the form $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ on $\gamma \in \tilde{\Gamma}\backslash\tilde{\Gamma}'$. Equation (119) implies that this extends to the limit points, completing the proof of (118) and with it the Lemma. ■

Note that the function $\bar{\boldsymbol{B}} : (0,1) \times (0,1) \to \mathbb{R}$ as introduced above allows for certain jumps at posteriors at the two sided directional derivatives fail to exist. In further characterizing the implications of Compression, such cases will be ruled out.

A5.2.12: Full Additive Separability

We have now established that directional derivatives at any posterior depends only on the probabilities of the two involved states. We now establish that the corresponding function can be defined based on a fixed function of each probability alone. This is what we refer to as full additive separability. The result is connected with a triangular pattern in two-sided directional derivatives. Given $\gamma \in \tilde{\Gamma}$ we know that $\tilde{T}_{(ji)}(\gamma), \tilde{T}_{(ik)}(\gamma), \tilde{T}_{(jk)}(\gamma)$ all exist, and furthermore that they are interdependent,

$$\tilde{T}_{(ji)}(\gamma) = \tilde{T}_{(jk)}(\gamma) + \tilde{T}_{(ki)}(\gamma). \tag{120}$$

In Lemma 5.25 we show that this relationship rests only on existence of any two of these three two-sided directional derivatives. The lemma also uses the negative inverse feature of these directional derivatives to point to the method for identifying the appropriate form of the function that generates the sought after representation.

Lemma 5.25 Given $\gamma \in \tilde{\Gamma}$, suppose that there exist three distinct indices $1 \le i, j, k \le J$ such that $\tilde{T}_{(ki)}(\gamma)$ and $\tilde{T}_{(kj)}(\gamma)$ both exist. Then $\tilde{T}_{(ji)}(\gamma)$ exists and,

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(k)) - \bar{\boldsymbol{B}}(\gamma(j), \gamma(k)). \tag{121}$$

Proof. Given Lemma 5.7, we may take $i = 1$, $j = 2$ and $k = 3$ without loss of generality.

Given $\gamma \in \tilde{\Gamma}$, define the set X by

$$X \equiv \left\{ x \in \mathbb{R}^2 | x_1, x_2 > 0 \text{ and } x_1 + x_2 < 1 - \sum_{l \neq i,j,k} \gamma(l) \right\}. \tag{122}$$

Define $\eta(x) \in \tilde{\Gamma}$

$$[\eta(x)](l) = \begin{cases} x_1 \text{ if } l = 1; \\ x_2 \text{ if } l = 2; \\ 1 - \sum_{l \neq i,j,k} \gamma(l) - x_1 - x_2 \text{ if } l = 3; \\ \gamma(l) \text{ otherwise} \end{cases} \tag{123}$$

Finally, define $H : X \to \mathbb{R}$ by

$$H(x) = \tilde{T}(\eta(x)) \tag{124}$$

Note that

$$\eta(\gamma(1), \gamma(2)) = \gamma$$

Note also that, given $\tilde{T}_{(32)}(\gamma)$ exists,

$$\begin{aligned} \tilde{T}_{(32)}(\gamma) &= \lim_{\epsilon \to 0} \frac{\tilde{T}(\gamma + \epsilon(e_2 - e_3)) - \tilde{T}(\gamma)}{\epsilon} = \lim_{\epsilon \to 0} \frac{\tilde{T}(\eta(\gamma(1), \gamma(2) + \epsilon)) - \tilde{T}(\eta(\gamma(1), \gamma(2)))}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{H(\gamma(1), \gamma(2) + \epsilon) - H(\gamma(1), \gamma(2))}{\epsilon} \end{aligned}$$

Hence,

$$\tilde{T}_{(32)}(\gamma) = H_2(\gamma(1), \gamma(2)). \tag{125}$$

Analogously,

$$\tilde{T}_{(31)}(\gamma) = H_1(\gamma(1), \gamma(2)).$$

Since both partials exist, note from Rockafellar theorem 25.1 that H is differentiable at γ and from theorem 25.2 that the directional derivative function $H'(\gamma|y)$ is linear in direction $y \in \mathbb{R}^2$. Hence the directional derivative in direction $e_1 - e_2$ is the difference between the partials,

$$H'(\gamma|e_1 - e_2) = H_1(\gamma(1), \gamma(2)) - H_2(\gamma(1), \gamma(2)) = \tilde{T}_{(31)}(\gamma) - \tilde{T}_{(32)}(\gamma). \tag{126}$$

To complete the proof of (121), note directly from the definitions that $H'(\gamma|e_2 - e_1)$ is equal to $\tilde{T}_{(ji)}(\gamma)$,

$$\begin{aligned} H'(\gamma|e_1 - e_2) &= \lim_{\epsilon \to 0} \frac{H(\gamma(1) + \epsilon, \gamma(2) - \epsilon) - H(\gamma(1), \gamma(2))}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{\tilde{T}(\gamma + \epsilon(e_j - e_i)) - \tilde{T}(\gamma)}{\epsilon} \equiv \tilde{T}_{(ji)}(\gamma) \end{aligned} \tag{127}$$

Setting the right-hand sides of (126) and (127) to equality establishes that

$$\tilde{T}_{(ji)}(\gamma) = \tilde{T}_{(ki)}(\gamma) - \tilde{T}_{(kj)}(\gamma).$$

In light of Lemma 5.24 this completes the proof of (121). ■

Lemma 5.25 points the way to a possible method for expressing $\tilde{T}_{(ji)}(\gamma)$ in a fully additively separable manner. Given $k \neq i, j$ and $\bar{x} \in (0, 1 - \gamma(i) - \gamma(j))$, Lemma 5.25 implies

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \bar{x}) - \bar{\boldsymbol{B}}(\gamma(j), \bar{x})$$

so that $\tilde{T}_{(ji)}(\gamma)$ is additively separable in $\gamma(i)$ and $\gamma(j)$. The complication in establishing this form of additive separability is that the requirement that $\bar{x} < 1 - \gamma(i) - \gamma(j)$ means that that no single $\bar{x}$ works for all $\gamma \in \tilde{\Gamma}$. In the following, we establish additive separability on a subset of $\tilde{\Gamma}$ and then drive the value of $\bar{x}$ down to zero to establish additive separability on the whole of $\tilde{\Gamma}$.

Lemma 5.26: Given $\epsilon \in (0, 0.5)$, there exists $x(\epsilon) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ and a full measure set $I(\epsilon) \subset (0, 1-\epsilon)$ such that, for any distinct states $1 \le i, k \le J$, given $\gamma \in \tilde{\Gamma}$ with $\gamma(k) = x(\epsilon)$, $\tilde{T}_{(ik)}(\gamma)$ exists whenever $\gamma(i) \in I(\epsilon)$.

Proof. Pick distinct states $1 \le i, k \le J$ and $\epsilon \in (0, 0.5)$. Let $Y(k, \epsilon) = \{z \in \mathbb{R}^J | z(k) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)\}$ denote the set of vectors in $\mathbb{R}^J$ for which $z(k) \in (\frac{\epsilon}{4}, \frac{\epsilon}{2})$ and focus on posteriors with $\gamma(k)$ so restricted,

$$\tilde{\Gamma}(Y(k, \epsilon)) = \left\{\gamma \in \tilde{\Gamma} | \gamma(k) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)\right\}. \tag{128}$$

Per the general prescription, define the restricted function $\tilde{T}^{Y(k,\epsilon)} : \tilde{\Gamma}(Y(k, \epsilon)) \to \mathbb{R}$, and note for arbitrary indices $1 \le j \ne l \le J$,

$$\tilde{T}^{Y(k,\epsilon)}_{\overrightarrow{lj}}(\gamma) = \tilde{T}_{\overrightarrow{lj}}(\gamma),$$

given that there is suitable variation of the posterior in all directions.

By Lemma 5.16 we know that $\tilde{T}^{Y(k,\epsilon)}$ is almost every differentiable in the relative interior of $\tilde{\Gamma}(Y(k, \epsilon))$. At point of differentiability of $\tilde{T}^{Y(k,\epsilon)}$, we know by Lemma 5.9 that all two-sided directional derivatives exists. Hence, we know that $\tilde{T}_{(ik)}(\gamma)$ exists for almost all $\gamma \in \tilde{\Gamma}(Y(k, \epsilon))$. But we already know from Lemma 5.24 that such existence can only depend on the values $\gamma(i)$ and $\gamma(k)$, so that existence is ensured on a full measure subset of the corresponding domain defined by:

$$\gamma(k) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right) \text{ and } \gamma(i) \in (0, 1 - \gamma(k)).$$

Now fix $x \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ and define

$$I(x) = \{y \in (0, 1-x) | \tilde{T}_{(ik)}(\gamma) \text{ exists when } \gamma(k) = x \text{ and } \gamma(i) = y\} \subset (0, 1-x). \tag{129}$$

Note that the union of these sets across $x \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ is precisely the set of $\gamma \in \tilde{\Gamma}(Y(k, \epsilon))$ on which $\tilde{T}_{(ik)}(\gamma)$ exists, which we know to have the same measure as the relative interior of $\tilde{\Gamma}(Y(k, \epsilon))$. This means that there exists $x(\epsilon) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ such that the measure of $I(x(\epsilon))$ is $1 - x(\epsilon)$. As $x(\epsilon) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$, take $I(\epsilon) = I(x(\epsilon)) \cap (1, 1-\epsilon)$. This completes the proof. ■

We now show how to define an appropriate fully additively separable function of the form we seek for any given $\epsilon \in (0, 0.5)$.

Lemma 5.27: Given $\epsilon \in (0, 0.5)$ and there exists a dense subset $I(\epsilon) \subset (0, 1-\epsilon)$ and a function $f^\epsilon : I(\epsilon) \to \mathbb{R}$ such that,

$$\tilde{T}_{(ji)}(\gamma) = f^\epsilon(\gamma(i)) - f^\epsilon(\gamma(j)), \tag{130}$$

for all $\gamma \in \tilde{\Gamma}$ such that $\gamma(i), \gamma(j) \in I(\epsilon)$ and $\gamma(i) + \gamma(j) < 1 - \epsilon$.

Proof. Given $\epsilon \in (0, 0.5)$, define $x(\epsilon) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ and the dense subset $I(\epsilon)$ of $(0, 1-\epsilon)$ so that the conditions of the last Lemma are satisfied. Now consider $\tilde{\Gamma}(\epsilon)$, the set of posteriors for which

Lemma 5.26 tells us that both $\tilde{T}_{(ik)}(\gamma)$ and $\tilde{T}_{(jk)}(\gamma)$ are well-defined,

$$\tilde{\Gamma}(\epsilon) = \left\{\gamma \in \tilde{\Gamma} | \gamma(i), \gamma(j) \in I(\epsilon), \gamma(k) = x(\epsilon)\right\},$$

Since both $\tilde{T}_{(ik)}(\gamma)$ and $\tilde{T}_{(jk)}(\gamma)$ exist on $\tilde{\Gamma}(\epsilon)$, by Lemma 5.25,

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), x(k)) - \bar{\boldsymbol{B}}(\gamma(j), x(k)). \tag{131}$$

for $\gamma \in \tilde{\Gamma}(\epsilon)$. We define the candidate function,

$$f^{\epsilon}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), x(\epsilon)).$$

The Lemma requires one more step, which is to remove the condition that $\gamma(k) = x(\epsilon)$, which is absent in the conditions of the Lemma. The key observation here is that Hence given γ' such that $\gamma'(i) = \gamma(i)$ and $\gamma'(j) = \gamma(i)$,

$$\tilde{T}_{(ji)}(\gamma') = \tilde{T}_{(ji)}(\gamma) = f^{\varepsilon}(\gamma(i)) - f^{\varepsilon}(\gamma(j)).$$

Hence the characterization applies to all $\gamma \in \tilde{\Gamma}$ such that $\gamma(i), \gamma(j) \in I(\epsilon)$ and $\gamma(i) + \gamma(j) < 1 - \epsilon$ as required. ■

Lemma 5.28: There exists $f : \bar{I} \to \mathbb{R}$ with $\bar{I} \subset (0, 1)$ of full measure such that for all $\gamma \in \tilde{\Gamma}$ with $\gamma(i), \gamma(j) \in \bar{I}$, $\tilde{T}_{(ji)}(\gamma)$ exists and,

$$\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)) \tag{132}$$

Proof. We construct a diminishing sequence $\{\epsilon(n)\}_{n=1}^{\infty} > 0$ with $\epsilon(n+1) < \epsilon(n)$ by setting $\epsilon(1) \in (0.0.5)$ and thereupon successively halving,

$$\epsilon(n+1) = \frac{\epsilon(n)}{2},$$

on $n > 1$. For each n, Lemma 5.26 states that there exists $x(n) \in \left(\frac{\epsilon(n)}{8}, \frac{\epsilon(n)}{4}\right)$ and a set $I(n) \subset (0, 1-\epsilon(n))$ which is dense in $(0, 1-\epsilon(n))$ such that $\tilde{T}_{(ji)}$ exists whenever $\gamma(j) = x(n)$ and $\gamma(i) \in I(n)$.

We now show that $I(n) \subset I(n+1)$. Since $I(n)$ is dense in $(0, 1-\epsilon(n))$ and $I(n+1)$ is dense in $(0, 1-\epsilon(n+1))$ and $(0, 1-\epsilon(n)) \subset (0, 1-\epsilon(n+1))$, $I(n) \cap I(n+1)$ is not empty. Consider $y \in I(n) \cap I(n+1)$ and choose η such that $\gamma(i) = y$, $\gamma(j) = x(n)$. That this is possible follows from the fact that

$$y + x(n) + x(n+1) < y + \frac{\epsilon(n)}{4} + \frac{\epsilon(n+1)}{4} \leq y + \frac{\epsilon(n)}{4} + \frac{\epsilon(n)}{8} < 1$$

Since $\gamma(i) \in I(n)$, $\tilde{T}_{(ji)}$ exists, and since $\gamma(i) \in I(n+1)$, $\tilde{T}_{(ki)}$ exists. It follows from Lemma 5.25, that $\tilde{T}_{(kj)}$ exists. Now consider any $y \in I(n)$, and consider η such that $\gamma(i) = y$, $\gamma(j) = x(n)$, and $\gamma(k) = x(n+1)$. Since $\gamma(i) \in I(n)$, $\tilde{T}_{(ji)}$ exists, and since $\tilde{T}_{(kj)}$ exists, it follows from Lemma 5.25, that $\tilde{T}_{(ki)}$ exists. Hence $y \in I(n+1)$.

By Lemma 5.27,

$$\tilde{T}_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n))$$

for all γ such that $\gamma(i), \gamma(j) \in I(n)$. Fix $z \in I(1)$, since $I(n) \subset I(n+1)$, $z \in I(n)$. For $\gamma(i) \in I(n)$, define

$$\boldsymbol{G}^n(\gamma(i)) = \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(z, x(n))$$

It follows that

$$\begin{aligned}
\tilde{T}_{(ji)}(\gamma) &= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n)) \\
&= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(z, x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n)) + \bar{\boldsymbol{B}}(z, x(n)) \\
&= \boldsymbol{G}^n(\gamma(i)) - \boldsymbol{G}^n(\gamma(j))
\end{aligned}$$

We now compare $\boldsymbol{G}^n(\gamma(i))$ to $\boldsymbol{G}^{n+1}(\gamma(i))$. Consider $\gamma(i) \in I(n) \subset I(n+1)$,

$$\begin{aligned}
\boldsymbol{G}^{n+1}(\gamma(i)) &= \bar{\boldsymbol{B}}(\gamma(i), x(n+1)) - \bar{\boldsymbol{B}}(z, x(n+1)) \\
&= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(x(n+1), x(n)) - \bar{\boldsymbol{B}}(z, x(n)) + \bar{\boldsymbol{B}}(x(n+1), x(n)) \\
&= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(z, x(n)) \\
&= \boldsymbol{G}^n(\gamma(i))
\end{aligned}$$

where the second equality follows from Lemma 5.25 applied to $\tilde{\gamma}$ with $\tilde{\gamma}(i) = x(n+1)$, $\tilde{\gamma}(j) = x(n)$ and $\tilde{\gamma}(k) = \gamma(i)$:

$$\tilde{T}_{(ji)}(\tilde{\gamma}) = \bar{\boldsymbol{B}}(x(n+1), x(n)) = \bar{\boldsymbol{B}}(x(n+1), \gamma(i)) - \bar{\boldsymbol{B}}(x(n), \gamma(i)) = -\bar{\boldsymbol{B}}(\gamma(i), x(n+1)) + \bar{\boldsymbol{B}}(\gamma(i), x(n))$$

and similarly for z in place of $\gamma(i)$.

Hence we can define a limit function $f : \cup_{n=1}^{\infty} I(n) \to \mathbb{R}$ unambiguously by taking any $x \in \cup_{n=1}^{\infty} I(n)$, selecting a particular $\bar{n}$ such that $x \in I(\bar{n})$, and defining,

$$f(x) = \boldsymbol{G}^{\bar{n}}(x).$$

By Lemma 5.27, we know that with $\gamma(i), \gamma(j) \in I(n)$, a dense subset of $(0, 1 - \epsilon(n))$, (132) holds,

$$\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j). \tag{133}$$

Since $\lim_{n\to\infty} \epsilon(n) = \lim_{n\to\infty} \epsilon(n) = 0$, note that

$$\bar{I} \equiv \cup_{n=1}^{\infty} I(n)$$

is a dense subset of $(0, 1)$, establishing the Lemma. ■

Lemma 5.29: The function $f : \bar{I} \longrightarrow \mathbb{R}$ defined in Lemma 5.28 for which (132) holds is non-decreasing, and can be extended to a function $f : (0, 1) \longrightarrow \mathbb{R}$ that is non-decreasing.

Proof. We pick arbitrary $x, x + \epsilon \in \cup_{n=1}^{\infty} I(n) = \bar{I}$ with $\epsilon > 0$ and show that $f(x + \epsilon) \geq f(x)$. Consider $\gamma \in \tilde{\Gamma}$ with $\gamma(i) = x$ and $\gamma(j) = x + \epsilon$, by Lemma 5.28

$$\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)) = f(x) - f(x + \epsilon).$$

If we now define $\gamma' \in \tilde{\Gamma}$ as,

$$\gamma' = \gamma + \epsilon(e_i - e_j),$$

Lemma 5.28 implies that,

$$\tilde{T}_{(ji)}(\gamma') = f(\gamma'(i)) - f(\gamma'(j)) = f(x+\epsilon) - f(x).$$

By the monotonicity lemma 5.14, $\gamma' = \gamma + \epsilon(e_i - e_j)$ for $\epsilon > 0$ implies $\tilde{T}_{(ji)}(\gamma') \geq \tilde{T}_{(ji)}(\gamma)$, which translates to,

$$f(x+\epsilon) - f(x) \geq f(x) - f(x+\epsilon),$$

which directly implies $f(x+\epsilon) \geq f(x)$, completing the proof that f is non-decreasing on $\bar{I}$.

To complete the proof, pick $x \in \bar{I}\backslash(0,1)$. Since $\bar{I} \subset (0,1)$ is of full measure in $(0,1)$, we can find sequence $\{x(n)\}_{n=1}^{\infty} > x$ with $x(n+1) < x(n)$ and $\lim_{n\to\infty} x(n) = x$ such that $x(n) \in \bar{I}$. We define $f(x)$ as the corresponding limit,

$$f(x) = \lim_{n\to\infty} f(x(n)$$

Since we have just shown f to be non-decreasing, the limit is well-defined, and also non-decreasing. ■

We now show a connection between continuity properties of f and existence of two-sided directional derivatives.

Lemma 5.30: Given $\eta \in \tilde{\Gamma}$, $\tilde{T}_{(ji)}(\eta)$ exists if and only if $f(\gamma(i)) - f(\gamma(j))$ is continuous at η.

Proof. Consider $\eta \in \tilde{\Gamma}$. Pick distinct states $1 \leq i, j \leq J$ and define

$$Y(\eta, i, j) = \{\gamma \in \mathbb{R}^J | \gamma(k) = \eta(k), k \neq i, j\}.$$

Note $Y(\eta, i, j)$ is convex set. Per the general prescription, define the restricted function $\tilde{T}^{\ Y(\eta,i,j)}$. By Lemma 5.16 this function is differentiable almost everywhere in the relative interior of the restricted domain $\tilde{\Gamma}(Y(\eta, i, j))$. Hence the directional derivative in the only relevant direction,

$$\tilde{T}^{Y(k,\epsilon)}_{(ji)}(\gamma) = \tilde{T}_{(ji)}(\gamma),$$

exists almost everywhere. Hence we can find sequences approaching from η both corresponding directions. We now select $\{\epsilon(n)\}_{n=1}^{\infty} > 0$ with $\lim_{n\to\infty} \epsilon(n) = 0$ such that given $\gamma_n = \eta + \epsilon(n)(e_i - e_j)$, $\tilde{T}_{(ji)}(\gamma_n)$ exists. We select also $\{\epsilon'(n)\}_{n=1}^{\infty} < 0$ with $\lim_{n\to\infty} \epsilon'(n) = 0$ such that, defining $\gamma'_n = \eta + \epsilon'(n)(e_i - e_j)$, $\tilde{T}_{(ji)}(\gamma'_n)$ exists.

Since $\tilde{T}$ is convex we know that the one-sided directional derivatives are monotonically increasing,

$$\lim_{n\to\infty} \tilde{T}_{(ji)}(\gamma'_n) \leq -\tilde{T}_{\overrightarrow{ji}}(\eta) \leq \tilde{T}_{\overrightarrow{ji}}(\eta) \leq \lim_{n\to\infty} \tilde{T}_{(ji)}(\gamma_n). \tag{134}$$

We now use Lemma 5.28 to substitute in (134) at all γ_n and γ'_n since $\tilde{T}_{(ji)}(\circ)$ is well-defined at these points, to arrive at,

$$\begin{aligned}\tilde{T}_{(ji)}(\gamma_n) &= f(\eta(i)+\epsilon_n) - f(\eta(j)-\epsilon_n);\\ \tilde{T}_{(ji)}(\gamma'_n) &= f(\eta(i)+\epsilon'_n) - f(\eta(j)-\epsilon'_n);\end{aligned}$$

Now suppose that $f(\gamma(i)) - f(\gamma(j))$ is continuous at η. In this case,

$$\lim_{n\to\infty}\left[f(\eta(i)+\epsilon'_n) - f(\eta(j)-\epsilon'_n)\right] = \lim_{n\to\infty}\left[f(\eta(i)+\epsilon_n) - f(\eta(j)-\epsilon_n)\right],$$

so that correspondingly,

$$\lim_{n\to\infty}\tilde{T}_{(ji)}(\gamma'_n) = \lim_{n\to\infty}\tilde{T}_{(ji)}(\gamma_n),$$

hence by (134),

$$-\tilde{T}_{\overrightarrow{ji}}(\eta) = \tilde{T}_{\overrightarrow{ji}}(\eta),$$

establishing through Lemma 5.9 that $\tilde{T}_{(ji)}(\eta)$ exists.

Suppose conversely that $\tilde{T}_{(ji)}(\eta)$ does not exist. In this case we know by Lemma 5.9 that $\tilde{T}_{\overrightarrow{ji}}(\eta) < \tilde{T}_{\overrightarrow{ji}}(\eta)$, so that by (134),

$$\lim_{n\to\infty}\tilde{T}_{(ji)}(\gamma_n) = \lim_{n\to\infty} f(\eta(i)+\epsilon'_n) - f(\eta(j)-\epsilon'_n) < \lim_{n\to\infty}\tilde{T}_{(ji)}(\gamma_n) = \lim_{n\to\infty} f(\eta(i)+\epsilon_n) - f(\eta(j)-\epsilon_n),$$

establishing that $f(\gamma(i)) - f(\gamma(j))$ is discontinuous at η, and completing the proof. ■

A5.2.13: Existence of Directional Derivatives

Lemma 5.31: Given η and given $\alpha = \frac{\eta(k)}{\eta(l)}$, if $\tilde{T}_{(kl)}(\eta)$ exists then $\tilde{T}$ is differentiable for almost all $\gamma \in \Gamma_{kl}(\alpha)$.

Proof. Consider $\eta \in \tilde{\Gamma}$ such that $\tilde{T}_{(kl)}(\eta)$ exists and set $\alpha = \frac{\eta(k)}{\eta(l)}$. Since $\Gamma_{kl}(\alpha)$ is convex, $\tilde{T}^{\Gamma_{kl}(\alpha)}$ almost everywhere differentiable on the relative interior of $\Gamma_{kl}(\alpha)$ by Lemma 5.16. At points of differentiability, we know from Lemma 5.9 that $\tilde{T}_{(ji)}(\gamma) = \tilde{T}^{\Gamma_{kl}(\alpha)}_{(ji)}(\gamma)$ exists provided $\Gamma_{kl}(\alpha)$ contains a line segment through γ in direction $(e_i - e_j)$, By definition of $\Gamma_{kl}(\alpha)$, this holds for all directions except that defined by the pair of states (lk) whose posterior belief ratio is held fixed through the set.

Consider $\gamma \in \Gamma_{kl}(\alpha)$ at which $\tilde{T}^{\Gamma_{kl}(\alpha)}$ is differentiable. As $\tilde{T}_{(kl)}(\eta)$ exists, Lemma 5.18 implies that $\tilde{T}_{(kl)}(\gamma)$ also exists. Hence at all such γ, we know that all 2-sided directional derivatives exist. Following precisely the steps in Lemma 5.25, we can remove an arbitrary state $k \neq i, j$ from the domain and construct set X as in (122), then define $\eta(x) \in \tilde{\Gamma}$ on $x \in X$ as in (123) and function $H(x)$ on X by (124), whose partial derivatives are precisely the directional derivatives $\tilde{T}_{(km)}(\gamma)$,

$$\tilde{T}_{(km)}(\gamma) = H_1(\gamma(m), \gamma(i)),$$

all $m \neq k$.

Since all partials of this function therefore exist, we note from Rockafellar theorem 25.2 that $H(\gamma)$ is differentiable at γ and that the directional derivative function $H'(\gamma|y)$ is linear in direction $y \in \mathbb{R}^2$. Re-application of Rockafellar theorem 25.2 implies that $\tilde{T}$ is differentiable at γ, completing the proof. ■

Lemma 5.32: $\tilde{T}_{(ji)}(\gamma)$ exists for all i, j and $\gamma \in \tilde{\Gamma}$.

Proof. The proof is by contradiction. Consider a posterior η at which $\tilde{T}_{(ji)}(\eta)$ does not exist. It follows from Lemma 5.30 that $f(\eta(i)) - f(\eta(j))$ is discontinuous at this point:

$$\lim_{\epsilon \uparrow 0} f(\eta(i) + \epsilon) - f(\eta(j) + \epsilon) \leq \lim_{\epsilon \downarrow 0} f(\eta(i) + \epsilon) - f(\eta(j) + \epsilon)$$

Without loss of generality, suppose that it is $f(\eta(i))$ that is discontinuous.

Since f is monotonic, $f(\gamma(j))$ is continuous for almost all $\gamma(j) \in (0, 1 - \eta(i))$ (Rudin (1976) Theorem 4.30). The discontinuity of f at $\eta(i)$ and the continuity of f almost everywhere else implies that $f(\eta(i)) - f(\gamma(j))$ is discontinuous in the direction (ji) for almost all $\gamma(j) \in (0, 1 - \eta(i))$. Hence by Lemma 5.30, $\tilde{T}_{(ji)}(\gamma)$ does not exist for almost all γ such that $\gamma(i) = \eta(i)$ and $\gamma(j) \in (0, 1 - \eta(i))$. It follows that for almost all $\alpha \in \left(\frac{\eta(i)}{1-\eta(i)}, \infty\right)$, there exists $\gamma \in \Gamma^{\alpha}_{ji}$ such that $\tilde{T}_{(ji)}(\gamma)$ does not exist. But by Lemma 5.18, if $\tilde{T}_{(ji)}(\gamma)$ exists for any $\eta \in \Gamma^{\alpha}_{ji}$, then $\tilde{T}_{(ji)}(\gamma)$ exists for all $\gamma \in \Gamma^{\alpha}_{ji}$. Hence for almost all $\alpha \in \left(\frac{\eta(i)}{1-\eta(i)}, \infty\right)$, $\tilde{T}_{(ji)}(\gamma)$ does not exist for any $\gamma \in \Gamma^{\alpha}_{ji}$. But $\left\{\gamma | \gamma \in \Gamma_{ji}(\alpha), \alpha \in \left(\frac{\eta(i)}{1-\eta(i)}, \infty\right)\right\}$ is a set of positive measure and $\tilde{T}$ is differentiable almost everywhere. This contradiction establishes the result. ■

Lemma 5.33: $\tilde{T}$ is continuously differentiable on $\gamma \in \tilde{\Gamma}$ and $f(\gamma(j))$ is continuous on $\tilde{\Gamma}$

Proof. Lemma 5.32 establishes that the directional derivatives $\tilde{T}_{(ji)}(\gamma)$ exist for all (ji) and all $\gamma \in \tilde{\Gamma}$. It follows from Rockafellar (1970) Theorem 25.2 that $\tilde{T}$ is differentiable and from Rockafellar (1970) Corollary 25.5.1 that $\tilde{T}$ is continuously differentiable on $\gamma \in \tilde{\Gamma}$. Since $\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j))$ continuity of f follows. ■

A5.2.14: Existence of Cross-Directional Derivatives

The twice differentiability of a convex function such as $\tilde{T}$ normally a subtle object as convexity alone is not sufficient to establish differentiability on any open set. In our case, however, we know from Lemma 5.33 that $\tilde{T}$ is continuously differentiable on $\gamma \in \tilde{\Gamma}$ and hence standard notions of twice differentiability apply. We now introduce the cross derivatives of $\tilde{T}$, which are of the essence in the proof of theorem 1.

Definition 14 *Given $\gamma \in \tilde{\Gamma}$ and any two pair of states $1 \leq i \neq j \leq J$ and $1 \leq k \neq l \leq J$ we define the corresponding **cross derivative in direction** lk of $\tilde{T}_{(ji)}$, as the corresponding directional derivative of $\tilde{T}_{(ji)}$, should it exist:*

$$\tilde{T}_{(ji)(lk)}(\gamma) = \lim_{\epsilon \to 0} \frac{\tilde{T}_{(ji)}(\eta + \epsilon(e_k - e_l)) - \tilde{T}_{(ji)}(\eta)}{\epsilon}$$

The next results show that the cross-directional derivatives exist almost everywhere.

Lemma 5.34: $\tilde{T}_{(ji)(il)}$ exists almost everywhere in $\tilde{\Gamma}$

Proof. Rockafellar (1999) Theorem 2.3 states that for almost all $\gamma \in \tilde{\Gamma}$ the gradient of $\tilde{T}(\gamma)$, $\nabla\tilde{T}(\gamma)$, exists and

$$\nabla\tilde{T}(\gamma) = \nabla\tilde{T}(\gamma + \omega) + A\omega + o(|\omega|)$$

holds with respect to $\{\omega | \gamma + \omega \in \text{dom}\nabla\tilde{T}\}$ for some $J-1$ by $J-1$ matrix A. By Lemma 5.33, $\tilde{T}$ is differentiable everywhere on $\tilde{\Gamma}$, and so $\text{dom}\nabla\tilde{T} = \tilde{\Gamma}$. The Lemma follows from considering the direction in $\nabla\tilde{T}$ associated with (il) after the direction associated with (ji). ■

Lemma 5.35: Given $\eta \in \tilde{\Gamma}$ at which $\tilde{T}_{(ij)(lk)}(\eta)$ exists and given $\alpha = \frac{\eta(k)}{\eta(l)}$, $\tilde{T}_{(ij)(lk)}(\nu)$ exists for all $\nu \in \Gamma_{kl}(\alpha)$.

Proof. Choose η at which $\tilde{T}_{(ij)(lk)}$ exists and set $\alpha = \frac{\eta(k)}{\eta(l)}$. Consider $\nu \in \Gamma_{kl}(\alpha)$. Lemma 5.12 establishes the existence of a parameterized set of problems (μ_t, A) indexed by $t \in [0,1]$ where:

$$\mu_t(j) = \begin{cases} t\,[\mu(k) + \mu(l)] & \text{for } j = k; \\ (1-t)\,[\mu(k) + \mu(l)] & \text{for } j = l; \\ \mu(j) & \text{otherwise;} \end{cases}$$

and η_t is the revealed posterior for action a and ν_t is the revealed posterior for action b where

$$\eta_t(j) = \frac{\eta(j)}{\mu(j)}\mu_t(j) \qquad \text{and} \qquad \nu_t(j) = \frac{\nu(j)}{\mu(j)}\mu_t(j).$$

By Lemma 5.31, $\tilde{T}$ is differentiable everywhere in $\tilde{\Gamma}$. It follows from Lemma 5.9 that $\tilde{T}_{(ij)}(\eta_t)$ and $\tilde{T}_{(ij)}(\nu_t)$ exist for all t. Lemma 5.11 implies therefore that,

$$\tilde{N}^a_{(ij)}(\eta_t) = \tilde{N}^b_{(ij)}(\nu_t)$$

for all t. Substituting the definition of net utility yields

$$\tilde{T}_{(ij)}(\nu_t) = \tilde{T}_{(ij)}(\eta_t) - u(a,k) + u(a,l) + u(b,k) - u(b,l) \tag{135}$$

Differencing (135) evaluated at t and $\bar{t}$ and taking $\mu_t = \mu + \epsilon_t(e_k - e_l)$ implies $\eta_t = \eta + \epsilon_t\frac{\eta(k)}{\mu(k)}(e_k - e_l)$ and $\nu_t = \nu + \epsilon_t\frac{\nu(k)}{\mu(k)}(e_k - e_l)$ yields

$$\lim_{\epsilon_t \to 0} \frac{\tilde{T}_{(ij)}\left(\nu + \epsilon_t\frac{\nu(k)}{\mu(k)}(e_k - e_l)\right) - \tilde{T}_{(lk)}(\eta)}{\epsilon_t\frac{\nu(k)}{\mu(k)}} = \lim_{\epsilon_t \to 0} \frac{\tilde{T}_{(ij)}\left(\eta + \epsilon_t\frac{\eta(k)}{\mu(k)}(e_k - e_l)\right) - \tilde{T}_{(lk)}(\eta)}{\epsilon_t\frac{\eta(k)}{\mu(k)}} \tag{136}$$

Since $\tilde{T}_{(ij)(lk)}(\eta)$ exists, the right-hand side limit exists, and so the left-hand side limit exists, establishing the result. ■

Lemma 5.36: Given $\gamma \in \tilde{\Gamma}$, $1 \leq i \neq j \leq J$ and $1 \leq k \neq l \leq J$, all cross-derivatives $\tilde{T}_{(ji)(lk)}(\gamma)$ exist.

Proof. Note first that for all $\gamma \in \tilde{\Gamma}$ for any distinct states i, j, k, and l, $\tilde{T}_{(ji)(lk)}(\gamma) = 0$, since Lemma 5.29 and 4.33 imply that $\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j))$ which is independent of $\gamma(k)$ and $\gamma(l)$.

Now consider cases with overlap. Consider first the case in which $k = i$ and $l \neq j$. The proof is by contradiction. Consider a posterior $\eta \in \tilde{\Gamma}$. By Lemma 5.32, $\tilde{T}$ is differentiable at η, and hence by Lemma 5.29, $\tilde{T}_{(ji)}(\eta) = f(\eta(i)) - f(\eta(j))$. It follows that

$$\frac{\tilde{T}_{(ji)}(\eta + \epsilon(e_i - e_l)) - \tilde{T}_{(ji)}(\eta)}{\epsilon} = \frac{f(\eta(i) + \epsilon) - f(\eta(i))}{\epsilon}$$

so that $\tilde{T}_{(ji)}$ is differentiable in the direction (li) if and only if f is differentiable at $\eta(i)$. Suppose now that f is not differentiable at $\eta(i)$. Consider the set of posteriors ν such that $\nu(i) = \eta(i)$ and $v(j) \in (1, 1 - \eta(i))$. Since

$$\frac{\tilde{T}_{(ji)}(\nu + \epsilon(e_i - e_l)) - \tilde{T}_{(ji)}(\nu)}{\epsilon} = \frac{f(\eta(i) + \epsilon) - f(\eta(i))}{\epsilon}$$

$\tilde{T}_{(ji)(il)}(\gamma)$ does not exist for all such ν. It follows that for each $\alpha \in (\frac{\eta(i)}{1-\eta(i)}, \infty)$, there exists $\gamma \in \Gamma^{\alpha}_{ji}$ such that $\tilde{T}_{(ji)(il)}(\gamma)$ does not exist, namely any γ such that $\gamma(i) = \eta(i)$ and $\gamma(j) = \eta(j)/\alpha$. But by Lemma 5.35, if $\tilde{T}_{(ji)(il)}(\gamma)$ exists at any $\gamma \in \Gamma_{ji}(\alpha)$ then $\tilde{T}_{(ji)(il)}$ exists for all $\gamma \in \Gamma_{ji}(\alpha)$. Hence $\tilde{T}_{(ji)(il)}(\gamma)$ does not exist for any $\gamma \in \Gamma_{ji}(\alpha)$ such that $\alpha \in \left(\frac{\eta(i)}{1-\eta(i)}, \infty\right)$. But this is a set of positive measure in $\tilde{\Gamma}$ which contradicts the result of Lemma 5.34. This contradiction establishes that $\tilde{T}_{(ji)(il)}(\gamma)$ exists for $k = i$ and $l \neq j$.

Finally, note that the above proves the differentiability of f which establishes that $\tilde{T}_{(ji)(il)}(\gamma)$ exists in the case that $k = i$ and $l = j$. ■

A5.3: Theorem 1 (Sufficiency)

Theorem 1: *If data set $C \in \mathcal{C}$ with a UPS representation satisfies IUC, it has a Shannon representation.*

Proof. We are looking to show that if $C \in \mathcal{C}$ *has a UPS representation* $K \in \mathcal{K}^{UPS}$ *and satisfies IUC*, then there exists $\kappa > 0$ such that, given $(\mu, Q) \in \mathcal{F}$ such that $Q \in \hat{\mathcal{Q}}(\mu|K)$,

$$K(\mu, Q) = \Sigma_{\gamma \in \Gamma(Q)} Q(\gamma) T(\gamma),$$

where,

$$T(\gamma) = \kappa \sum_{\omega \in \Gamma(\gamma)} \gamma(\omega) \ln(\gamma(\omega)). \tag{137}$$

The proof has three parts. The first establishes that, given any fixed set of states $\bar{\Omega}$ of cardinality $J \geq 4$, there exists $\kappa^J > 0$ such that, for all corresponding interior posteriors $\gamma \in \tilde{\Gamma}$ with $\bar{\Omega}$, the corresponding strictly convex function $\tilde{T} : \tilde{\Gamma} \to \mathbb{R}$ in the UPS representation has the form,

$$\tilde{T}(\gamma) = \kappa^J \sum_{j=1}^{J} \gamma(j) \ln \gamma(j). \tag{138}$$

The second part of the proof shows all optimal strategies are precisely as if κ^J applied to all posteriors $\gamma \in \Gamma(\mu)$ with $|\Omega(\gamma)| = L \leq J$. This implies that one can identify all optimal strategies, and hence the full data in this CIR, using the necessary and sufficient conditions for optimality when this function to all feasible posteriors. The final step is to apply IUC (A1) not only to show that $\kappa^{J+1} = \kappa^J$ for $J \geq 4$, but also that the same Shannon functional form and multiplier applies for all $J \geq 2$, which completes the proof.

To prove the first part, we pick any fixed set of states $\bar{\Omega}$ of cardinality $J \geq 4$, and define the corresponding interior posteriors $\tilde{\Gamma}$. We choose $\eta \in \tilde{\Gamma}$, set $\alpha = \frac{\eta(k)}{\eta(l)}$, and consider $\nu \in \Gamma_{kl}(\alpha)$, so that,

$$\frac{\eta(k)}{\eta(l)} = \frac{\nu(k)}{\nu(l)} > 0.$$

By Lemma 5.7 (Symmetric Costs), we can order the arguments such that $l = J$. Define the mean belief $\bar{\mu} = \frac{\eta+\nu}{2}$, and μ_t for $t \in [0,1]$ by:

$$\mu_t(j) = \begin{cases} t[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = k; \\ (1-t)[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = l; \\ \bar{\mu}(j) & \text{otherwise.} \end{cases}$$

By Lemma 5.12 we know that there exists $a, b \in \mathcal{A}$ with $u(a,k) = u(a,l)$ and $u(b,k) = u(b,l)$ such that $C(\mu_t, \{a,b\}) = \{P_t\}$ and for $t \in (0,1)$, η_t is the revealed posterior for action a and ν_t is the revealed posterior for action b where,

$$\eta_t(j) = \left[\frac{\eta(j)}{\bar{\mu}(j)}\right] \mu_t(j) \qquad \text{and} \qquad \nu_t(j) = \left[\frac{\nu(j)}{\bar{\mu}(j)}\right] \mu_t(j)$$

for $1 \leq j \leq J$. Since $\mu_t(j) = \bar{\mu}(j)$ for $j \neq k, l$ and $\mu_t(k) = t[\bar{\mu}(k) + \bar{\mu}(l)]$, note that $\mu_{\bar{t}} = \bar{\mu}$ if and only it,

$$\bar{t} = \frac{\bar{\mu}(k)}{\bar{\mu}(k) + \bar{\mu}(l)},$$

in which case,

$$\eta_{\bar{t}}(j) = \eta(j) \text{ and } \nu_{\bar{t}}(j) = \nu(j).$$

By Lemma 5.33, $\tilde{T}$ is differentiable for all $\gamma \in \tilde{\Gamma}$. By Lemma 5.11, we know that (85) holds for all $t \in (0,1)$,

$$u(a,i) - u(a,j) - \tilde{T}_{(ji)}(\eta_t) \equiv \tilde{N}^a_{(ji)}(\eta_t) = \tilde{N}^b_{(ji)}(\nu_t) \equiv u(b,i) - u(b,j) - \tilde{T}_{(ji)}(\nu_t).$$

By Lemma 5.36, $\tilde{T}_{(ji)}$ is differentiable in the direction (lk). Since this equation holds for all $t \in (0,1)$, we can differentiate this identity with respect to t at $\bar{t}$,

$$\frac{d\tilde{T}_{(ji)}(\eta_{\bar{t}})}{dt} = \frac{d\tilde{T}_{(ji)}(\nu_{\bar{t}})}{dt}. \tag{139}$$

A change in t at $\bar{t}$ raises $\eta_t(k)$ and lowers $\eta_t(l)$ each by $\eta(k)+\eta(l)$ and leaves $\eta_t(j)$ unchanged for all $j \neq k, l$. The chain rule then implies,

$$\frac{dT_{(ji)}(\eta_{\bar{t}})}{dt} = \tilde{T}_{(ji)(lk)}(\eta)\left[\eta(k)+\eta(l)\right] \tag{140}$$

Combining (139) and (140) and setting $t = \bar{t}$ yields,

$$\tilde{T}_{(ji)(lk)}(\eta)\left[\eta(k)+\eta(l)\right] = \tilde{T}_{(ji)(lk)}(\nu)\left[\nu(k)+\nu(l)\right].$$

Since,

$$\frac{\nu(k)+\nu(l)}{\eta(k)+\eta(l)} = \frac{\bar{t}\left[\nu(k)+\nu(l)\right]}{\bar{t}\left[\eta(k)+\eta(l)\right]} = \frac{\nu(k)}{\eta(k)}$$

we have

$$\eta(k)\tilde{T}_{(ji)(lk)}(\eta) = \nu(k)\tilde{T}_{(ji)(lk)}(\nu). \tag{141}$$

Equation (141) must hold for all η and ν in $\Gamma_{kl}(\alpha)$, therefore $\gamma(k)\tilde{T}_{(ji)(lk)}(\gamma)$ is constant across $\gamma \in \tilde{\Gamma}$. Note that by Lemma 5.7, this constant is independent of the states i, j, k, and l, although at this point it may depend on the dimension J. Since the additive separability of $\tilde{T}$ implies $\tilde{T}_{(ji)(lk)} = 0$ for distinct states i, j, k, and l. The interesting cases involve overlap. Taking $i = k$.

$$\gamma(i)\tilde{T}_{(ji)(li)}(\gamma) = \kappa^J. \tag{142}$$

for some constant κ^J.

We look for the general form of $\tilde{T}_{(ji)}$ that solves (142). We know from Lemma 5.28 that,

$$\tilde{T}_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)).$$

Hence

$$\gamma(i)f'(\gamma(i)) = \kappa^J$$

The solution to this differential equation is

$$f(\gamma(i)) = \kappa^J \ln\gamma(i) + \varsigma$$

for some constant of integration ς. It follows that

$$\tilde{T}_{(ji)}(\gamma) = \kappa^J \ln\gamma(i) - \kappa^J \ln\gamma(j). \tag{143}$$

Note that we can rule out the dependence of ς on $\gamma(m)$ for $m \neq i$ as $f(\gamma(i))$ depends only on $\gamma(i)$.

The general solution to (143) is

$$\tilde{T}(\gamma) = \kappa^J\gamma(i)[\ln(\gamma(i)) - 1] + \kappa^J\gamma(j)[\ln(\gamma(j)) - 1] + G(\gamma(i)+\gamma(j), \{\gamma(k)\}_{k\neq i,j}).$$

Here G combines constants of integration with the potential for a shift between $\gamma(i)$ and $\gamma(j)$ to offset each other. Lemma 5.7 states that $\tilde{T}$ is symmetric. Hence,

$$\tilde{T}(\gamma) = \sum_j \kappa^J \gamma(j)[\ln(\gamma(j)) - 1] + G\left(\sum_j \gamma(j)\right).$$

As $\sum_j \gamma(j) = 1$ and $\tilde{T}$ is defined only to an affine transformation,

$$\tilde{T}(\gamma) = \kappa^J \sum_j \gamma(j)\ln(\gamma(j)). \tag{144}$$

This completes the proof of (138).

For the second part of the proof, the first key observation is that, given $J \geq 4$, all optimal strategies are precisely as if κ^J applied to all posteriors $\gamma \in \Gamma$ with $|\Omega(\gamma)| = L \leq J$. Note that we replace $\tilde{T}$ with T from this point forward in the proof since this function is designed to apply to boundary as well as to interior posteriors. What we require then is that cost cannot be strictly lower than this formula implies,

$$T(\gamma) \geq \kappa^J \sum_{l=1}^{L} \gamma(l)\ln\gamma(l). \tag{145}$$

To see that this is sufficient, suppose that this has been established and replace the costs of all such posteriors with precisely the lower bound. Note that in this case one can apply the standard necessary and sufficient conditions for optimal choices to conclude that, even with this lower bound imposed it is never optimal to select any such posteriors. The equal likelihood ratio necessary conditions for $\lambda \in \Lambda(\mu, A)$ to be optimal in the Shannon model with general cost parameter $\kappa^J > 0$ in Caplin *et al.* [2016]) which asserts that, for $\lambda \in \Lambda(\mu, A)$ to be an optimal strategy requires, that for all chosen actions $a, b \in \mathcal{A}(\lambda)$,

$$\frac{\gamma_\lambda^a(j)}{\exp(u(a,j)/\kappa^{J+1})} = \frac{\gamma_\lambda^b(j)}{\exp(u(b,j)/\kappa^J + 1)} \text{ all } j. \tag{146}$$

Note that this is inconsistent with there being any posterior with $\gamma_\lambda^a(j) = 0$, since we know that $\mu(j) > 0$, so that by Bayes' rule there must be some strictly positive values $\gamma_\lambda^b(j) > 0$ for some chosen action, hence all must be strictly positive for (146) to hold. Hence replacement of $\kappa^J \sum_{l=1}^{L} \gamma(l)\ln\gamma(l)$ when some of the ex ante possible states are ruled out cannot make strategies with such posteriors optimal: hence there is no loss from the view point of optimization and hence data observed in the CIR in applying this cost function.

We now demonstrate the validity of (145). Suppose to the contrary that there exists some $\gamma \in \Gamma$ with $|\Omega(\gamma)| = L < J$ such that the opposite holds,

$$T(\gamma) = \kappa^J \sum_{l=1}^{L} \gamma(l)\ln\gamma(l) - \delta, \tag{147}$$

for $\delta > 0$. Note that there is no loss of generality in supposing that $L = (J-1)$, since there must be some highest number of states $\bar{L} < J$ for which this is true, which implies that (145) held for all $\bar{L}+1$ yet not for $\bar{L}$, so that the comparison below works at least as well for $\bar{L}+1$ relative to $\bar{L}$, as it does for J relative to $\bar{L}$. It then simplifies to set $\bar{L} = J - 1$. We show that this contradicts completeness, since it is cheaper and cannot lower expected utility to rule out an ex ante possible state state than to leave minimal ignorance.

Consider a particular posterior $\bar{\gamma}$ of this form such that (147) holds with difference $\bar{\bar{\delta}} > 0$. Without loss of generality, suppose that the prior possible state that is impossible state in this posterior is state 1, $\bar{\gamma}(1) = 0$. In this case we can express costs in simple form,

$$T(\bar{\gamma}) = \kappa^J \sum_{j=2}^{J} \bar{\gamma}(j) \ln \bar{\gamma}(j) - \bar{\bar{\delta}}. \tag{148}$$

We now define a second posterior $\bar{\eta}$ that permutes $\bar{\gamma}$ by reversing the posteriors associated with states 1 and 2,

$$\bar{\eta}(j) = \begin{cases} \bar{\gamma}(2) \text{ if } j = 1; \\ 0 \text{ if } j = 2; \\ \bar{\gamma}(j) \text{ if } j \geq 3. \end{cases}$$

By the Symmetry of Costs Lemma, we know that $T(\bar{\gamma}) = T(\bar{\eta})$.

We now define $\bar{\mu}$ to be the average posterior,

$$\bar{\mu} = \frac{\bar{\gamma} + \bar{\eta}}{2} = \begin{cases} \frac{\bar{\gamma}(2)}{2} \text{ if } j = 1, 2; \\ \bar{\gamma}(j) \text{ if } j \geq 3. \end{cases}.$$

and consider the parameterized families of posteriors $\bar{\gamma}^\alpha, \bar{\eta}^\alpha \in \Gamma(\bar{e})$ on $\alpha \in [0, 1]$ by:

$$\bar{\gamma}^\alpha(j) = \begin{cases} \alpha\bar{\gamma}(2) \text{ if } j = 1; \\ (1-\alpha)\bar{\gamma}(2) \text{ if } j = 2 \\ \bar{\gamma}(j) \text{ if } j \geq 3. \end{cases} \text{ and } \bar{\eta}^\alpha(j) = \begin{cases} (1-\alpha)\bar{\gamma}(2) \text{ if } j = 1; \\ \alpha\bar{\gamma}(2) \text{ if } j = 2 \\ \bar{\gamma}(j) \text{ if } j \geq 3. \end{cases} \tag{149}$$

By construction, for all $\alpha \in [0, 1]$ the simple average of the posteriors $\bar{\gamma}^\alpha, \bar{\eta}^\alpha$ is precisely $\bar{\mu}$,

$$\frac{1}{2}\left[\bar{\gamma}^\alpha(j) + \bar{\eta}^\alpha(j)\right] = \begin{cases} \frac{\bar{\gamma}(2)}{2} \text{ if } j = 1, 2 \\ \bar{\gamma}(j) \text{ if } j \geq 3. \end{cases}$$

Given that the data has a PS representation, theorem 3 implies that Axiom A4 (Completeness) holds. Hence we know that, for any $\alpha > 0$, there exists action set $\{a^\alpha, b^\alpha\}$ such that decision problem $(\bar{\mu}, \{a^\alpha, b^\alpha\}) \in \mathcal{D}$ has data $P \in C(\bar{\mu}, \{a^\alpha, b^\alpha\})$ that assigns equal likelihood $\frac{1}{2}$ to each of $\bar{\gamma}^\alpha, \bar{\eta}^\alpha$. Since this is a CIR, there exists an optimal strategy $\lambda^\alpha \in \hat{\Lambda}(\bar{\mu}, \{a^\alpha, b^\alpha\})$ with the corresponding property,

$$Q_{\lambda^\alpha}(\bar{\gamma}^\alpha) = Q_{\lambda^\alpha}(\bar{\eta}^\alpha) = \frac{1}{2}.$$

We show now that if $\kappa^J < \kappa^{J+1}$, this cannot hold for small enough $\alpha > 0$. The proof involves demonstrating that alternative strategy λ' that assigns equal likelihood to posteriors $\bar{\gamma}$ and $\bar{\eta}$

$$Q_{\lambda^\alpha}(\bar{\gamma}) = Q_{\lambda^\alpha}(\bar{\eta}) = \frac{1}{2},$$

strictly dominates in this limit.

That such a strategy produces no lower expected utility follows directly from the fact that the posterior distribution in λ^α is a garbling of λ', so that λ' is Blackwell more informative than λ^α. With regard to the costs, note that, for any $\alpha \in (0,1)$, the relevant parameter in the Shannon function is $\kappa^J > 0$, since both possible posteriors have all prior possible states still possible,

$$|\Omega(\bar{\gamma}^\alpha)| = |\Omega(\bar{\eta}^\alpha)| = J.$$

Given that the posteriors are permutations of one another, the corresponding Shannon costs are simple to compute as,

$$K(\bar{\mu}, Q_{\lambda^\alpha}) = \kappa^J \left[\sum_j \bar{\gamma}(j)\ln\bar{\gamma}(j) + \alpha\bar{\gamma}(2)\ln(\alpha\bar{\gamma}(2)) + (1-\alpha)\bar{\gamma}(2)\ln[(1-\alpha)\bar{\gamma}(2)]\right] \tag{150}$$

By construction, equation (148) and the symmetry of costs shows that the corresponding computation for λ' has cost strictly $\bar{\delta} > 0$ below what it would be according to the corresponding Shannon cost,

$$K(\bar{\mu}, Q_{\lambda'}) = \kappa^J \left[\sum_j \bar{\gamma}(j)\ln\bar{\gamma}(j) + \bar{\gamma}(2)\ln(\bar{\gamma}(2))\right] - \bar{\delta}. \tag{151}$$

Subtraction of (151) from (150) reveals that the costs of λ^α are strictly higher than those of λ' provided,

$$\bar{\delta} > \alpha\bar{\gamma}(2)\ln(\alpha\bar{\gamma}(2)) + (1-\alpha)\bar{\gamma}(2)\ln[(1-\alpha)\bar{\gamma}(2)] - \bar{\gamma}(2)\ln(\bar{\gamma}(2)). \tag{152}$$

Note that the LHS of (152) is a fixed strictly positive constant independent of α. With regard to the RHS, note that in the limit as $\alpha \downarrow 0$ it approaches zero, since,

$$\begin{aligned} \lim_{\alpha\downarrow 0}(1-\alpha)\bar{\gamma}(2)\ln[(1-\alpha)\bar{\gamma}(2)] &= \bar{\gamma}(2)(2)\ln[\bar{\gamma}(2)]; \\ \lim_{\alpha\downarrow 0}[\alpha\bar{\gamma}(2)\ln(\alpha\bar{\gamma}(2))] &= 0. \end{aligned}$$

For $\alpha > 0$ but sufficiently small we conclude that,

$$K(\bar{\mu}, Q_{\lambda'}) < K(\bar{\mu}, Q_{\lambda^\alpha}),$$

contradicting optimality of strategy $\lambda(\alpha)$ and thereby establishing (145).

We know now that the applicable cost function for working out all optimal strategies and hence all observed data for priors involving $J \geq 4$ possible states the Shannon cost function with parameter κ^J as defined for posteriors γ with $|\Omega(\gamma)| = J$ in equation (144). The final part of the proof uses IUC (A1) to iterate down in dimension. To be precise, define K^J to be the Shannon cost function with parameter κ^J for $J \geq 4$ as defined on all posteriors with that state space or below,

$$K^J(\gamma) \equiv \kappa^J \sum_{j\in\Omega(\gamma)} \gamma(j)\ln\gamma(j), \text{ all } \gamma \in \Gamma \text{ with } |\Omega(\gamma)| \leq J. \tag{153}$$

The precise result we establish is that, given any decision problem $(\mu, A) \in \mathcal{D}$ with a prior of cardinality one lower, $|\Omega(\mu)| = J - 1$,

$$P \in C(\mu, A) \text{ iff } \exists\lambda \in \hat{\Lambda}(\mu, A|K^J) \text{ such that } \mathbf{P}_\lambda = P. \tag{154}$$

Note that establishing this completes the proof of the theorem, since it directly implies that $\kappa^J = \kappa^{J-1}$ for $J \geq 4$, where the Shannon form was already established, and that the Shannon form and the corresponding parameter apply also to $J = 3$, then iteratively to $J = 2$, completing the logic.

Consider first $P \in C(\mu, A)$ where $|\Omega(\mu)| = J - 1$ and label the states in $\Omega(\mu)$ by $2 \leq j \leq J$. Consider now an additional state, $j = 1$, that in payoff terms is a replica of state $j = 2$,

$$u(a, 1) = u(a, 2) \text{ all } a \in A;$$

and the prior μ' that divides up the $\mu(1)$ equally between states $j = 1, 2$,

$$\mu'(j) = \begin{cases} \frac{\mu(1)}{2} \text{ if } j = 1, 2; \\ \mu(j) \text{ if } j \geq 3. \end{cases}$$

By construction (μ, A) is a basic version of (μ', A), $(\mu, A) \in \mathcal{B}(\mu', A)$. Hence by IUC we know that data set P' that agrees with P in all states and repeats in state $j = 1$ what P specifies in state $j = 1$,

$$P'(a|j) = \begin{cases} P(a|2) \text{ if } j = 1; \\ P(a|j) \text{ if } j \geq 2; \end{cases}$$

satisfies $P' \in C(\mu, A)$. Consider any chosen action $a \in \mathcal{A}(P')$ and define the corresponding revealed posterior $\bar{\gamma}^a_{P'}$. Since this is a CIR, we know that $\exists \lambda' = (Q', q') \in \hat{\Lambda}(\mu', A|K^J)$ such that $\mathbf{P}_\lambda = P'$ and with this posterior possible, and action a chosen deterministically at this posterior (by FIO),

$$Q'(\bar{\gamma}^a_{P'}) > 0 \text{ and } q'(a|\bar{\gamma}^a_{P'}) = 1.$$

Applying this to all chosen actions and noting that the strategy is optimal, we know that it satisfies the full bank of necessary and sufficient conditions for an optimal strategy. Specifically, given $a, b \in \mathcal{A}(P')$ we therefore know that the ILR equality holds:

$$\frac{\bar{\gamma}^a_{P'}(j)}{\exp(u(a,j)/\kappa^J)} = \frac{\bar{\gamma}^b_{P'}(j)}{\exp(u(b,j)/\kappa^J)} \text{ all } j; \tag{155}$$

together with the corresponding inequality: given $a \in \mathcal{A}(P')$ and $c \in A$,

$$\sum_{j=1}^{J} \left[\frac{\bar{\gamma}^a_{P'}(j)}{\exp(u(a,j)/\kappa^J)} \right] \exp(u(c,j)/\kappa^J) \leq 1. \tag{156}$$

Note that the revealed posteriors defined by data set P on chosen actions $a \in \mathcal{A}(P') = \mathcal{A}(P)$ are readily derived from those associated with data set P' by application of Bayes' rule,

$$\bar{\gamma}^a_P = \begin{cases} 0 & \text{if } j = 1; \\ \bar{\gamma}^a_{P'}(1) + \bar{\gamma}^a_{P'}(2) & \text{if } j = 2; \\ \gamma^a_{\lambda^{J-1}}(j) & \text{if } j \geq 3. \end{cases}$$

Now consider the strategy $\lambda = (Q_\lambda, q_\lambda) \in \Lambda(\mu, A)$ designed to precisely mirror $\lambda' = (Q', q') \in \hat{\Lambda}(\mu', A|K^J)$,

$$Q_\lambda(\bar{\gamma}^a_P) = Q'(\bar{\gamma}^a_{P'}) \text{ and } q_\lambda(a|\bar{\gamma}^a_P) = q'(a|\bar{\gamma}^a_{P'}).$$

Note by construction that $\mathbf{P}_\lambda = P$. Note also that validity of (155) and (156) survives this amalgamation,

$$\begin{aligned} a, b &\in \mathcal{A}(P) \Longrightarrow \frac{\bar{\gamma}_P^a(j)}{\exp(u(a,j)/\kappa^J)} = \frac{\bar{\gamma}_P^b(j)}{\exp(u(b,j)/\kappa^J)} \text{ all } j; \\ a &\in \mathcal{A}(P), c \in A \Longrightarrow \sum_{j=1}^{J} \left[\frac{\bar{\gamma}_P^a(j)}{\exp(u(a,j)/\kappa^J)} \right] \exp(u(c,j)/\kappa^J) \leq 1. \end{aligned}$$

Hence this strategy satisfies the necessary and sufficient conditions for $\lambda \in \hat{\Lambda}(\mu, A|K^J)$ as required.

The final step in the proof is to show the converse implication: given $\lambda = (Q_\lambda, q_\lambda) \in \hat{\Lambda}(\mu, A|K^J)$, the corresponding data satisfies $\mathbf{P}_\lambda \in C(\mu, A)$. We work again with the necessary and sufficient conditions (155) and (156) that characterize such optimal strategies: given $a, b \in \mathcal{A}(\lambda)$:

$$\begin{aligned} a, b &\in \mathcal{A}(\lambda) \Longrightarrow \frac{\gamma_\lambda^a(j)}{\exp(u(a,j)/\kappa^J)} = \frac{\gamma_\lambda^b(j)}{\exp(u(b,j)/\kappa^J)} \text{ all } j \geq 2; \\ a &\in \mathcal{A}(\lambda), c \in A \Longrightarrow \sum_{j=2}^{J} \left[\frac{\gamma_\lambda^a(j)}{\exp(u(a,j)/\kappa^J)} \right] \exp(u(c,j)/\kappa^J) \leq 1. \end{aligned}$$

We define the additional state $j = 1$ and the corresponding prior μ' precisely as above. We then derive strategy $\lambda' = (Q', q') \in \Lambda(\mu', A)$ from $\lambda = (Q_\lambda, q_\lambda) \in \Lambda(\mu, A)$ by reversing the process above. Given $a \in \mathcal{A}(\lambda)$ we first define corresponding posteriors,

$$\gamma_{\lambda'}^a = \begin{cases} \frac{\gamma_\lambda^a(2)}{2} & \text{if } j = 1, 2; \\ \gamma_\lambda^a(j) & \text{if } j \geq 3. \end{cases}$$

We then define the strategy $\lambda' = (Q', q') \in \Lambda(\mu', A)$ by

$$Q'(\gamma_{\lambda'}^a) = Q_\lambda(\gamma_\lambda^a) \text{ and } q_\lambda(a|\gamma_{\lambda'}^a) = q'(a|\gamma_\lambda^a).$$

To round out the proof, we note first that this strategy satisfies the full conditions (155) and (156) characterizing optimal strategies for the Shannon model. Hence we conclude that $\lambda' = (Q', q') \in \hat{\Lambda}(\mu', A|K^J)$. Hence, since this is a CIR, we conclude that $\mathbf{P}_{\lambda'} \in C(\mu, A')$. Finally, we apply IUC, which shows that the corresponding data satisfies $\mathbf{P}_\lambda \in C(\mu, A)$. This completes the proof. ■

Corollary 3: Data set $C \in \mathcal{C}$ has a Shannon representation if and only if it satisfies A1 through A9.

Proof. Suppose first that $C \in \mathcal{C}$ satisfies A1 through A9. By theorem 4 sufficiency we know that, since $\mathcal{C}$ satisfies A2 through A9, it has a UPS representation. At this point theorem 1 sufficiency applies, whereby, since $C \in \mathcal{C}$ has a UPS representation and satisfies A1, it has a Shannon representation.

To complete the proof, we show that having a Shannon representation implies satisfaction of A1 through A9. Theorem 1 necessity shows directly that if $C \in \mathcal{C}$ has a Shannon representation it satisfies A1. To prove that A2 through A9 are implied, we show that any $C \in \mathcal{C}$ that has Shannon representation is regular, $C \in \mathcal{C}^R$. This will complete the proof, since by theorem 4 necessity, we know that, since $C \in \mathcal{C}^R$ has a Shannon representation, it also has a UPS representation, hence satisfies A2 through A9.

To establish that any $C \in \mathcal{C}$ that has Shannon representation is regular, $C \in \mathcal{C}^R$, consider $\mu_1 \in \Gamma$ and $Q \in \Delta(\Gamma(\mu_1))$ with $\Gamma(Q) \subset \Gamma^C(\mu_1)$ and $\sum_{\gamma \in \Gamma(\mu_2)} \gamma Q(\gamma) = \mu_2$. It is implied directly from the invariant likelihood ratio characterization of optimal strategies that, given prior $\mu_1 \in \Gamma$, the set of observed posteriors associated with a Shannon representation is precisely the interior set of posteiors, $\Gamma^C(\mu_1) = \tilde{\Gamma}(\mu_1)$. Hence given $Q \in \Delta(\Gamma(\mu_1))$ with $\Gamma(Q) \subset \Gamma^C(\mu_1)$, we know that $\Omega(\gamma) = \Omega(\mu_1)$ all $\gamma \in \Gamma(Q)$. Hence the same applies to their weighted average,

$$\sum_{\gamma \in \Gamma(\mu_2)} \gamma Q(\gamma) = \mu_2.$$

Given that $\Omega(\mu_2) = \Omega(\mu_1)$, we know that $\Gamma^C(\mu_2) = \tilde{\Gamma}(\mu_2)$. Overall,

$$\Gamma(Q) \subset \Gamma^C(\mu_1) = \tilde{\Gamma}(\mu_1) = \tilde{\Gamma}(\mu_2),$$

establishing regularity, $C \in \mathcal{C}^R$, and completing the proof of the corollary. ■

References

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

Andrew Caplin and Daniel Martin. A testable theory of imperfect perception. *The Economic Journal*, 125(582):184–202, 2015.

Andrew Caplin, Mark Dean, and John Leahy. Rational inattention, optimal consideration sets and stochastic choice. 2016.

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

Jean-Charles Rochet. A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics*, 16(2):191–200, April 1987.

R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.