Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy

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

Online Appendix

1 Posterior-Separable Cost Functions

The starting point for our analysis is a data set **D** that can be represented by a posterior separable cost function. Here we reproduce the definition of posterior separability.

Definition 1 *An attention cost function K is* ***posterior separable*** *if, given $\mu \in \Delta\Omega$, there exists a function $T_\mu : \Delta(supp\ \mu) \to \bar{\mathbb{R}}$, such that*

1. *$T_\mu(\gamma)$ is strictly convex in γ.*
2. *$T_\mu(\gamma)$ is real-valued on int $\Delta(supp\ \mu)$*
3. *$T_\mu(\gamma) \geq 0$ with $T_\mu(\mu) = 0$*
4. *given any Bayes' consistent posteriors $Q \in \mathcal{Q}(\mu)$,*

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

The next section introduces notation. We then present and prove three important lemmas: the Lagrangian Lemma, Feasibility Implies Optimality, and Continuity.

1.1 Notation and Terminology

In stating and proving these lemmas we fix $\mu \in \Delta\Omega$. Recall that given a decision problem (μ, A), the set of feasible posterior-based strategies is $\Lambda(\mu, A)$

$$\Lambda(\mu, A) \equiv \{(Q, q) | Q \in \mathcal{Q}(\mu),\ q : \text{supp } Q \to \Delta A\}$$

[*]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

$$\mathcal{Q}(\mu) = \left\{ Q \in \Delta\Delta \text{supp } \mu \;\middle|\; \mu = \sum_{\gamma \in \text{supp } Q} \gamma Q(\gamma) \right\}$$

Recall that the value of strategy $(Q, q) \in \Lambda(\mu, A)$ given cost function K is

$$\begin{aligned} V(Q, q|\mu, A, K) &= U(Q, q|\mu, A) - K(\mu, Q) \\ &= U(Q, q|\mu, A) - \sum_{\gamma \in \text{supp } Q} Q(\gamma) T_\mu(\gamma), \end{aligned}$$

where

$$U(Q, q|\mu, A) = \sum_{\gamma \in \text{supp } Q} \sum_{a \in A} Q(\gamma) q(a|\gamma) u(\gamma, a).$$

Recall also that $N_\mu^a(\gamma)$ is the net utility of action a given posterior γ,

$$N_\mu^a(\gamma) \equiv u(\gamma, a) - T_\mu(\gamma),$$

where $u(\gamma, a) \equiv \sum_{\omega \in \text{supp } \gamma} \gamma(\omega) a(\omega)$ denotes the expected utility of action a given posterior γ. This provides an alternative method of identifying the value of $(Q, q) \in \Lambda(\mu, A)$ as,

$$V(Q, q|\mu, A, K) = \sum_{\gamma \in \text{supp } Q} \sum_{a \in A} Q(\gamma) q(a|\gamma) N_\mu^a(\gamma).$$

Finally, recall that $\hat{V}(\mu, A|K)$ refers to the supremum of V over all feasible strategies (Q, q) and $\hat{\Lambda}(\mu, A|K)$ is the set of strategies that achieve the supremum:

$$\begin{aligned} \hat{V}(\mu, A|K) &\equiv \sup_{(Q,q) \in \Lambda(\mu, A)} V(Q, q|\mu, A, K); \\ \hat{\Lambda}(\mu, A|K) &\equiv \left\{ (Q, q) \in \Lambda(\mu, A) \,|\, V(Q, q|\mu, A, K) = \hat{V}(\mu, A|K) \right\}. \end{aligned}$$

It is useful for the results that follow to assign "values" to strategies that are feasible from the prior $\mu' \in \Delta(\text{supp } \mu)$ rather than for the given prior $\mu \in \Delta(\Omega)$. We first define in a standard manner the gross utility of strategy $(Q', q') \in \Lambda(\mu', A)$ for $\mu' \in \Delta(\text{supp } \mu)$ as,

$$U(Q', q'|\mu', A) = \sum_{\gamma' \in \text{supp } Q'} \sum_{a \in A} Q'(\gamma') q'(a'|\gamma') u(\gamma', a')$$

We then define a "μ-based" net utility function for strategies $(Q', q') \in \Lambda(\mu', A)$ for $\mu' \in \Delta(\text{supp } \mu)$ using the costs associated with μ,

$$V_\mu(Q', q'|\mu', A, K) \equiv U(Q', q'|\mu', A) - \sum_{\gamma \in \text{supp } Q'} Q'(\gamma) T_\mu(\gamma). \tag{1.1}$$

Note that, as supp $\mu' \subseteq$supp μ, we can treat $\Delta(\text{supp } \mu')$ as a subset of $\Delta(\text{supp } \mu)$ by assuming that any distribution $\gamma \in \Delta(\text{supp } \mu')$ puts zero probability on any state $\omega \in$supp $\mu/$supp μ'. Thus T_μ is well defined for all relevant posteriors.

If $\mu' = \mu$ and therefore $(Q', q') \in \Lambda(\mu, A)$ then this is the standard value function,

$$\begin{aligned} V_\mu(Q', q'|\mu, A, K) &= U(Q', q'|\mu, A) - K(\mu, Q') \\ &= V(Q', q'|\mu, A, K). \end{aligned}$$

Finally, we use the following notation to denote the set of posteriors that are ever used as part of an optimal strategy from some prior given cost function K

$$\tilde{\Gamma}(\mu|K) = \{\gamma \in \text{supp } \mu | \exists A \subset \mathcal{A} \text{ and } (Q, q) \in \hat{\Lambda}(\mu, A|K) \text{ with } \gamma \in \text{supp } Q\}. \tag{1.2}$$

We will show in Corollary 1.1 below that $\tilde{\Gamma}(\mu|K)$ is equal to the set $\hat{\Gamma}(\mu|K)$, which is the of posteriors at which T_μ is subdifferentiable.[1]

The posterior separable cost function K and choice set A are fixed for much of the analysis that follows, so we suppress them whenever possible, for example writing $V_\mu(Q', q'|\mu')$ in place of $V_\mu(Q', q'|\mu', A, K)$ and $\Lambda(\mu')$ in place of $\Lambda(\mu', A)$ for feasible strategies.

1.2 The Lower Epigraph of V_μ

Key to the proof of the Lagrangian lemma is the lower epigraph of V_μ for a fixed A. To specify we reduce the dimension of the state space by defining $J = |\text{supp } \mu|$, correspondingly index states by $1 \leq j \leq J$, and identify each prior by $\mu \in \mathbb{R}_+^{J-1}$ with $\sum_{j=1}^{J-1} \mu(j) \leq 1$ and with $\mu(J) = 1 - \sum_{j=1}^{J-1} \mu(j)$ left as implicit. Let Δ^{J-1} denote the set of vectors in $\mathbb{R}_+^{J-1}$ with $\sum_{j=1}^{J-1} \mu(j) \leq 1$.

Definition 2 *The lower epigraph of $V_\mu(Q', q'|\mu')$ is*

$$\mathcal{E}(\mu) \equiv \left\{(y, \mu') \in \mathbb{R} \times \Delta^{J-1} \ |y \leq V_\mu(Q', q'|\mu') \text{ some } (Q', q') \in \Lambda(\mu')\right\}.$$

What makes this set of interest is that it is convex and therefore we can apply the separating hyperplane theorem. This is a key step in the Lagrangian characterization of optimal strategies in Lemma 1.

Remark 1 *For every $\mu \in \Delta(\Omega)$, $\mathcal{E}(\mu)$ is a convex subset of $\mathbb{R}^J$.*

To establish convexity we introduce a mixture operation on the collection of sets $\Lambda(\mu')$ for all $\mu' \in \Delta(\text{supp } \mu)$. Given any finite set of strategies, $\{(Q_l, q_l) \in \Lambda(\mu_l)\}_{1 \leq l \leq L}$, and strictly positive probability weights $\{\alpha_l\}_{1 \leq l \leq L}$, we define the corresponding mixture strategy (Q_α, q_α) by:

$$\begin{aligned} Q_\alpha(\gamma) &= \sum_l \alpha_l Q_l(\gamma) \text{ all } \gamma \in \cup_l \text{supp } Q_l \equiv \text{supp } Q_\alpha; \\ q_\alpha(a|\gamma) &= \frac{\sum_l \alpha_l q_l(a|\gamma) Q_l(\gamma)}{Q_\alpha(\gamma)} \text{ all } \gamma \in \text{supp } Q_\alpha, \ a \in A; \end{aligned}$$

[1] See Rockafeller [1970] for an extensive discussion of the subdifferentiability of convex functions.

Defining $\mu(\alpha) = \sum_l \alpha_l \mu_l$ as the corresponding weighted average of priors it is immediate that $(Q_\alpha, q_\alpha) \in \Lambda(\mu(\alpha))$ and it can readily be confirmed that V_μ is linear in the mixture operation,

$$V_\mu(Q_\alpha, q_\alpha|\mu(\alpha)) = \sum_l \alpha_l V_\mu(Q_l, q_l|\mu_l).$$

This directly implies that $\mathcal{E}(\mu)$ is a convex.

1.3 Three Lemmas

In this section we establish three lemmas concerning posterior-separable models. The first two, the Lagrangian lemma and Feasibility Implies Optimality, present properties of the solution to the problem of maximizing expected utility subject to a posterior separable cost function. The third, Continuity, presents a property of posterior-separable cost functions that represent data.

1.3.1 The Lagrangian Lemma

We now establish the Lagrangian lemma, which makes direct use of the above mixture operation as well as of the convexity of $\mathcal{E}(\mu)$. In stating the lemma we reintroduce the fixed posterior-separable cost function K and choice set A. We suppress them where clear in the body of the proof.

Lemma 1 (Lagrangean): Given a posterior-separable cost function K and decision problem (μ, A), $(Q, q) \in \hat{\Lambda}(\mu, A|K)$ if and only if $\exists \theta \in \mathbb{R}^{J-1}$ such that, given $\gamma \in$supp Q and $a \in A$ with $q(a|\gamma) > 0$,

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

all $\gamma' \in \Delta(\text{supp } \mu)$ and $a' \in A$.

Proof. We first show necessity: that this holds for any optimal strategy $(Q, q) \in \hat{\Lambda}(\mu)$. We note first that optimality implies that $(V(Q, q|\mu), \mu)$ is an upper boundary point of $\text{cl}(\mathcal{E}(\mu))$, the closure of $\mathcal{E}(\mu)$. If not then there exists $y > V(Q, q|\mu)$ such that $(y, \mu) \in \text{cl}(\mathcal{E}(\mu))$, and hence also $(y', \mu) \in \mathcal{E}(\mu)$ such that $y' > V(Q, q|\mu)$. Hence there exists $(Q', q') \in \Lambda(\mu)$ with $V(Q', q'|\mu) > V(Q, q|\mu)$, contradicting the optimality of (Q, q). Since $(V(Q, q|\mu), \mu)$ is a boundary point of the closure of $\mathcal{E}(\mu)$, which inherits convexity, we can identify a supporting hyperplane $\theta(j)$ for $0 \leq j \leq J-1$ with $\theta(j) \neq 0$ for some j such that,

$$\theta(0)y - \sum_{j=1}^{J-1} \theta(j)\mu'(j) \leq \theta(0)V(Q, q|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j), \tag{1.3}$$

all $(y, \mu') \in \mathcal{E}(\mu)$. We show now that $\theta(0) > 0$.

First, suppose to the contrary that $\theta(0) < 0$. In this case we can renormalize to $\theta(0) = -1$ to conclude that,

$$-y - \sum_{j=1}^{J-1} \theta(j)\mu'(j) \leq -V(Q, q|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j),$$

which is a clear contradiction, since the set of feasible y's is unbounded below and so the left hand side is unbounded above.

Now suppose instead that $\theta(0) = 0$. If so we conclude from (1.3) that,

$$\sum_{j=1}^{J-1} \theta(j)\mu(j) \leq \sum_{j=1}^{J-1} \theta(j)\mu'(j). \tag{1.4}$$

all $\mu' \in \Delta(\text{supp } \mu)$. Note that by definition $\mu(j) > 0$ all j. Hence,

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

To minimize the expression on the RHS in (1.4), one can set $\mu'(\bar{j}) = 1$ where the index $\bar{j}$ is chosen so that,

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

Hence (1.4) can be valid only if $\theta(j) = \theta(\bar{j}) = \bar{\theta}$ for all j. Hence what is required is,

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

all $\mu' \in \Delta(\text{supp } \mu)$. Since $0 < \sum_{j=1}^{J-1}\mu(j) < 1$ while $\sum_{j=1}^{J-1}\mu'(j)$ has a range that includes 0 and 1, this is impossible unless $\bar{\theta} = 0$, a contradiction to the non-zero separating plane in (1.3).

Given that $\theta(0) > 0$, we renormalize to $\theta(0) = 1$ in (1.3) to conclude that,

$$y - \sum_{j=1}^{J-1}\theta(j)\mu'(j) \leq V(Q, q|\mu) - \sum_{j=1}^{J-1}\theta(j)\mu(j). \tag{1.5}$$

all $(y, \mu') \in \mathcal{E}(\mu)$. Given $\gamma' \in \Delta(\text{supp } \mu)$ this applies for all $y \leq V_\mu(Q', q'|\gamma')$ for some $(Q', q') \in \Lambda(\gamma')$. In particular this includes those strategies which are inattentive, and so $Q'(\gamma') = 1$, and involve deterministic choice of any action $a' \in A$, so $q'(a'|\gamma') = 1$. For these particular strategies,

$$V_\mu(Q', q'|\gamma') = N_\mu^{a'}(\gamma').$$

This implies that $(N_\mu^{a'}(\gamma'), \gamma') \in \mathcal{E}(\mu)$, hence by (1.5)

$$N_\mu^{a'}(\gamma') - \sum_{j=1}^{J-1}\theta(j)\gamma'(j) \leq V(Q, q|\mu) - \sum_{j=1}^{J-1}\theta(j)\mu(j). \tag{1.6}$$

all $a' \in A$ and $\gamma' \in \Delta\text{supp } \mu$.

All that remains is to show that we can replace $V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j)$ with $N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j)$ for any $\gamma\in$supp Q and $a\in A$ with $q(a|\gamma)>0$ on the right hand side of (1.6). First observe that

$$V(Q,q|\mu)=\sum_{\gamma\in suppQ}Q(\gamma)\sum_{a\in A}q(a|\gamma)N_\mu^a(\gamma)$$

and

$$\sum_{j=1}^{J-1}\theta(j)\mu(j)=\sum_{\gamma\in suppQ}Q(\gamma)\sum_{a\in A}q(a|\gamma)\left[\sum_{j=1}^{J-1}\theta(j)\gamma(j)\right],$$

as $\sum_{a\in A}q(a|\gamma)=1$ for every $\gamma\in$supp Q and $\sum_{\gamma\in suppQ}Q(\gamma)\gamma(j)=\mu(j)$ for every $1\leq j\leq J-1$. Thus we have

$$V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j)=\sum_{\gamma\in suppQ}Q(\gamma)\sum_{a\in A}q(a|\gamma)\left[N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j)\right]. \tag{1.7}$$

This expresses $V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j)$ as a weighted average of terms $N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j)$ with strictly positive weights. By (1.6), none of these terms can be strictly above $V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j)$.

$$N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j)\leq V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j).$$

Hence they must all be equal to it for any weighted average with strictly positive weights to be equal to it,

$$N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j)=V(Q,q|\mu)-\sum_{j=1}^{J-1}\theta(j)\mu(j),$$

all $\gamma\in$supp Q and $a\in A$ with $q(a|\gamma)>0$. Combined with (1.6) this establishes that indeed, given $\gamma\in$supp Q and $a\in A$ with $q(a|\gamma)>0$,

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

all $\gamma'\in\Delta$(supp μ) and $a'\in A$, and with it necessity.

With regard to sufficiency, consider $(Q,q)\in\Lambda(\mu,A)$ for which there exists $\exists\theta\in\mathbb{R}^{J-1}$ such that, given $\gamma\in$supp Q and $a\in A$ with $q(a|\gamma)>0$,

$$N_\mu^{a'}(\gamma')-\sum_{j=1}^{J-1}\theta(j)\gamma'(j)\leq N_\mu^a(\gamma)-\sum_{j=1}^{J-1}\theta(j)\gamma(j), \tag{1.8}$$

all $\gamma' \in \Delta(\text{supp } \mu)$ and $a' \in A$. As established in (1.7) we can rewrite

$$V(Q, q|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j) = \sum_{\gamma \in suppQ} Q(\gamma) \sum_{a \in A} q(a|\gamma) \left[N_\mu^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \right].$$

Moreover, condition (1.8) implies that for any $\gamma, \gamma' \in$supp Q and $a, a' \in A$ with $q(a|\gamma) > 0$ and $q(a'|\gamma') > 0$ it must be the case that

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

and so, as $\sum\limits_{a \in A} q(a|\gamma) = 1 \; \forall \; \gamma \in$supp Q and $\sum_{\gamma \in suppQ} Q(\gamma) = 1$

$$N_\mu^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) = V(Q, q|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j), \tag{1.9}$$

all $\gamma \in$supp Q and $a \in A$ with $q(a|\gamma) > 0$.

Now consider arbitrary strategy $(Q', q') \in \Lambda(\mu, A)$ and repeat precisely the sequence of steps that led to (1.7) to obtain

$$V(Q', q'|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j) = \sum_{\gamma' \in \text{supp } Q'} Q'(\gamma') \sum_{a' \in A} q'(a'|\gamma') \left[N_\mu^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \right]. \tag{1.10}$$

Picking an arbitrary $\gamma \in$supp Q and $a \in A$ with $q(a|\gamma) > 0$ we have from (1.8) that

$$\begin{aligned} & \sum_{\gamma' \in \text{supp } Q'} Q'(\gamma') \sum_{a' \in A} q'(a'|\gamma') \left[N_\mu^{a'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \right] \\ \leq \; & \sum_{\gamma' \in \text{supp } Q'} Q'(\gamma') \sum_{a' \in A} q'(a'|\gamma') \left[N_\mu^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \right] \\ = \; & N_\mu^a(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \\ = \; & V(Q, q|\mu) - \sum_{j=1}^{J-1} \theta(j)\mu(j), \end{aligned} \tag{1.11}$$

where the third line follows from the fact that $\sum\limits_{a \in A} q(a|\gamma) = 1 \; \forall \; \gamma \in$supp Q' and $\sum_{\gamma \in \text{supp } Q'} Q'(\gamma) = 1$, and the fourth follows from (1.9) above. Combining (1.10) and (1.11) and subtracting $\sum_{j=1}^{J-1} \theta(j)\mu(j)$ from each side gives

$$V(Q', q'|\mu) \leq V(Q, q|\mu),$$

establishing that indeed $(Q, q) \in \hat{\Lambda}(\mu)$ and with it the theorem. ■

1.3.2 Feasibility Implies Optimality

The next Lemma establishes that one can assign to any posterior that is ever optimal, an action such that the net utility of using that action at that posterior is zero, while the net utility of any other posterior when choosing that action is less than zero. Coupled with the Lagrangian lemma that means that one can construct decision problems in which it is optimal to use any collection of such posteriors.

Lemma 1.2 (Feasibility Implies Optimality – first version): Given $\mu \in \Delta\Omega$, a posterior-separable cost function K, and a corresponding $T_\mu : \Delta(\text{supp } \mu) \to \bar{\mathbb{R}}$, there exists a one-to-one function $f : \tilde{\Gamma}(\mu|K) \to \mathcal{A}$ with image $\mathcal{A}^f_\mu$ such that, given $\gamma \in \tilde{\Gamma}(\mu|K)$,

$$N^{f(\gamma)}_\mu(\gamma') \leq N^{f(\gamma)}_\mu(\gamma) = 0, \tag{1.12}$$

all $\gamma' \in \Delta(\text{supp } \mu)$. Moreover, given decision problem (μ, A) with $A \subset \mathcal{A}^f_\mu$, any $(Q, q) \in \Lambda(\mu, A)$ with supp $Q \subset \left\{f^{-1}(a)\right\}_{a \in A}$ that deterministically picks actions defined by this function is optimal,

$$q(f(\gamma)|\gamma) = 1 \text{ for all } \gamma \in \text{supp } Q \Longrightarrow (Q, q) \in \hat{\Lambda}(\mu, A|K).$$

Proof. We define the function satisfying (1.12) in two phases. First we consider interior beliefs γ with $\gamma(\omega) > 0$ all $\omega \in$supp μ (we describe why the same approach will not necessarily work for boundary posteriors below). We note that since $-T_\mu(\gamma)$ is a strictly concave function, its lower epigraph is a convex set. Hence $(-T_\mu(\gamma), \gamma)$ is an upper boundary point of the closure of this lower epigraph (as detailed when applying the supporting hyperplane theorem in the first paragraph of Lemma 1). Since $(-T_\mu(\gamma), \gamma)$ is a boundary point of the closure of $\mathcal{E}(\mu)$, which inherits convexity, we can then apply the supporting hyperplane theorem to this point to identify a supporting hyperplane $\theta(j)$ for $0 \leq j \leq J - 1$ with $\theta(j) \neq 0$ for some j such that, Hence we can apply the supporting hyperplane theorem at $(-T_\mu(\gamma), \gamma)$ to identify multipliers $\beta(0)$ and $-\beta(j)$ on $1 \leq j \leq J-1$, not all zero, such that,

$$-\beta(0)T_\mu(\gamma') - \sum_{j=1}^{J-1} \beta(j)\gamma'(j) \leq -\beta(0)T_\mu(\gamma) - \sum_{j=1}^{J-1} \beta(j)\gamma(j), \tag{1.13}$$

all $\gamma' \in \Delta(\text{supp } \mu)$.

Given that $\gamma(\omega) > 0$ all $\omega \in$supp μ we can mimic the proof in the third paragraph of Lemma 1 above to establish that $\beta(0) \neq 0$. It is also not possible that $\beta(0) < 0$. To see this suppose this were so. In this case we could renormalize to $\beta(0) = -1$ in (1.13), implying

$$T_\mu(\gamma') - \sum_{j=1}^{J-1} \beta(j)\gamma'(j) \leq T_\mu(\gamma) - \sum_{j=1}^{J-1} \beta(j)\gamma(j), \tag{1.14}$$

all $\gamma' \in \Delta(\text{supp } \mu)$. Given that $\gamma(\omega) > 0$ all $\omega \in$supp μ, we can find two distinct beliefs $\gamma'_1, \gamma'_2 \in \Delta(\text{supp } \mu)$ such that $\gamma = \frac{\gamma'_1+\gamma'_2}{2}$. Averaging inequality (1.14) applied to each of γ'_1, γ'_2 separately, we conclude that,

$$\begin{aligned}
\frac{T_\mu(\gamma'_1) - \sum_{j=1}^{J-1}\beta(j)\gamma'_1(j) + T_\mu(\gamma'_2) - \sum_{j=1}^{J-1}\beta(j)\gamma'_2(j)}{2} &= \frac{T_\mu(\gamma'_1)+T_\mu(\gamma'_2)}{2} - \sum_{j=1}^{J-1}\beta(j)\gamma(j) \\
&\leq T_\mu(\gamma) - \sum_{j=1}^{J-1}\beta(j)\gamma(j).
\end{aligned}$$

We conclude therefore that,

$$0.5T_\mu(\gamma'_1) + 0.5T_\mu(\gamma'_2) \leq T_\mu(\gamma) = T_\mu(\frac{\gamma'_1+\gamma'_2}{2}), \tag{1.15}$$

which contradicts strict convexity of T_μ.

With $\beta(0) > 0$, we can renormalize to $\beta(0) = 1$ in (1.13) and flip signs to conclude that,

$$T_\mu(\gamma') + \sum_{j=1}^{J-1}\beta(j)\gamma'(j) \geq T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j). \tag{1.16}$$

all $\gamma' \in \Delta(\text{supp } \mu)$. Now define action $f(\gamma) \in \mathcal{A}$ so that, for $1 \leq k \leq J$,

$$f^\gamma(k) = \begin{cases} T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j) - \beta(k) \text{ for } 1 \leq k \leq J-1; \\ T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j) \text{ for } k = J; \end{cases} .$$

where $f^\gamma(k)$ is the k-th coordinate of $f(\gamma)$. By construction note that,

$$\begin{aligned}
N_\mu^{f(\gamma)}(\gamma) &= \sum_{k=1}^{J} f^\gamma(k)\gamma(k) - T_\mu(\gamma) \\
&= \sum_{k=1}^{J-1}\left[T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j) - \beta(k)\right]\gamma(k) + \left[T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j)\right]\gamma(J) - T_\mu(\gamma) \\
&= T_\mu(\gamma) + \sum_{j=1}^{J-1}\beta(j)\gamma(j) - \sum_{k=1}^{J-1}\beta(k)\gamma(k) - T_\mu(\gamma) \\
&= 0.
\end{aligned}$$

Also by construction note that for any $\gamma' \in \Delta(\text{supp } \mu)$,

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

where the last inequality derives directly from equation (1.16).

Given a boundary posterior $\gamma \in$supp Q with $\gamma(\omega) = 0$ some $\omega \in$supp μ we cannot guarantee that the multiplier $\beta(0)$ in (1.13) is non-zero (costs based on Shannon mutual information are a counterexample). The remaining cases therefore involve boundary posteriors that are part of an optimal strategy for some decision problem - i.e. $\gamma \in \tilde{\Gamma}(\mu|K)$. By definition there exists a decision problem (μ, A) and an optimal strategy $(Q, q) \in \hat{\Lambda}(\mu, A)$ with $\gamma \in$supp Q. Hence by the Lemma 1 there exists $\theta \in \mathbb{R}^{J-1}$ such that, for any a with $q(a|\gamma) > 0$,

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

all $\gamma' \in \Delta(\text{supp } \mu)$. Rearrangement yields,

$$\sum_{j=1}^{J} a(j)\gamma'(j) - \sum_{j=1}^{J-1}\theta(j)\gamma'(j) - \left[\sum_{j=1}^{J} a(j)\gamma(j) - \sum_{j=1}^{J-1}\theta(j)\gamma(j)\right] + T_\mu(\gamma) - T_\mu(\gamma') \leq 0. \tag{1.17}$$

Now define action $f(\gamma) \in \mathcal{A}$ so that,

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

By construction, given $\gamma' \in \Delta(\text{supp } \mu)$,

$$N_\mu^{f(\gamma)}(\gamma') = \sum_{j=1}^{J} a(j)\gamma'(j) - \sum_{j=1}^{J-1}\theta(j)\gamma'(j) - \left[\sum_{j=1}^{J} a(j)\gamma(j) - \sum_{j=1}^{J-1}\theta(j)\gamma(j) - T_\mu(\gamma)\right] - T_\mu(\gamma') \leq 0,$$

by (1.17). It is direct that this value is zero at $\gamma' = \gamma$,

$$N_\mu^{f(\gamma)}(\gamma) = \sum_{j=1}^{J} a(j)\gamma(j)) - \sum_{j=1}^{J-1}\theta(j)\gamma(j) - \left[\sum_{j=1}^{J} a(j)\gamma(j) - \sum_{j=1}^{J-1}\theta(j)\gamma(j) - T_\mu(\gamma)\right] - T_\mu(\gamma) = 0.$$

Hence we have established existence of a function f :intΔ(supp μ) $\cup\, \tilde{\Gamma}(\mu|K) \to \mathcal{A}$ with the properties called for in (1.12). We define $\mathcal{A}_\mu^f$ as the union of these actions,

$$\mathcal{A}_\mu^f = \cup_{\text{int}\Delta(\text{supp }\mu)\cup\tilde{\Gamma}(\mu|K)} f(\gamma).$$

We now pick an arbitrary Bayes' consistent distribution for prior μ, $Q \in \mathcal{Q}(\mu)$ with supp $Q \subset$intΔ(supp μ) $\cup\, \tilde{\Gamma}(\mu|K)$ and consider the decision problem (μ, A) with $\cup_{\gamma\in\text{supp }Q} f(\gamma) \subset A \subset \mathcal{A}_\mu^f$. We then specify the strategy $(Q, q) \in \Lambda(\mu, A)$ that deterministically picks the action defined by this function,

$$q(f(\gamma)|\gamma) = 1,$$

all $\gamma \in$supp Q.

We show now that $(Q, q) \in \hat{\Lambda}(\mu, \cup_{\gamma\in\text{supp }Q} f(\gamma))$. This follows from the sufficiency aspect of the Lagrangian Lemma (Lemma 1) with $\theta(j) = 0\ \forall\ 1 \leq j \leq J-1$. What we need to establish is that,

$$N_\mu^{f(\gamma')}(\gamma'') \leq N_\mu^{f(\gamma)}(\gamma).$$

all $\gamma, \gamma' \in$supp Q and $\gamma'' \in \Delta$(supp μ). This is immediate from the construction since,

$$N_\mu^{f(\gamma')}(\gamma'') \leq N_\mu^{f(\gamma')}(\gamma') = N_\mu^{f(\gamma)}(\gamma) = 0.$$

At this point application of the Lagrangian lemma establishes that indeed $(Q, q) \in \hat{\Lambda}(\mu, \cup_{\gamma\in\text{supp }Q} f(\gamma))$. To complete the proof, note that the 1-1 nature of $f(\gamma)$ follows since otherwise there exists an optimal strategy that selects the same action at two different posteriors. Given that T_μ is strictly convex, it would strictly lower costs without changing gross expected utility to pool the two posteriors and take this action only at the corresponding weighted average posterior. ■

The theorem is proved by finding a supporting hyperplane at each observed posterior and constructing a problem that satisfies the Lagrangian lemma. An immediate implication is that the set of observed posteriors, $\tilde{\Gamma}$, is equal to the set of posteriors at which T_μ is subdifferentiable, $\hat{\Gamma}$.

Corollary 1.1: $\hat{\Gamma}(\mu|K) = \tilde{\Gamma}(\mu|K)$.

As a convex function is subdifferentiable on the interior of its effective domain (Rockafellar [970], Theorem 23.4), it follows that intΔ(supp μ) $\subset \hat{\Gamma}(\mu|K)$.

Corollary 1.2: intΔ(supp μ) $\subset \tilde{\Gamma}(\mu|K)$.

Together Lemma 1.1, Corollary 1.1 and Corollary 1.2 imply Lemma 2 in the text of the paper.

Lemma 2 (Feasibility Implies Optimality – final version): Fix $\mu \in \Delta\Omega$ and a posterior-separable cost function K then

1. Given $Q \in \mathcal{Q}(\mu)$ such that $\gamma \in$supp Q implies $\gamma \in \hat{\Gamma}(\mu|K)$, there exist A and q such that $(Q, q) \in \hat{\Lambda}(\mu, A|K)$ and each action $a \in A$ is chosen deterministically from one posterior $\gamma \in$supp Q.
2. Given $\gamma \in \hat{\Gamma}(\mu|K)$, it is always possible to find an action $a \in A$ such that for all $\gamma' \in \Delta$supp μ we have $N^a_\mu(\gamma') \leq 0$ with equality if $\gamma' = \gamma$.
3. int$(\Delta(\text{supp } \mu)) \subset \hat{\Gamma}(\mu|K)$.

1.3.3 Continuity

Convex functions are continuous on the interior of their effective domain, but may be discontinuous on the boundary. The following lemma states that if **D** has a posterior-separable representation K, then, without loss of generality, we may take T_μ to be continuous on the boundary. It rests on a direct implication of existence of a representation of the data, which is that all decision problems have solutions, since behavior is always observed. Note that we allow the cost function to be infinite, so that the limit below can correspondingly be infinity.

Lemma 3 (Continuity): Consider a data set with **D** a posterior-separable representation

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

Let

$$\hat{K}(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma) \hat{T}_\mu(\gamma).$$

where $\hat{T}_\mu(\bar{\gamma}) = T_\mu(\bar{\gamma})$ for all $\bar{\gamma} \in$int$(\Delta(\text{supp } \mu))$ and for $\bar{\gamma}$ on the boundary,

$$\hat{T}_\mu(\bar{\gamma}) = \lim_{\gamma \to \bar{\gamma}} T_\mu(\gamma) \in \bar{\mathbb{R}}_+,$$

where the limit is taken with respect to $\gamma \in$int$(\Delta(\text{supp } \mu))$. Then $\hat{K}$ also represents **D**.

Proof. Fix an arbitrary prior μ and and consider an arbitrary posterior $\bar{\gamma}$ on the boundary of $\Delta(\text{supp } \mu)$, $\bar{\gamma} \in \partial\left(\Delta\left(\text{supp } \mu\right)\right)$, such that $T(\bar{\gamma}) \neq \hat{T}_\mu(\bar{\gamma}) = \lim_{\gamma \to \bar{\gamma}} T_\mu(\gamma)$ for $\gamma \in$int$(\Delta($supp $\mu))$. Since T is discontinuous at $\bar{\gamma}$, T is not subdifferentiable at $\bar{\gamma}$. It follows from Corollary 1.1 that $\bar{\gamma} \notin \hat{\Gamma}(\mu|K)$. If also $\bar{\gamma} \notin \hat{\Gamma}(\mu|\hat{K})$ then this discontinuity does not affect the ability of $\hat{K}$ to represents **D**. Therefore suppose $\bar{\gamma} \in \hat{\Gamma}(\mu|\hat{K})$ and consider $\gamma' = \frac{\mu - \alpha\gamma}{1-\alpha}$. For a suitable choice of α, $\gamma' \in$int$(\Delta(\text{supp } \mu))$, and hence by Corollary 1.2, $\gamma' \in \hat{\Gamma}(\mu|\hat{K})$. Since both $\bar{\gamma}$ and γ' are in $\hat{\Gamma}(\mu|\hat{K})$, Feasibility Implies Optimality (Lemma 1.2) states that there exists a problem a decision problem (μ, A) with two actions and an optimal policy $(Q, q) \in \hat{\Lambda}(\mu, A|\hat{K})$ such that supp $Q = \{\bar{\gamma}, \gamma'\}$ and each action is chosen from one of these two posteriors. Let $\bar{a}$ denote the action chosen from $\bar{\gamma}$ and a' the action chosen from γ'.

Now consider $\mathbf{P}_{(\mu,A)} \in \mathbf{D}$ with revealed policy $(\mathbf{Q}, \mathbf{q})$. Since there are two actions we may suppose that there are at most two revealed posteriors: $\boldsymbol{\gamma}_{\bar{a}}$ and $\boldsymbol{\gamma}_{a'}$. Since K represents **D** and $\bar{\gamma} \notin \hat{\Gamma}(\mu|K)$, neither $\boldsymbol{\gamma}_{\bar{a}}$ nor $\boldsymbol{\gamma}_{a'}$ is equal to $\bar{\gamma}$. Since $(Q, q) \in \hat{\Lambda}(\mu, A|\hat{K})$, and $(\mathbf{Q}, \mathbf{q})$ is feasible, $V(Q, q|\mu, A, \hat{K}) \geq V(\mathbf{Q}, \mathbf{q}|\mu, A, \hat{K})$. If this inequality is strict then the continuity of $\hat{K}$ implies that there exist better choices than $(\mathbf{Q},\mathbf{q})$ in the neighborhood of (Q,q), a contradiction. Hence $V(Q,q|\mu,A,\hat{K}) = V(\mathbf{Q},\mathbf{q}|\mu,A,\hat{K})$ or

$$Q_{\bar{\gamma}}\,(\bar{a}\cdot\bar{\gamma}) + Q_{\gamma'}(a'\cdot\gamma') - Q_{\bar{\gamma}}T_\mu(\bar{\gamma}) - Q_{\gamma'}T_\mu(\gamma') = \mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}\,(\bar{a}\cdot\boldsymbol{\gamma}_{\bar{a}}) + \mathbf{Q}_{\boldsymbol{\gamma}_{a'}}(a'\cdot\boldsymbol{\gamma}_{a'}) - \mathbf{Q}_{\bar{\gamma}}T_\mu(\boldsymbol{\gamma}_{\bar{a}}) - \mathbf{Q}_{\gamma'}T_\mu(\boldsymbol{\gamma}_{a'})$$

Construct the policy $(\tilde{Q},\tilde{q})$ so that $\bar{a}$ is chosen from the posterior $\frac{Q_{\bar{\gamma}}\bar{\gamma}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}\boldsymbol{\gamma}_{\bar{a}}}{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}$ with probability $\frac{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}{2}$ and $\frac{Q_{\gamma}\gamma'+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}\boldsymbol{\gamma}_{a'}}{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}$ with probability $\frac{Q_{\gamma'}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}{2}$. This policy has gross expected utility,

$$\frac{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}{2}\left(\bar{a}\cdot\frac{Q_{\bar{\gamma}}\bar{\gamma}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}\boldsymbol{\gamma}_{\bar{a}}}{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}\right) + \frac{Q_{\gamma'}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}{2}\left(a'\cdot\frac{Q_{\gamma'}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}{2}\right),$$

which is the equally weighted average of that associated with (Q,q) and $(\mathbf{Q},\mathbf{q})$. The cost of this strategy is correspondingly,

$$\frac{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}{2}\hat{T}_\mu\left(\frac{Q_{\bar{\gamma}}\bar{\gamma}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}\boldsymbol{\gamma}_{\bar{a}}}{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{\bar{a}}}}\right) + \frac{Q_{\gamma'}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}{2}\hat{T}_\mu\left(\frac{Q_{\gamma}\gamma'+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}\boldsymbol{\gamma}_{a'}}{Q_{\bar{\gamma}}+\mathbf{Q}_{\boldsymbol{\gamma}_{a'}}}\right),$$

Given the strict convexity of $\hat{T}_\mu$ the cost of this policy is strictly less than the equally weighted average of that associated with (Q,q) and $(\mathbf{Q},\mathbf{q})$. Hence the value of this strategy is strictly higher than the weighted average of the two supposedly optimal policies $(\mathbf{Q},\mathbf{q})$ and (Q,q)

$$V(\tilde{Q},\tilde{q}|\mu,A,\hat{K}) > \frac{V(\mathbf{Q},\mathbf{q}|\mu,A,\hat{K}) + V(Q,q|\mu,A,\hat{K})}{2}.$$

Given that $V(Q,q|\mu,A,\hat{K}) = V(\mathbf{Q},\mathbf{q}|\mu,A,\hat{K})$, this contradicts the assumption that (Q,q) is optimal given $\hat{K}$ and establishes the lemma. ■

2 Proof of Theorem 1

Theorem 1 states that the axiom Locally Invariant Posteriors is necessary and sufficient for a data set with a posterior separable representation to have a uniformly posterior separable representation. We reproduce the definitions of uniform posterior separability, Locally Invariant Posteriors, and the theorem here.

Definition 3 *A posterior-separable cost function K is* ***uniformly posterior separable****, if for each finite subset $\bar{\Omega}\subset\Omega$ there exists a strictly convex function $T_{\bar{\Omega}}:\Delta(\bar{\Omega})\to\bar{\mathbb{R}}$ such that, for all $\mu\in\Delta(\Omega)$ and $Q\in\mathcal{Q}(\mu)$,*

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

Axiom 1 Locally Invariant Posteriors: Consider any decision problem (μ,A) and state dependent stochastic choice data $\mathbf{P}_{(\mu,A)}\in\mathbf{D}_{(\mu,A)}$ with revealed strategy $(\mathbf{Q},\mathbf{q})$ such that $\mathbf{q}$ is deterministic in the sense that for all $\gamma\in\text{supp }\mathbf{Q}$ there exists $a\in A$ such that $\mathbf{q}(a|\gamma)=1$. Consider (Q',q') with $\sum_{\gamma\in\text{ supp }Q'}\gamma Q'(\gamma)=\mu'$. If supp Q' ⊂supp $\mathbf{Q}$, supp μ' =supp μ,and $q'(\gamma)=q(\gamma)$ for all γ ∈supp Q', then $P_{(Q',q')}\in\mathbf{D}_{(\mu',A)}$.

Theorem 1: A data set **D** with a posterior-separable representation has a uniformly posterior-separable representation if and only if it satisfies Locally Invariant Posteriors (Axiom 1).

As a preliminary observation, we note that, related to any posterior separable cost function is a class of representations based on affine transforms of T_μ. This class is identified in the following remark.

Remark 2 *Let K be a* ***posterior-separable*** *cost function with associated $T_\mu : \Delta(supp\ \mu) \to \bar{\mathbb{R}}$ for some $\mu \in \Delta(\Omega)$. Given $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^{|supp\ \mu|}$, define $\hat{T}_\mu : \Delta(supp\ \mu) \to \bar{\mathbb{R}}$ as*

$$\hat{T}_\mu(\gamma) = \begin{cases} \infty & \textit{if } T_\mu(\gamma) = \infty \\ T_\mu(\gamma) + \alpha + \beta \cdot \gamma & \textit{otherwise} \end{cases}$$

Then

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

2.1 Theorem 1: Sufficiency

Theorem 1 (Sufficiency) A data set **D** with a posterior-separable representation has a uniformly posterior-separable representation if it satisfies Locally Invariant Posteriors (Axiom 1).

Proof. The result relates to a fixed set of possible states which we will refer to for simplicity as $\bar{\Omega}$. We establish the result by defining μ_1 be the uniform prior over Ω and picking any distinct prior μ_2 with the same support. We identify the corresponding strictly convex functions $T_i : \Delta\bar{\Omega} \to \mathbb{R}$ for which,

$$K(\mu_i, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma)T_i(\gamma),$$

where K is the posterior-separable cost function in the representation of **D**. We will use K_i to refer to $K(\mu_i, .)$.

On the basis of Remark 2, the proof strategy involves showing that Locally Invariant Posteriors implies two things. First, T_2 is a strictly positive affine transform of T_1 for all γ that are optimal in some decision problem from both both μ_1 and μ_2. Second the set of never optimal posteriors is the same for both priors, meaning that we can without loss of generality set the cost of these to be infinite according to both T_1 and T_2. Remark 1 then implies that, for any $Q \in \mathcal{Q}(\mu_2)$

$$K(\mu_2, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma)T_2(\gamma) = \sum_{\gamma \in \text{supp } Q} Q(\gamma)T_1(\gamma) - T_1(\mu_2),$$

completing the proof.

Both parts of the proof rely on the construction of $1 \leq k \leq J = |\Omega|$ interior "basis" posteriors that allow us to construct distributions $\bar{Q}_1$ and $\bar{Q}_2$ over them that generate both μ_1 and μ_2. To do this, we weight together the unit posteriors e_k with 1 in position k and zeroes elsewhere with the uniform posterior $\bar{e}$ that puts probability $1/J$ on each state, to arrive at a set of interior posteriors $\bar{\gamma}_k \in \Delta\bar{\Omega}$ that span (in the linear algebra sense) the set $\Delta\bar{\Omega}$ 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 $\Delta\bar{\Omega}$. Technically, we find $\delta \in (0,1)$ such that, when we define the corresponding posteriors $\bar{\gamma}_k^{\delta}$ for $1 \leq k \leq J$,

$$\bar{\gamma}_k^{\delta}(j) = \begin{cases} \frac{\delta}{J} + (1-\delta) \text{ if } k = j; \\ \frac{\delta}{J} \text{ if } k \neq j \ . \end{cases} \tag{2.1}$$

the unique probability weights $\bar{Q}_i(k)$ for $i = 1, 2$ and $1 \leq k \leq J$ 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, $\bar{Q}_i(\bar{\gamma}_k^{\delta}) > 0$ for $i = 1, 2$ and $1 \leq k \leq J$. This can be done by setting δ such that $0 < \frac{\delta}{J} < \min_{j=1..J} \{\min\{\mu_1(j), \mu_2(j)\}\}$, where the latter term is greater than 0 by construction. Going forward we suppress the δ parameter and set $\bar{\gamma}_k^{\delta} = \bar{\gamma}_k$.

Note that, as these basis vectors are interior, by Corollary 1.2 they must be optimal in some decision problem for both μ_1 and μ_2 - i.e. $\bar{\gamma}_k \in$int$\Delta\bar{\Omega}$ implies $\bar{\gamma}_k \in \hat{\Gamma}(\mu_1|K)$ and $\bar{\gamma}_k \in \hat{\Gamma}(\mu_2|K)$

We focus on prior μ_1 and invoke from Feasibility Implies Optimality (Lemma 1,2), the 1-1 function $f : \hat{\Gamma}(\mu_1|K) \to \mathcal{A}$ with $f(\bar{\gamma}_k) \equiv \bar{f}_k$ for $1 \leq k \leq J$ such that, when we substitute into the corresponding net utility function,

$$N_1^{\bar{f}_k}(\gamma') \leq N_1^{\bar{f}_k}(\bar{\gamma}_k) = 0,$$

all $\gamma' \in \Delta\bar{\Omega}$. Defining $\bar{A} = \cup_k f(\bar{\gamma}_k)$, Feasibility Implies Optimality then implies that the strategy $(\bar{Q}_1, \bar{q}_1) \in \Lambda(\mu_1, \bar{A}|K_1)$ having posteriors supp $\bar{Q}_1 = \cup_{k=1}^{J} \bar{\gamma}_k$, placing probability weights on them according to $\bar{Q}_1$, and involving deterministic choice at each possible posterior of the corresponding action,

$$\begin{aligned} \bar{Q}_1(\bar{\gamma}_k) &= \bar{Q}_1(k) > 0; \\ \bar{q}_1(\bar{f}_k|\bar{\gamma}_k) &= 1. \end{aligned}$$

is an optimal strategy,

$$(\bar{Q}_1, \bar{q}_1) \in \hat{\Lambda}(\mu_1, \bar{A}|K_1)$$

Note that since K_1 represents the data and $(\bar{Q}_1, \bar{q}_1) \in \hat{\Lambda}(\mu_1, \bar{A}|K_1)$, the corresponding SDSC data satisfies $P_{(\bar{Q}_1, \bar{q}_1)} \in \mathbf{D}(\mu_1, \bar{A})$. Let $(\mathbf{Q}_1, \mathbf{q}_1)$ be the corresponding revealed strategy. Note that the one-to-one nature of the function f means that each action in $\bar{A}$ is chosen from exactly one posterior, meaning that the set of revealed posteriors is the same as those used in the strategy - i.e. supp $\mathbf{Q}_1$ =supp $\bar{Q}_1$ - and $\bar{q}_1$ is deterministic - i.e. $\mathbf{q}_1(f(\bar{\gamma}_k)|\bar{\gamma}_k) = 1 \ \forall \ \bar{\gamma}_k \in$int$\Delta\bar{\Omega}$.

We now consider the strategy $(\bar{Q}_2, \bar{q}_2)$ which has the same possible posteriors, the same deterministic choice at each possible posterior, yet places probability weights on them according to $\bar{Q}_2$,

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

Since supp $\bar{Q}_2 \subset$supp $\mathbf{Q}_1$, supp μ_1 =supp μ_2, and $\bar{q}_2$ has the same deterministic choice strategy as $\mathbf{q}_1$, Locally Invariant Posteriors implies that $P_{(\bar{Q}_2, \bar{q}_2)} \in \mathbf{D}(\mu_2, \bar{A})$.

It is straightforward to show that $(\bar{Q}_2, \bar{q}_2) \in \hat{\Lambda}(\mu_2, \bar{A}|K_2)$. Note first that since **D** has a posterior-separable representation, there is an optimal strategy so that $\hat{\Lambda}(\mu_2, \bar{A}|K_2)$ is non-empty. Moreover, that optimal strategy must generate the data $P_{(\bar{Q}_2, \bar{q}_2)}$. What is unique about $(\bar{Q}_2, \bar{q}_2)$ is that each action is chosen at only one posterior. Any alternative strategy to $(\bar{Q}_2, \bar{q}_2)$ that produces the same data must have at least one action chosen at strictly more than one posterior. By strict convexity of T and linearity of EU, one can strictly improve upon this by using the strategy that coalesces choice onto the average of these. Hence indeed $(\bar{Q}_2, \bar{q}_2) \in \hat{\Lambda}(\mu_2, \bar{A}|K_2)$.

The Lagrangian Lemma then ensures there are multipliers $\theta \in \mathbb{R}^{J-1}$ s.t. for $1 \leq k, l \leq J$

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

or,

$$\sum_{j=1}^{J} \bar{f}_k(j)\bar{\gamma}_k(j) - T_2(\bar{\gamma}_k) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_k(j) = \sum_{j=1}^{J} \bar{f}_l(j) - T_2(\bar{\gamma}_l) - \sum_{j=1}^{J-1} \theta(j)\bar{\gamma}_l(j). \tag{2.2}$$

By construction of the function f we know that $N_1^{\bar{f}_k}(\bar{\gamma}_k) = N_1^{\bar{f}_l}(\bar{\gamma}_l) = 0$ and so,

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

Substitution in (2.2) 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 $1 \leq k, l \leq J$,

$$\sum_{j=1}^{J-1} \theta(j) \left[\bar{\gamma}_k(j) - \bar{\gamma}_l(j)\right] = T_1(\bar{\gamma}_k) - T_2(\bar{\gamma}_k) - T_1(\bar{\gamma}_l) + T_2(\bar{\gamma}_l). \tag{2.3}$$

We show now that these equations have unique solutions. 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 δ as defined in (2.1),

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

The next key claim is that with Lemma 1.2 the equation above applies not only to the spanning posteriors but to all pairs of posteriors, $\gamma, \gamma' \in \hat{\Gamma}(\mu_1|K)$. To see this, set $\gamma = \bar{\gamma}_{J+1}$ and $\gamma' = \bar{\gamma}_{J+2}$ and repeat the above argument for the larger set of posteriors $\cup_{k=1}^{J+2}\bar{\gamma}_k$ and the corresponding actions defined by $f : \hat{\Gamma}(\mu_1|K) \to \mathcal{A}$ as identified in Lemma 1.2,

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

We now identify 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 $\Delta\bar{\Omega}$, 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 possible 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 that uniquely pin down $\theta(j)$ also characterize

$\eta(j)$. This implies that, for any $\gamma, \gamma' \in \hat{\Gamma}(\mu_1|K)$,

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

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 that, on the subdomain $\hat{\Gamma}(\mu_1|K)$, T_2 can be obtained by adding an affine function of γ to $T_1 : \hat{\Gamma}(\mu_1|K) \to \bar{\mathbb{R}}$ so that we can define,

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

on the domain of posteriors that are ever optimal from μ_1.

The final observation is that the above argument ensures that any posteriors that are ever optimal at μ_1 must also be be chosen, and therefore optimal, in some decision problem for μ_2 - i.e. $\hat{\Gamma}(\mu_1|K) \subset \hat{\Gamma}(\mu_2|K)$. However, the argument can also be reversed to ensure that $\hat{\Gamma}(\mu_1|K) \supset \hat{\Gamma}(\mu_2|K)$, and so the two sets are equal. Thus, we can set $T_1(\gamma) = T_2(\gamma) = \infty$ for any $\gamma \in \Delta\bar{\Omega}/\hat{\Gamma}(\mu_1|K)$. Remark 2 can then be applied to complete the proof. ■

2.2 Theorem 1: Necessity

Theorem 1 (Necessity) A data set $\mathbf{D}$ with a posterior-separable representation has a uniformly posterior-separable representation only if it satisfies Locally Invariant Posteriors (Axiom 1).

Proof. We first note that, an obvious corollary of the Lagrangian Lemma is that, if a cost function is uniformly posterior-separable, for any decision problem (μ, A), $(Q, q) \in \hat{\Lambda}(\mu, A|K)$ if and only if $\exists \theta \in \mathbb{R}^{J-1}$ such that, given $\gamma \in$supp Q and $a \in A$ with $q(a|\gamma) > 0$,

$$N^{a'}_{\text{supp } \mu}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \leq N^{a}_{\text{supp } \mu}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j), \tag{2.4}$$

all $\gamma' \in \Delta(\text{supp } \mu)$ and $a' \in A$, where

$$N^{a}_{\text{supp } \mu}(\gamma) \equiv u(\gamma, a) - T_{\text{supp } \mu}(\gamma).$$

Now consider any decision problem (μ, A) and state dependent stochastic choice data $\mathbf{P}_{(\mu,A)} \in \mathbf{D}_{(\mu,A)}$ with deterministic revealed strategy $(Q, q) = (\mathbf{Q}_{\mathbf{P}_{(\mu,A)}}, \mathbf{q}_{\mathbf{P}_{(\mu,A)}})$. Note that since this is a posterior-separable representation and $\mathbf{P}_{(\mu,A)} \in \mathbf{D}_{(\mu,A)}$ the corresponding revealed strategy is optimal, since any other strategy that produces this data involves picking the same action at two different posteriors, hence cannot be optimal. Hence,

$$(Q, q) \in \hat{\Lambda}(\mu, A|K).$$

Let $\theta \in \mathbb{R}^{J-1}$ be the multipliers that satisfy the inequalities defined in (2.4) above. Now consider (Q', q') with supp $Q' \subset$supp Q and supp $\mu' =$ supp μ, and the same deterministic choice strategy as $\mathbf{q}_{\mathbf{P}_{(\mu,A)}}$ at all $\gamma \in$supp Q'. Noting that supp μ' = supp μ implies that $N^a_{\text{supp }\mu}(\gamma) = N^a_{\text{supp }\mu'}(\gamma)$ for all $\gamma \in \Delta(\text{supp }\mu) = \Delta(\text{supp }\mu')$, these conditions imply that, given $\gamma \in$supp Q' and $a \in A$ with $q'(a|\gamma) > 0$,

$$N^{a'}_{\text{supp }\mu'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j) \leq N^a_{\text{supp }\mu'}(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j)$$

all $\gamma' \in \Delta(\text{supp }\mu)$ and $a' \in A$, guaranteeing that $(Q', q') \in \hat{\Lambda}(\mu', A|K)$

Since the data has a posterior-separable representation, we know therefore that $P_{(Q',q')} \in \mathbf{D}_{(\mu',A)}$. Moreover, as we know that (Q', q') implies picking distinct actions at each posterior $\mathbf{P}_{(Q',q')} = P_{(Q',q')}$, completing the proof. ■

3 Proof of Theorem 2

Theorem 2: A data set **D** with a posterior-separable representation has an invariant posterior-separable representation if and only if it satisfies Invariance under Compression (Axiom 2).

We begin by defining terms. We then prove sufficiency and necessity, in turn. We close the section by showing that Invariance under Compression, or equivalently an invariant posterior-separable cost function, implies that the cost function K is symmetric.

3.1 Definitions

3.1.1 Invariant Cost Functions

Consider a fixed state space $\bar{\Omega} \subset \Omega$ of finite dimension, and consider a partition of $\bar{\Omega}$ into disjoint events $\{\bar{\Omega}_z\}_{z=1,...Z}$ such that $\bar{\Omega}_i \cap \bar{\Omega}_j = \varnothing$ and $\cup_i \bar{\Omega}_i = \bar{\Omega}$. Consider μ such that $\mu(\bar{\Omega}_z) > 0$ for all $\bar{\Omega}_z$, and $Q \in \mathcal{Q}(\mu)$. Define the function $\nu_{\bar{\Omega}_z,\mu} : \mathcal{Q}(\mu) \to \mathcal{Q}(\mu)$ as follows. For each $\gamma \in$supp Q, construct γ' such that γ' assigns the same probability to each subset $\bar{\Omega}_z$,

$$\gamma'(\bar{\Omega}_z) \equiv \sum_{\omega \in \bar{\Omega}_z} \gamma'(\omega) = \sum_{\omega \in \bar{\Omega}_z} \gamma(\omega) \equiv \gamma(\bar{\Omega}_z), \tag{3.1}$$

and within each subset $\bar{\Omega}_z$, the conditional probability of each state is equal to that of the prior,

$$\gamma'(\omega|\bar{\Omega}_z) = \mu(\omega|\bar{\Omega}_z) \tag{3.2}$$

Finally let

$$Q'(\gamma') = Q(\gamma) \tag{3.3}$$

where γ is the posterior in Q used in the construction of γ'. Then $Q' = \nu_{\bar{\Omega}_z,\mu}(Q)$. It will be useful to use the notation $\nu_{\bar{\Omega}_z,\mu}\gamma$ for γ' constructed as above.

Definition 4 *A cost function K is **invariant** if for all finite sets of states $\bar{\Omega} \subset \Omega$, all partitions of $\bar{\Omega}$, all pairs of priors μ and μ' that place equal probability on each partition subset, and all feasible strategies $Q \in \mathcal{Q}(\mu)$:*

$$K(\mu, Q) \geq K(\mu, \nu_{\bar{\Omega}_z,\mu}(Q))$$

and

$$K(\mu, \nu_{\bar{\Omega}_z,\mu}(Q)) = K(\bar{\mu}, \nu_{\bar{\Omega}_z,\bar{\mu}}(Q)).$$

A cost function is ***invariant posterior-separable*** *if it is both invariant and posterior-separable.*

3.1.2 Basic Forms

Our second theorem states that a single invariance axiom takes us from a posterior-separable representation to an invariant posterior-separable representation. This axiom insists that choices not change when payoff equivalent states are "compressed" into a single state. In a sense, the payoffs are what define the states that are relevant for choice. Having multiple states with similar payoffs has no effect on the propensity for any action to be chosen. We first define compression and then present out axiom.

Consider a set of actions A, and label these actions $(a_1, \ldots a_N)$. Let $A(\omega) = (a_1(\omega), \ldots a_N(\omega))$ denote the vector of payoffs in state ω. We say that a decision problem $(\bar{\mu}, A)$ is a **reduction of** (μ, A) if (1) the support of $\bar{\mu}$ is contained in the support of μ, supp $\bar{\mu}$ $\subset$supp μ, and (2) the probability of each payoff profile is the same under both measures, $\bar{\mu}\{\omega|A(\omega) = f\} = \mu\{\omega|A(\omega) = f\}$ for all observed payoff vectors f. The idea of a reduction is that we have reduced the number of states with strictly positive probability (point 1) without altering the frequency with which any vector of outcomes is observed (point 2). We say that a decision problem (μ, A) is **basic** if there exists no decision problem $(\bar{\mu}, A)$ that is a reduction of (μ, A). Intuitively, basic decision problems are those in which no two states have the same profile of payoffs. Let $\mathcal{B}(\mu, A)$ denote the set of basic decision problems that are reductions of (μ, A). We will call elements of $\mathcal{B}(\mu, A)$ **basic forms** of (μ, A).

The following characterization of basic forms is useful in the proof

Lemma 3.1 Given (μ, A), $(\bar{\mu}, A) \in \mathcal{B}(\mu, A)$ if and only if:

1. There exists a partition $\{\Omega_z\}_{1 \le l \le L}$ of supp $\mu \in \Omega$ such that given $\omega \in \Omega_l$ and $\omega' \in \Omega_m$,
$$l = m \text{ iff } a(\omega) = a(\omega') \text{ for all } a \in A.$$
2. There exists a selection $\xi : \{\Omega_z\}_{1 \le z \le Z} \to \Omega$ such that $\xi(\Omega_z) \in \Omega_z$ such that
$$\bar{\mu}(\omega) = \begin{cases} \sum_{\omega' \in \Omega_z} \mu(\omega') & \text{if } \omega = \xi(\Omega_z) \\ 0 & \text{otherwise} \end{cases}$$

Proof. Immediate. ■

Given a decision problem (μ, A), Point 1 states that the payoff equivalent, or redundant, states define a partition $\{\Omega_z\}_{1 \le z \le Z}$ of the support of μ. Point 2 defines a mapping ξ which selects a state from each subset Ω_z. All of the prior probability of Ω_z is assigned to this state.

3.1.3 Invariance under Compression

Our Invariance under Compression axiom insists that patterns of choice are equivalent in all decision problems with a common basic form. Given a decision problem (μ, A) and a basic form $(\bar{\mu}, A) \in \mathcal{B}(\mu, A)$, we can define a reduction mapping ξ :supp $\mu \rightarrow$supp $\bar{\mu}$ which assigns to each state in the support of μ the unique state in of the support of $\bar{\mu}$ with the same payoff profile: $A(\omega) = f$ implies $A(\xi(\omega)) = f$. Invariance under Compression states that the observed frequency of action a in state ω in the decision problem (μ, A) is the same as the observed frequency of action a in state $\xi(\omega)$ in the basic form $(\bar{\mu}, A)$.

Axiom 2 Invariance under Compression: *Given any* (μ, A) and $(\bar{\mu}, A)$ such that $(\bar{\mu}, A) \in \mathcal{B}(\mu, A)$, *then* $\mathbf{P}_{(\mu,A)} \in \mathbf{D}_{(\mu,A)}$ if and only if there exists $\bar{\mathbf{P}}_{(\bar{\mu},A)} \in \mathbf{D}_{(\bar{\mu},A)}$ s.t. $\mathbf{P}_{(\mu,A)}(a|\omega) = \bar{\mathbf{P}}_{(\bar{\mu},A)}(a|\xi(\omega))$, *for all* $\omega \in$*supp* μ.

While we state the axiom in terms of a decision problem and its basic form, an immediate implication is that behavior is similar in any two problems that share a basic form. Given (μ_1, A) and (μ_2, A) with a common basic form $(\bar{\mu}, A)$ and the associated reduction mappings ξ_1 and ξ_2, $\mathbf{P}_{(\mu_1,A)}(a|\omega) = \mathbf{P}_{(\mu_2,A)}(a|\omega')$ whenever $\xi_1(\omega) = \xi_2(\omega')$. The axiom therefore relates all problems with comparable payoff profiles.

3.2 Theorem 2: Sufficiency

Theorem 2 (Sufficiency): A data set $\mathbf{D}$ with a posterior-separable representation has an invariant posterior-separable representation if it satisfies Invariance under Compression (Axiom 2).

Proof. Suppose that $\mathbf{D}$ has a posterior-separable representation, then $\mathbf{D}$ has representation with a cost function

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma) T_\mu(\gamma)$$

where T_μ is a strictly convex function. Suppose further that $\mathbf{D}$ satisfies Invariance under Compression, we show that the cost function K must be invariant.

Fix an arbitrary set of states $\bar{\Omega}$ and an consider an arbitrary partition of $\bar{\Omega}$, $\{\bar{\Omega}_z\}_{1 \leq z \leq Z}$. Consider two priors μ and μ' that place equal probability on each partition subset, and a feasible strategy $Q \in \mathcal{Q}(\mu)$. We wish to show that

$$K(\mu, Q) \geq K(\mu, \nu_{\bar{\Omega}_z,\mu}(Q)) \tag{3.4}$$

and

$$K(\mu, \nu_{\bar{\Omega}_z,\mu}(Q)) = K(\bar{\mu}, \nu_{\bar{\Omega}_z,\mu'}(Q)). \tag{3.5}$$

We begin assuming that all $\gamma \in$supp Q lie in the interior of $\Delta\bar{\Omega}$. We first consider (3.4). Consider a selection $\xi : \{\Omega_z\}_{1 \leq l \leq L} \rightarrow \bar{\Omega}$. Let $\bar{\Omega}_\xi$ denote the image of ξ, and consider the probability density $\bar{\mu} \in \Delta\bar{\Omega}_\xi$ such that

$$\bar{\mu}(\omega_l) = \begin{cases} \sum_{\omega \in \Omega_z} \mu(\omega) & \text{if } \omega_l = \xi(\Omega_z) \\ 0 & \text{otherwise} \end{cases}$$

Consider $\bar{Q} = \nu_{\bar{\Omega}_z,\bar{\mu}}(Q)$ constructed by defining for each $\gamma \in Q$ posterior $\bar{\gamma}(\gamma)$ such that

$$\bar{\gamma}(\omega_l) = \begin{cases} \sum_{\omega \in \Omega_z} \mu(\omega) & \text{if } \omega_l = \xi(\Omega_z) \\ 0 & \text{otherwise} \end{cases}$$

and setting $\bar{Q}(\bar{\gamma}(\gamma)) = Q(\gamma)$. Note that $Q \in \mathcal{Q}(\mu)$ implies $\bar{Q} \in \mathcal{Q}(\bar{\mu})$.

Since each $\gamma \in$supp Q lies in the interior of $\Delta\bar{\Omega}$, each $\bar{\gamma} \in$supp $\bar{Q}$ lies in the interior of Δsupp $\bar{\mu}$. Corollary 1.2 then implies that the $\bar{\gamma} \in \hat{\Gamma}(\bar{\mu}|K)$. Feasibility Implies Optimality (Lemma 1.2) then states that there exists a decision problem $(\bar{\mu}, A)$ such that $(\bar{Q}, \bar{q})$ is the optimal policy (where $\bar{q}$ associates each each posterior in $\bar{Q}$ with a single action in A and all actions in A are associated with some posterior). Also since only payoffs on supp $\bar{\mu}$ are relevant for the decision problem $(\bar{\mu}, A)$, we can choose A such that the payoffs on all states $\omega \in \Omega_z$ are identical to the payoff on $\xi(\Omega_z)$. By Lemma 3.1, $(\bar{\mu}, A)$ is a basic form for (μ, A).

By Invariance under Compression, $P_{(\bar{Q},\bar{q})} \in \mathbf{D}$ implies that there exists $\mathbf{P}_{(\mu,A)} \in \mathbf{D}$ such that

$$\mathbf{P}_{(\mu,A)}(a|\omega) = P_{(\bar{Q},\bar{q})}(a|\xi(\omega))$$

for all $a \in A$. Let $(\mathbf{Q}, \mathbf{q})$ denote the revealed policy associated with $\mathbf{P}_{(\mu,A)}$. Since K represents $\mathbf{D}$, $(\mathbf{Q}, \mathbf{q}) \in \hat{\Lambda}(\mu, A|K)$. By Bayes rule, the posterior associated with the choice of $a \in A$ is

$$\boldsymbol{\gamma}^a(\omega) = \frac{\mathbf{P}_{(\mu,A)}(a|\omega)\mu(\omega)}{\mathbf{P}(a)}$$

It follows that

$$\boldsymbol{\gamma}^a(\omega) = \bar{\gamma}^a(\xi(\omega))\frac{\mu(\omega)}{\bar{\mu}(\xi(\omega))}$$

were $\bar{\gamma}^a$ is the posterior associated with action a in the problem $(\bar{\mu}, A)$. It follows immediately that $\mathbf{Q} = \nu_{\bar{\Omega}_z,\mu}(Q)$. Hence $(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q}) \in \hat{\Lambda}(\mu, A|K)$.

Since $(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q})$ is optimal for (μ, A),

$$V(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q}|\mu, A, K) \geq V(Q, q|\mu, A, K)$$

where (Q, q) was the strategy that we started with. Note that the expected utility to strategies $(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q})$ and (Q, q) are identical as the posteriors only differ on redundant states. It follows that

$$K(\mu, Q) \geq K(\mu, \nu_{\bar{\Omega}_z,\mu}(Q))$$

which establishes (3.4).

To establish (3.5), consider also an action a' that pays off $V(\bar{Q}, \bar{q}|\bar{\mu}, A, K)$ in all states of the world, so that strategy of not learning, $(Q_{\bar{\mu}}, q_{\bar{\mu}})$ with $Q_{\bar{\mu}}(\bar{\mu}) = 1$ and $q_{\bar{\mu}}(a'|\bar{\mu}) = 1$, yields the same payoff as $(\bar{Q}, \bar{q})$. Both $(\bar{Q}, \bar{q})$ and $(Q_{\bar{\mu}}, q_{\bar{\mu}})$ are optimal in the augmented problem $(\bar{\mu}, A \cup a')$.

Now consider the problem $(\mu, A \cup a')$. By Invariance under Compression, the strategies $(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q})$ and $(\nu_{\bar{\Omega}_z,\mu}(Q_{\bar{\mu}}), q'_{\bar{\mu}})$ where $q'_{\bar{\mu}}(a'|\mu) = 1$ are both optimal for this problem. Note that by construction $(\nu_{\bar{\Omega}_z,\mu}(Q_{\bar{\mu}}), q'_{\bar{\mu}})$ pays $V(\bar{Q}, \bar{q}|\bar{\mu}, A, K)$ in all states of the world and has no information cost. Moreover, by construction the expected utility to $(\nu_{\bar{\Omega}_z,\mu}(Q), \mathbf{q})$ is identical to the expected utility of $(\bar{Q},\bar{q})$. It follows that,

$$K(\bar{\mu},\bar{Q}) = K(\mu,\nu_{\bar{\Omega}_z,\mu}(Q))$$

As μ is an arbitrary prior that places positive probability on each subset $\bar{\Omega}_z$, this establishes (3.5).

We now consider the case in which some $\gamma \in$supp Q, lies on the boundary of $\Delta\bar{\Omega}$. By Lemma 1.3 we may take K to be continuous on the boundary. (3.4) and (3.5) therefore hold on the boundary as well. ■

3.3 Theorem 2: Necessity

Theorem 2 (Necessity) A data set **D** with a posterior-separable representation has an invariant posterior-separable representation only if it satisfies Invariance under Compression (Axiom 2).

Proof. Suppose **D** has a representation with a cost function K that is posterior separable with strictly convex function T_μ. Suppose that K is invariant. We show that **D** satisfies Axiom 2.

Consider an arbitrary decision problem (μ, A) with redundant states, and consider a basic form $(\bar{\mu}, A) \in \mathcal{B}(\mu, A)$. Let $\{\Omega_z\}$ denote the partition of supp μ into payoff equivalent states and let $\xi : \{\Omega_z\} \to$supp μ denote the selection associated with the basic form $(\bar{\mu}, A)$.

Consider an arbitrary optimal strategy $(Q, q) \in \hat{\Lambda}(\mu, A)$. Consider the policy $(\nu_{\bar{\Omega}_z,\mu}(Q), q')$ where q' is defined in the natural way as $q'(a|\gamma') = q(a|\gamma)$ where $\gamma' = \nu_{\bar{\Omega}_z,\mu}\gamma$. Suppose that $(Q, q) \neq (\nu_{\bar{\Omega}_z,\mu}(Q), q')$. We will derive a contradiction.

Note that since γ and γ' only differ on redundant states, the expected utility of (Q, q) and (Q', q') are equal. Since K is invariant, and hence monotonic,

$$K(\mu, Q') \leq K(\mu, Q).$$

Since $(Q, q) \in \hat{\Lambda}(\mu, A)$, we must have $K(\mu, Q') = K(\mu, Q)$ and (Q', q') is optimal as well.

Since $Q \neq Q'$, consider the policy (Q'', q'') which is the convex combination of (Q, q) and (Q', q') formed as follows. Given $\gamma \in$supp Q and $\alpha \in (0, 1)$, there exists $\gamma'' \in$supp Q'' such that $\gamma'' = \alpha\gamma + (1 - \alpha)\nu_{\bar{\Omega}_z,\mu}\gamma$. Set (Q'', q'') such that for each $\gamma \in$supp Q, $Q''(\gamma'') = Q(\gamma)$ and $q''(\gamma'') = q(\gamma)$. Since T_μ is strictly convex in γ, $K(\mu, Q'') < K(\mu, Q') = K(\mu, Q)$. The expected payoff to (Q'', q'') is the same as (Q, q) and (Q', q') and the cost is strictly lower. Hence (Q'', q'') dominates both (Q, q) and (Q', q') contradicting the assumption that (Q, q) was optimal. It follows that $(\nu_{\bar{\Omega}_z,\mu}(Q), q')$ is the optimal policy.

Bayes rule then implies that state dependent stochastic choice the same redundant states. Given $\omega, \omega' \in \Omega_z$,

$$P(a|\omega) = \frac{P(a)\gamma(\omega)}{\mu(\omega)} = \frac{P(a)\gamma(\omega')}{\mu(\omega')} = P(a|\omega')$$

where the second inequality follows from $Q = Q'$. It follows that the data set satisfies Invariance under Compression. ■

3.4 Symmetry

Our notion of symmetry is based on bijections between sets of states, actions and posteriors. We say that two probability distributions $\gamma_1, \gamma_2 \in \Delta\Omega$ are symmetric, if there exists a bijection σ_Ω :supp γ_1 →supp γ_2 such that, for all $\omega \in$supp γ_1,

$$\gamma_1(\omega) = \gamma_2(\sigma_\Omega(\omega)).$$

Correspondingly, we say that two decision problems (μ_1, A_1) and (μ_2, A_2) are symmetric if there exists a bijection σ_ω :supp μ_1 →supp μ_2 such that, for all $\omega \in$supp μ_1,

$$\mu_1(\omega) = \mu_2(\sigma_\Omega(\omega))$$

and a bijection $\sigma_A : A_1 \to A_2$ such that for $b = \sigma_A(a)$

$$a(\omega) = b(\sigma_\Omega(\omega)).$$

We say (μ_1, Q_1) and (μ_2, Q_2) are symmetric where μ_1 and μ_2 are priors, $Q_1 \in \mathcal{Q}(\mu_1)$ and $Q_2 \in \mathcal{Q}(\mu_2)$, if there exists a bijection σ_Ω :supp μ_1 →supp μ_2 such that, for all $\omega \in$supp μ_1,

$$\mu_1(\omega) = \mu_2(\sigma_\Omega(\omega))$$

and a bijection σ_Q :supp Q_1 →supp Q_2 such that for $\hat{\gamma} = \sigma_Q(\gamma)$

$$Q_1(\gamma) = Q_2(\hat{\gamma})$$

and

$$\gamma(\omega) = \hat{\gamma}(\sigma(\omega))$$

for all $\omega \in$supp μ_1.

Finally, we say that a cost function $K(\mu, Q)$ is symmetric if given two symmetric learning strategies (μ_1, Q_1) and (μ_2, Q_2) imply equal costs, $K(\mu_1, Q_1) = K(\mu_2, Q_2)$.

Lemma 3.2 (Symmetric Costs) If **D** has a posterior-separable representation K and satisfies Invariance under Compression, then K is symmetric.

Proof. By Theorem 2, If **D** has a posterior-separable representation K and satisfies Invariance under Compression then K is invariant.

Consider symmetric (μ_1, Q_1) and (μ_2, Q_2). By assumption there exists a bijection σ_Ω :supp μ_1 →supp μ_2 such that, for all $\omega \in$supp μ_1,

$$\mu_1(\omega) = \mu_2(\sigma_\Omega(\omega))$$

and a bijection σ_Q :supp Q_1 →supp Q_2 such that for $\hat{\gamma} = \sigma_Q(\gamma)$

$$Q_1(\gamma) = Q_2(\hat{\gamma})$$

and

$$\gamma(\omega) = \hat{\gamma}(\sigma(\omega)).$$

Suppose for the moment that supp μ_1 and supp μ_3 are disjoint. Consider now a third learning strategy (μ_3, Q_3) on supp $\mu_1 \cup$supp μ_2 such that

$$\mu_3(\omega) = \frac{1}{2}\mu_1(\omega) + \frac{1}{2}\mu_2(\omega)$$

and for all $\gamma \in$supp Q_1, there exists $\gamma^{'} \in$supp Q_3

$$Q_3(\gamma^{'}) = Q_1(\gamma)$$

and

$$\gamma^{'}(\omega) = \frac{1}{2}\gamma(\omega) + \frac{1}{2}\hat{\gamma}(\omega)$$

where $\hat{\gamma} = \sigma_Q(\gamma)$. Let $\gamma' = \sigma_3(\gamma)$. Consider also the partition that groups each $\omega \in$supp μ_1 with its image under σ_Ω, $\Omega_z = \{\omega, \sigma_\Omega(\omega)\}$. Note that $(\mu_1, Q_1) = (\mu_1, \nu_{\bar{\Omega}_z,\mu_1}(Q_3)$ and $(\mu_2, Q_2) = (\mu_2, \nu_{\bar{\Omega}_z,\mu_2}(Q_3))$. Invariance implies that

$$K(\mu_1, \nu_{\bar{\Omega}_z,\mu_1}(Q_3) = K(\mu_2, \nu_{\bar{\Omega}_z,\mu_2}(Q_3))$$

It follows that $K(\mu_1, Q_1) = K(\mu_2, Q_2)$ for any given symmetric information strategies (μ_1, Q_1) and (μ_2, Q_2) over disjoint states.

Now if supp μ_1 and supp μ_3 are not disjoint, we can consider a learning strategy (μ_4, Q_4) which is symmetric to both (μ_1, Q_1) and (μ_2, Q_2) such that supp $\mu_4 \cap \{$supp $\mu_1 \cup$supp $\mu_2\} = \varnothing$. By the above, it follows that

$$K(\mu_1, Q_1) = K(\mu_4, Q_4) = K(\mu_2, Q_2)$$

Hence K is symmetric. ∎

4 Proof of Theorem 3

In this section we prove that Axioms 1 and 2 are necessary and sufficient for a Shannon representation. We know from Theorems 1 and 2 that Axiom 1 is equivalent to the cost function being uniformly posterior-separable, and that Axiom 2 is equivalent to the cost function being invariant posterior-separable. Proving that these qualities characterize the Shannon cost function is equivalent to proving that Axioms 1 and 2 characterize the Shannon cost function.

Theorem 3: *The Shannon cost function is unique in that it is invariant and uniformly posterior-separable.*

4.1 Theorem 3: Necessity

Theorem 3 (Necessity): *The Shannon cost function is invariant and uniformly posterior-separable.*

Necessity is immediate. The Shannon cost function is

$$K^S_\kappa(\mu, Q) \equiv \kappa \left[\sum_{\gamma \in \text{supp } Q} Q(\gamma) \sum_{\omega \in \text{supp } \gamma} \gamma(\omega) \ln \gamma(\omega) - \sum_{\omega \in \text{supp } \mu} \mu(\omega) \ln \mu(\omega) \right]. \quad (4.1)$$

This is clearly uniformly posterior separable. The Kullback-Leibler divergence is invariant (Amari, 2016 , pp. 54-57). It follows that the Shannon cost function, as the expected Kullback-Leibler divergence, is invariant.

4.2 Theorem 3: Sufficiency

Theorem 3 (Sufficiency): *An invariant and uniformly posterior-separable cost function is the Shannon cost function.*

4.2.1 A sketch of the argument

The proof is long and builds on a series of lemmas. We use a result from the information geometry literature that, when considering probability distributions over a fixed set of states the Kullback-Leibler divergence is unique in that it is at once invariant, differentiable, and a type of function known as a Bregman divergence. There are four gaps that lie between the standard result from information geometry and the sufficiency statement in Theorem 3. We need to show that T_μ is differentiable. We need to show that T_μ rather than K is invariant. We need to write T_μ as a Bregman divergence. Finally, once we have established the form of T_μ for one set of states, we need to show that that form is constant across states. This boils down to showing that κ is constant across decision problems. We address each gap in turn. Before we show that T_μ is differentiable, however, it is useful to show that T_μ is symmetric.

We initially consider a fixed state space $\bar{\Omega} \subset \Omega$ of cardinality $J \geq 4$ and consider priors μ with support $\bar{\Omega}$. Since K is uniformly posterior separable, we can write

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma) T_\mu(\gamma)$$

with

$$T_\mu(\gamma) = T_{\bar{\Omega}}(\gamma) - T_{\bar{\Omega}}(\mu)$$

Since the state space $\bar{\Omega}$ is fixed we will simplify notation and simply write $T(\gamma)$ for $T_{\bar{\Omega}}(\gamma)$.

While K is unique, T is not. We can add a linear function to T without changing K (See remark in Appendix Section 1). Since K is invariant, Lemma 4 states that K is symmetric. We first show that there exists a choice of T that is symmetric (Lemma 4.1).

The proof that T is differentiable rests on two key insights. The first observation is that the Lagrangian Lemma (Lemma 1) implies that there is a common hyperplane tangent to each of the net utility functions at each chosen posterior. This links the subdifferentials of the net utility function at distinct optimal posteriors (Lemma 4.6). If net utility is differentiable, then the partial derivatives must be equal

$$N^a_{(ji)}(\gamma^a) = N^b_{(ji)}(\gamma^b)$$

where $N^a_{(ji)}$ is the change in the net utility of action a from raising $\gamma(\omega_i)$ and reducing $\gamma(\omega_j)$.

A second observation is that Invariance under Compression places structure on the sets of posteriors that can be linked by considering decision problems with equivalent states. In Figure A1, we illustrate this implication of Invariance under Compression with three states, but the intuition applies generally. Consider a decision problem with three states $(\omega_1, \omega_2, \omega_3)$ and two actions $A = \{a, b\}$, in which the payoffs in states ω_1 and ω_2 are equivalent. Figure A1 displays the space of potential priors and posteriors. Suppose that $\bar{\mu}_1$ is the prior in the basic problem in which all of the combined probability of ω_1 and ω_2 is assigned to ω_1 and $\bar{\mu}_2$ is the prior in the case in which ω_2 receives all of the weight. Since $\bar{\mu}_1(\omega_1) = \bar{\mu}(\omega_2)$ the line segment connecting these two priors is parallel to the segment connecting $(1, 0, 0)$ and $(0, 1, 0)$. The line segment connecting $\bar{\mu}_1$ and $\bar{\mu}_2$ represents the set of potential priors for which,

$$\mu(\omega_1) + \mu(\omega_2) = \bar{\mu}_1(\omega_1) = \bar{\mu}_2(\omega_2),$$

so that $(\bar{\mu}_1, A)$ and $(\bar{\mu}_2, A)$ are basic versions of (μ, A).

The above shows that, letting μ to be an arbitrary prior in this set, Invariance under Compression places restrictions on the relationship between the optimal posteriors for the problems (μ, A), $(\bar{\mu}_1, A)$ and $(\bar{\mu}_2, A)$ (Lemma 4.7). Consider γ^a. Bayes rule states that $\gamma^a(\omega) = P(a|\omega)\mu(\omega)/P(a)$. Invariance under Compression implies that $P(a|\omega)$ and $P(a)$ are the same for all μ on the segment connecting $\bar{\mu}_1$ and $\bar{\mu}_2$, including $\bar{\mu}_1$ and $\bar{\mu}_2$ themselves. This implies that as μ moves from $\bar{\mu}_1$ to $\bar{\mu}_2$, γ^a and γ^b are always proportionate to μ. It follows that γ^a and γ^b lie at the intersection of a line through μ and $(0, 0, 1)$, the dashed grey line in the figure, and a line parallel to the segment connecting $(1, 0, 0)$ and $(0, 1, 0)$, the solid red and blue lines in the figure. $\bar{\gamma}_1^a$ and $\bar{\gamma}_1^b$ in the figure denote the optimal posteriors for $(\bar{\mu}_1, A)$, and $\bar{\gamma}_2^a$ and $\bar{\gamma}_2^b$ the optimal posteriors for $(\bar{\mu}_2, A)$.

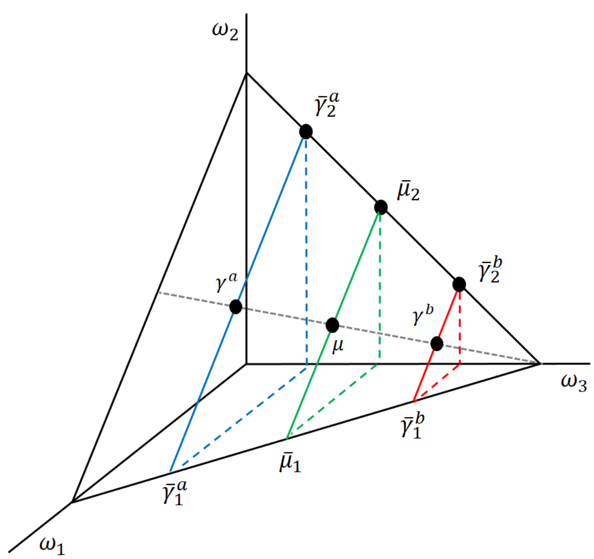

Figure A1: Implications of Compression

These two observations when combined relate the derivatives of T at γ^a and γ^b in the Figure. The Lagrangian Lemma implies that there is a hyperplane tangent to both $N(\gamma^a)$ and $N(\gamma^b)$. Suppose that the partial derivatives $T_{(ij)}(\gamma^a)$ and $T_{(ij)}(\gamma^b)$ both exist. Since prize-based expected utility is linear, the difference between $T_{(ji)}(\gamma^a)$ and $T_{(ji)}(\gamma^b)$ must equal $u(a,\omega_i) - u(a,\omega_j) - u(b,\omega_i)+u(b,\omega_j)$. Since shifts in μ from $\bar{\mu}_1$ to $\bar{\mu}_2$, do not affect prize based utility, $T_{(ji)}(\gamma^a)-T_{(ji)}(\gamma^b)$ must be independent of μ whenever both derivatives exist (Lemma 4.8).

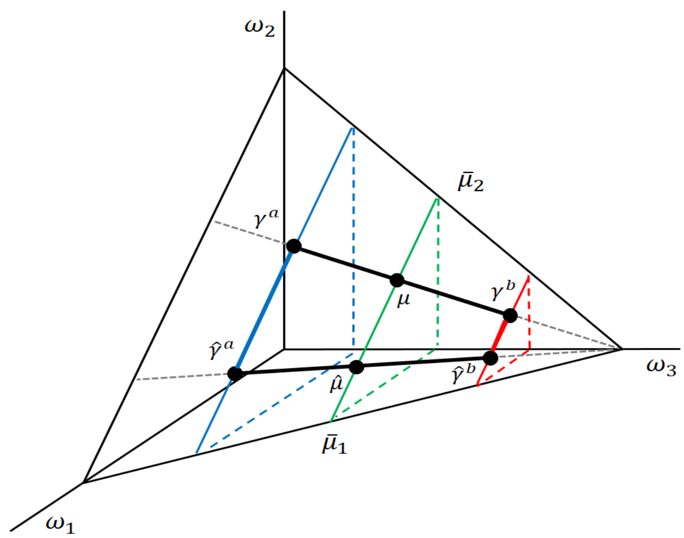

Figure A2: The Trapezoid

Consider now two priors μ and $\hat{\mu}$, each lying between $\bar{\mu}_1$ and $\bar{\mu}_2$. Figure A2 shows that the four posteriors γ^a,γ^b,$\hat{\gamma}^a$, and $\hat{\gamma}^b$ form a trapezoid. If T is differentiable at all four points we would know that:[2]

$$T_{(ji)}(\gamma^a) - T_{(ji)}(\gamma^b) = T_{(ji)}(\hat{\gamma}^a) - T_{(ji)}(\hat{\gamma}^b). \tag{4.2}$$

Equation (4.2) is close to the rectangle condition for additive separability. To apply the rectangle condition, we deform the simplex so that the trapezoid becomes a rectangle, and then return to the simplex. This results in the following characterization of the directional derivative which we state in terms of the dimension J since it requires $J \geq 4$ (Lemma 4.16):

$$T_{(ji)}(\gamma) = \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2),...,\gamma(J-1)\right), \tag{4.3}$$

for some functions $\boldsymbol{A} : \mathbb{R}_+ \longrightarrow \mathbb{R}$ and $\boldsymbol{B} : \mathbb{R}^{J-2} \longrightarrow \mathbb{R}$, and for all $2 \leq i \neq j \leq J-1$. As (4.2) must hold for a range of $\gamma(1)$ and $\gamma(J)$ we can show that $\boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right)$ must be constant (Lemma 5.17). The symmetry of T then implies that, if $T_{(ji)}(\gamma)$ does not depend on $\gamma(1)$ and $\gamma(J)$, $\boldsymbol{B}$ cannot depend on any $\gamma(k)$ other than $\gamma(i)$ and $\gamma(j)$ (Lemma 4.18). Finally, we use the fact that $T_{(ji)}(\gamma) = T_{(ki)}(\gamma) - T_{(kj)}(\gamma)$ whenever the latter are well defined to establish that there exists a function f on $(0,1)$ such that for all $\gamma \in$int$\Delta(\bar{\Omega})$ (Lemma 4.24):

$$T_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)).$$

The convexity of T implies that f is monotonic (Lemma 4.24). We complete the proof of differen-

[2]Lemma 5.12 establishes (4.2) everywhere for the directional derivatives $T_{\overrightarrow{ji}}$ by finding pairs of differentiable points that simultaneously converge to the four posteriors γ^a,γ^b,$\hat{\gamma}^a$, and $\hat{\gamma}^b$.

tiability by showing that if $T_{(ji)}(\gamma)$ does not exist, then f is discontinuous at either $\gamma(i)$ or $\gamma(j)$, but this implies that $T_{(ji)}(\hat{\gamma})$ does not exist for any $\hat{\gamma}$ with $\hat{\gamma}(i) = \gamma(i)$ or $\hat{\gamma}(j) = \gamma(j)$. But this contracts observation that T is differentiable along almost all rays such as those from ω_3 in Figure A1.

Given the differentiability of T, we employ the arguments of Hébert and La'0 [2020] to argue that T_μ is invariant. Our next task is to write T_μ as a Bregman divergence. We do so by taking limits of T_μ as μ and γ approach the boundary $\partial\Delta\bar{\Omega}$. At this point we have a differentiable, invariant, Bregman divergence. The theorem of Jiao, *et al.* [2015] implies that T_μ is the Kullback-Leibler divergence. The final step is to use invariance to equate the constant κ in the Shannon cost function across sets of states $\bar{\Omega}$.

We now turn to the details.

4.2.2 Symmetry of T

As previously stated we work with a fixed state space $\bar{\Omega}$ in sections 4.2.2 through 4.2.4. We will consider decision problem with priors μ such that supp $\mu = \bar{\Omega}$. We assume that the cardinality of $\bar{\Omega}$ is greater than or equal to four so that we can construct problems with redundant states. To conserve notation we will write T for $T_{\text{supp }\mu}$ where supp $\mu = \bar{\Omega}$. We will also write $\gamma(j)$ for $\gamma(\omega_j)$.

By assumption K has a uniform posterior-separable representation

$$K(\mu, Q) = \sum Q(\gamma)T(\gamma) - T(\mu)$$

We say that T is symmetric if given symmetric γ_1 and γ_2, $T(\gamma_1) = T(\gamma_2)$.

While the cost function K is unique, the function T is not necessarily unique. The requirement that $\sum Q(\gamma)\gamma = \mu$ implies that we can add any function that is linear in γ to T without affecting K. Since these linear functions tilt T, most T's are not symmetric. The next lemma shows that given any T we can always find a linear function $\bar{\alpha} \cdot \gamma$ so that $\hat{T}(\gamma) = T(\gamma) - \bar{\alpha} \cdot \gamma$ is symmetric.

Lemma 4.1 Given a uniformly posterior-separable cost function K, there exists a symmetric function $\hat{T}$ such that $K(\mu, Q) = \sum Q(\gamma)\hat{T}(\gamma) - \hat{T}(\mu)$.

Proof. Since $K(\mu, Q)$ is uniformly posterior separable, we can write

$$K(\mu, Q) = \sum Q(\gamma)T(\gamma) - T(\mu)$$

Suppose $J \geq 3$, and consider the J distributions $\bar{\gamma}_j = \left(\frac{1}{2(J-1)}, \frac{1}{2(J-1)} \cdots \frac{1}{2} \cdots \frac{1}{2(J-1)}\right)$ where the $1/2$ is in the jth place. Let $\bar{\mu}$ denote the uniform distribution on $\bar{\Omega}$. Consider the linear transformation of T,

$$\hat{T}(\gamma) = T(\gamma) - T(\bar{\mu}) - \bar{\alpha} \cdot (\gamma - \bar{\mu})$$

where $\bar{\alpha}$ solves

$$\hat{T}(\bar{\gamma}_j) = \bar{T}$$

Since the $\bar{\gamma}_j$ are independent such an $\bar{\alpha}$ exists. Note

$$K(\mu, Q) = \sum Q(\gamma)\hat{T}(\gamma) - \hat{T}(\mu)$$

since $\sum Q(\gamma)[\bar{\alpha} \cdot (\gamma - \bar{\mu})] = \bar{\alpha} \cdot [\sum Q(\gamma)(\gamma - \bar{\mu})] = 0$.

We now show that $\hat{T}(\gamma)$ is symmetric. Consider any γ_1 and γ_2 that are symmetric in that there exists a bijection $\sigma : \bar{\Omega} \to \bar{\Omega}$ such that $\gamma_1(\omega) = \gamma_2(\sigma(\omega))$. Consider the information strategy that places weight $Q(\gamma_1)$ on γ_1 and $(1 - Q(\gamma_1))\beta_j$ on each of the $\bar{\gamma}_j$ such that

$$\sum \beta_j \bar{\gamma}_j = \bar{\mu} - \frac{Q(\gamma_1)}{1 - Q(\gamma_1)}(\gamma_1 - \bar{\mu}) \tag{4.4}$$

Given the independence of the $\bar{\gamma}_j$, this set of β_j exists. Note that this policy lies in $\mathcal{Q}(\bar{\mu})$ by construction. Now consider the information strategy that places weight $Q(\gamma_1)$ on γ_2 and β_j on $\bar{\gamma}_{\sigma(j)}$. Given the symmetry of γ_1 and γ_2,

$$\sum \beta_j \bar{\gamma}_{\sigma(j)} = \bar{\mu} - \frac{Q(\gamma_1)}{1 - Q(\gamma_1)}(\gamma_2 - \bar{\mu})$$

and this policy is also in $\mathcal{Q}(\bar{\mu})$. The symmetry of K (Lemma 3.2) implies

$$Q(\gamma_1)\hat{T}(\gamma_1) + \sum \beta_j \hat{T}(\bar{\gamma}_j) - \hat{T}(\bar{\mu}) = Q(\gamma_1)\hat{T}(\gamma_2) + \sum \beta_j \hat{T}(\bar{\gamma}_{\sigma(j)}) - \hat{T}(\bar{\mu})$$

As the $\hat{T}(\bar{\gamma}_j) = \bar{T}$, this reduces to

$$\hat{T}(\gamma_1) = \hat{T}(\gamma_2)$$

so that $\hat{T}$ is symmetric. ■

From this point on in the proof we will assume that $T(\gamma)$ is chosen to be symmetric.

4.2.3 Differentiability of T

Our goal in this section is to prove the following lemma.

Lemma 4.2: Suppose that K is uniformly posterior separable and invariant, then $T_{\text{supp } \mu}$ is differentiable.

Basic Results The fundamental objects of interest in what follows are certain derivatives of the function T on $\bar{\Omega}$. Note that $\bar{\Omega}$ does not allow for independent variation in any single state-specific posterior $\gamma(j)$ due to the adding up constraint on probabilities. In general, given convex function $T : \Delta\bar{\Omega} \to \mathbb{R}$ and $\gamma \in \text{int}\Delta\bar{\Omega}$, the directional derivative at γ in direction $y \in \mathbb{R}^J$ is defined as,

$$\tilde{T}'(\gamma|y) = \lim_{\epsilon \downarrow 0} \frac{\tilde{T}(\gamma + \epsilon y) - \tilde{T}(\gamma)}{\epsilon}, \tag{4.5}$$

if it exists. We use special notation for the directional derivatives of interest. We are interested in derivatives in the direction of raising the probability of one state ω_i and lowering the probability of another ω_j. We define the **one-sided derivative in direction** ji, $T_{\overrightarrow{ji}}(\gamma)$, as the directional derivative associated with increasing the ith coordinate and equally reducing the jth:

$$T_{\overrightarrow{ji}}(\gamma) = \lim_{\epsilon \downarrow 0} \frac{T(\gamma + \epsilon(e_i - e_j)) - T(\gamma)}{\epsilon}; \tag{4.6}$$

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, $T_{(ji)}$, by:

$$T_{(ji)}(\gamma) = \lim_{\epsilon \to 0} \frac{T(\gamma + \epsilon(e_i - e_j)) - T(\gamma)}{\epsilon}. \tag{4.7}$$

We now translate a few standard results on derivatives of convex functions into our setting. Almost all of these results are adapted directly from Rockafellar's comprehensive treatise (Rockafellar [1970]). 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.

Lemma 4.3: The following hold:

1. T is continuous on $\gamma \in \text{int}\Delta\bar{\Omega}$.
2. Given $1 \leq i \neq j \leq J$, $T_{\overrightarrow{ji}}(\gamma)$ exists for all $\gamma \in \text{int}\Delta\bar{\Omega}$.
3. We have

$$-T_{\overrightarrow{ij}}(\gamma) \leq T_{\overrightarrow{ji}}(\gamma). \tag{4.8}$$

Proof. Continuity of T on its relative interior is Theorem 10.1 in Rockafellar [1970]. Given $\gamma \in \text{int}\Delta\bar{\Omega}$, $T_{\overrightarrow{ji}}(\gamma)$ is equal to the the one-sided directional derivative of T at γ with respect to the vector $y = e_i - e_j$ (Rockafellar [1970], p. 213). Rockafellar [1970] Theorem 23.1 establishes that, since $T : \mathbb{R}^J \to \bar{\mathbb{R}}$ is convex and $T(\gamma)$ is finite at $\gamma \in \text{int}\bar{\Omega}$, $T_{\overrightarrow{ji}}(\gamma)$ exists.

This same theorem establishes

$$-T_{\overrightarrow{ij}}(\gamma) = -T'(\gamma | e_j - e_i) \leq T'(\gamma | e_i - e_j) = T_{\overrightarrow{ji}}(\gamma).$$

■

We are particularly interested in posteriors $\gamma \in \text{int}\bar{\Omega}$ at which the inequality (4.8) 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 4.4: $T_{(ji)}(\gamma)$ exists if and only if (4.8) holds with equality,

$$-T_{\overrightarrow{ij}}(\gamma) = T_{\overrightarrow{ji}}(\gamma), \tag{4.9}$$

in which case

$$T_{(ji)}(\gamma) = T_{\overrightarrow{ji}}(\gamma) = -T_{\overrightarrow{ij}}(\gamma) = -T_{(ij)}(\gamma). \tag{4.10}$$

Moreover, if T is differentiable at $\gamma \in$int$\Delta\Omega$, $T_{(ji)}(\gamma)$ exists for all $1 \leq i \neq j \leq J$.

Proof. Note first that if (4.9) holds so that $T_{\overrightarrow{ji}}(\gamma) = -T_{\overrightarrow{ij}}(\gamma)$, this corresponds to equality of the limits from the left and right

$$\begin{aligned}\lim_{\epsilon\downarrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon} &= T_{\overrightarrow{ji}}(\gamma) = -T_{\overrightarrow{ij}}(\gamma) = -\lim_{\epsilon\downarrow 0}\frac{T(\gamma+\epsilon(e_j-e_i))-T(\gamma)}{\epsilon} \\ &= -\lim_{\delta=-\epsilon\uparrow 0}\frac{T(\gamma-\delta(e_j-e_i))-T(\gamma)}{-\delta} = \lim_{\delta\uparrow 0}\frac{T(\gamma+\delta(e_i-e_j))-T(\gamma)}{\delta}.\end{aligned}$$

It is standard that this implies that the equal left and right limits define the limit itself,

$$T_{(ji)}(\gamma) = \lim_{\epsilon\downarrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon},$$

establishing equivalence of (4.9) and existence of $T_{(ji)}(\gamma)$.

Conversely, note that if $T_{(ji)}(\gamma)$ exists,

$$T_{(ji)}(\gamma) = \lim_{\epsilon\longrightarrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon} = \lim_{\epsilon\uparrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon} = T_{\overrightarrow{ji}}(\gamma);$$

and,

$$\begin{aligned}T_{(ji)}(\gamma) &= \lim_{\epsilon\longrightarrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon} = \lim_{\epsilon\downarrow 0}\frac{T(\gamma+\epsilon(e_i-e_j))-T(\gamma)}{\epsilon} \\ &= -\lim_{\delta=-\epsilon\uparrow 0}\frac{T(\gamma+\delta(e_j-e_i))-T(\gamma)}{-\delta} = \lim_{\delta\uparrow 0}\frac{T(\gamma+\delta(e_j-e_i))-T(\gamma)}{\delta} = -T_{\overrightarrow{ij}}(\gamma).\end{aligned}$$

These equations together verify that (4.9) holds and also that,

$$T_{(ji)}(\gamma) = T_{\overrightarrow{ji}}(\gamma) = -T_{\overrightarrow{ij}}(\gamma). \tag{4.11}$$

To complete the proof that (4.10) holds, note that since $T_{\overrightarrow{ji}}(\gamma) = -T_{\overrightarrow{ij}}(\gamma)$, we know from (4.9) that $T_{(ij)}(\gamma)$ exists, and therefore that it satisfies the corresponding equality,

$$T_{(ij)}(\gamma) = T_{\overrightarrow{ij}}(\gamma) = -T_{\overrightarrow{ji}}(\gamma). \tag{4.12}$$

In combination, (4.12) and (4.11) imply (4.10).

With regard to the final clause of the Lemma, that $T_{(ji)}(\gamma)$ exists for all $1 \leq i \neq j \leq J$, if T is differentiable at $\gamma \in$int$\Delta\bar{\Omega}$, Rockafellar [1970] Theorem 25.2 shows that all directional derivatives $T'(\gamma|y)$ exist and are linear in $y = (y(1), ..., y(J))$. Hence they can be written in terms of partial derivatives $T_j(\gamma)$ as,

$$T'(\gamma|y) = \sum_{j=1}^{J} y(j)T_j(\gamma).$$

Hence, given $1 \leq i \neq j \leq J$,

$$-T_{\overrightarrow{ij}}(\gamma) = -\left[T_j(\gamma) - T_i(\gamma)\right] = T_i(\gamma) - T_j(\gamma) = T'(\gamma|e_i - e_j)) = T_{\overrightarrow{ji}}(\gamma),$$

verifying (4.9) 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 4.5: If $T_{(ji)}(\gamma)$ exists, then for any bijection $\sigma : \{1, .., J\} \to \{1, .., J\}$,

$$T_{(ji)}(\gamma) = T_{(\sigma(j)\sigma(i))}(\gamma^\sigma),$$

where,

$$\gamma^\sigma(j) = \gamma(\sigma^{-1}(j)).$$

Proof. Suppose that $T_{(ji)}(\gamma)$ exists and note by symmetry of T on $\Delta\bar{\Omega}$ (Lemma 4.1) and the bijective nature of σ,

$$T(\gamma) = T(\gamma^\sigma).$$

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, $\gamma^\sigma + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})$ are $\gamma^\sigma + \epsilon(e_i - e_j)$ symmetric, so that by the symmetry of T,

$$T\left[\gamma^\sigma + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})\right] = T(\gamma + \epsilon(e_i - e_j)).$$

Hence,

$$T_{(ji)}(\gamma^\sigma) = \lim_{\epsilon \to 0} \frac{T\left[\gamma^\sigma + \epsilon(e_{\sigma(i)} - e_{\sigma(j)})\right] - T(\gamma^\sigma)}{\epsilon} = \lim_{\epsilon \to 0} \frac{T\left[(\gamma + \epsilon(e_i - e_j))^\sigma\right] - T(\gamma^\sigma)}{\epsilon} = T_{(ji)}(\gamma),$$

establishing the Lemma. ■

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 1.1).

Lemma 4.6 (Optimality and Directional Derivatives): Consider a uniformly posterior-separable cost function K and an optimal strategy $(Q, q) \in \hat{\Lambda}(\mu, A|K)$ with supp $\mu = \bar{\Omega}$. Suppose $a, b \in A$ are chosen with positive probability and suppose γ^a is the posterior associated with choice a and γ^b is the posterior associated with the choice b where $\gamma^a, \gamma^b \in \text{int}\Delta\bar{\Omega}$. Consider $\theta \in \mathbb{R}^{J-1}$ s.t.,

$$N^c(\gamma) - \sum_{j=1}^{J-1} \theta(j)\gamma(j) \leq \sup_{c' \in A, \gamma' \in \bar{\Omega}} N^{c'}(\gamma') - \sum_{j=1}^{J-1} \theta(j)\gamma'(j),$$

all $\gamma \in \bar{\Omega}$ and $c \in A$. Then for all pairs of states $1 \leq i \neq j \leq J$:

1. For $c \in \{a, b\}$,
$$-T_{\overrightarrow{ij}}(\gamma^c) \leq c(i) - c(j) - [\theta(i) - \theta(j)] \leq T_{\overrightarrow{ji}}(\gamma^c); \tag{4.13}$$

2. For $c \in \{a, b\}$, if $T_{(ji)}(\gamma^c)$ exists,
$$T_{(ji)}(\gamma^c) = c(i) - c(j) - [\theta(i) - \theta(j)]; \tag{4.14}$$

3. If $T_{(ji)}(\gamma^a)$ and $T_{(ji)}(\gamma^b)$ both exist
$$N^a_{(ji)}(\gamma^a) = N^b_{(ji)}(\gamma^b). \tag{4.15}$$

Proof. Note that the Lagrangian Lemma (Lemma 1.1) implies the existence of the θ in the statement of the lemma, so the lemma is well posed. Note also that given that $J \geq 3$, we apply Lemma 4.5 to ensure that directional derivatives are invariant to re-indexing states if needed to make that state J is neither i nor j.

Define the function $F^c(\gamma)$ on $c \in A$ and $\gamma \in \Delta\bar{\Omega}$:

$$F^c(\gamma) \equiv N^c(\gamma) - \sum_{k=1}^{J-1} \theta(k)\gamma(k); \tag{4.16}$$

where,

$$N^c(\gamma) = \sum_{k=1}^{J} c(k)\gamma(k) - T(\gamma),$$

The Lagrangian Lemma implies that the supremal value of $F^c(\gamma)$ is achieved by the posteriors associated with any optimal policy. Hence

$$\sup_{c' \in A, \gamma' \in \bar{\Omega}} F^c(\gamma) = \sum_{k=1}^{J} c(k)\gamma^c(k) - T(\gamma^c) - \sum_{k=1}^{J-1} \theta(k)\gamma^c(k)$$

for $c \in \{a, b\}$.

By Lemma 4.4, if $T_{(ji)}(\gamma)$ 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

$$-T_{\overrightarrow{ij}}(\gamma^c) + c(i) - c(j) - [\theta(i) - \theta(j)] \leq 0 \leq T_{\overrightarrow{ji}}(\gamma^c) + c(i) - c(j) - [\theta(i) - \theta(j)],$$

confirming (4.13) .Conversely, if $T_{(ji)}(\gamma^c)$ exists for $c \in \{a, b\}$, this maximization implies that the corresponding derivative $F^c(\gamma^c)$ must equal zero,

$$F^c_{(ji)}(\gamma^c) = N^c_{(ij)}(\gamma) - [\theta(i) - \theta(j)] = c(i) - c(j) - T_{(ji)}(\gamma^c) - [\theta(i) - \theta(j)] = 0, \tag{4.17}$$

confirming (4.14).

To complete the proof, note that if T is differentiable at both posteriors, then then $T_{(ji)}(\gamma)$ exists as well. Applying (4.17) at both γ^a and γ^b yields,

$$N^a_{(ji)}(\gamma^a) = \theta(i) - \theta(j) = N^b_{(ji)}(\gamma^b),$$

confirming (4.15) and establishing the Lemma. ■

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 informative about the cost function. To do this we use decision problems for which with redundant states and show that the invariance of K places restrictions on how optimal posteriors behave as the prior is shifted across these redundant states. We first define the sets of posterior that we can relate through invariance. These are posteriors in which the odds ratio is equal to a fixed constant. We then present an extension of Feasibility Implies Optimality which shows that given any two posteriors with a common odds ratio for two states, we can construct a problem in which they are the optimal choice. Moreover, we can construct a set of problems indexed by t which differ only in how they allocate probability among these two states.

Definition 5 *Given $\alpha \in (0, \infty)$ and two states $1 \leq k \neq l \leq J$, we define the corresponding **ratio set** $\Gamma_{kl}(\alpha) \subset \Delta\bar{\Omega}$ as the set of posteriors in for which $\frac{\gamma(k)}{\gamma(l)} = \alpha$:*

$$\Gamma_{kl}(\alpha) = \left\{ \gamma \in \Delta\bar{\Omega} \,\middle|\, \frac{\gamma(k)}{\gamma(l)} = \alpha \right\}.$$

$\Gamma_{kl}(\alpha)$ is the intersection of $\Delta\bar{\Omega}$ and a $J-2$ dimensional linear subspace of $\mathbb{R}^J$. It is therefore convex and has dimension $J-2$. Figure A3 depicts $\Delta\bar{\Omega}$ for $J = 4$. The blue triangle represents $\Gamma_{12}(\alpha)$ for $\alpha = 2$ so that $\gamma \in \Delta\bar{\Omega}$ 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 $\gamma(2) = \frac{1}{3}$.

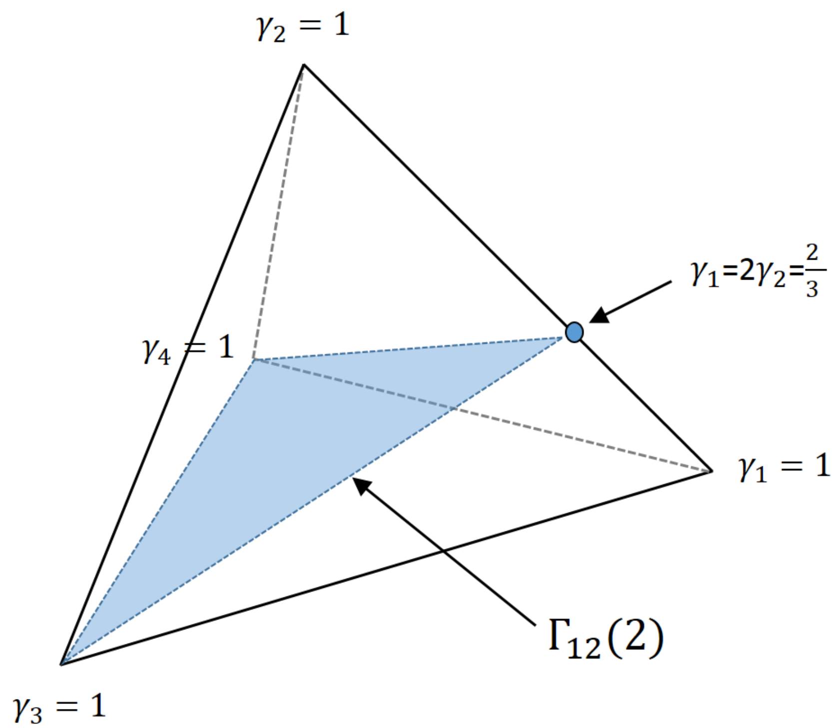

Figure A3

We now show that given γ and $\eta \in$ and $\Gamma_{kl}(\alpha)$ we can construct a problem (μ, A) such that γ and η are optimal posteriors.

Lemma 4.7: Suppose that K is uniformly posterior separable and invariant. Consider $\gamma \neq \eta \in$int$\bar{\Omega}$ with for $J \geq 3$, and $1 \leq k \neq l \leq J$ such that $\gamma, \eta \in \Gamma_{kl}(\alpha)$ some $\alpha \in (0, \infty)$,

$$\frac{\gamma(k)}{\gamma(l)} = \frac{\eta(k)}{\eta(l)} = \alpha. \tag{4.18}$$

Define the mean belief,

$$\bar{\mu}(j) = \frac{\gamma(j) + \eta(j)}{2},$$

for all j, and for $t \in [0, 1]$ define μ_t,:

$$\mu_t(j) = \begin{cases} t[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = k; \\ (1-t)[\bar{\mu}(k) + \bar{\mu}(l)] & \text{for } j = l; \\ \bar{\mu}(j) & \text{otherwise}; \end{cases} \tag{4.19}$$

define γ_t, and η_t analogously. Then there exists $a, b \in \mathcal{A}$ with $a(k) = a(l)$ and $b(k) = b(l)$ such that, $(Q_t, q_t) \in \hat{\Lambda}(\mu_t, A)$ where $Q_t(\gamma_t) = Q_t(\eta_t) = \frac{1}{2}$ and $q_t(a|\gamma_t) = 1$ and $q_t(b|\eta_t) = 1$.

Specifically, for $\bar{t} = \frac{\bar{\mu}(k)}{\bar{\mu}(k)+\bar{\mu}(l)} \in (0,1)$, $\mu_{\bar{t}} = \bar{\mu}$ and,

$$\chi_{\bar{t}}(j) = \chi(j)$$

for $\chi = \gamma, \eta$.

Proof. Fix $\gamma \neq \eta \in$int$\bar{\Omega}$ with for $J \geq 3$, and $1 \leq k \neq l \leq J$ such that (4.18) holds. Note that this is only possible if $J \geq 3$ since otherwise (4.18) implies $\gamma = \eta$.

Consider first the case $t = 1$, so that $\mu_1(l) = \gamma_1(l) = \eta_1(l) = 0$, $\mu_1(j) = \bar{\mu}(j)$ and $\chi_1(k) = \chi(j)$ for $j \neq k, l$, and

$$\begin{aligned} \mu_1(k) &= \bar{\mu}(k) + \bar{\mu}(l); \\ \chi_1(k) &= \chi(k) + \chi(l); \end{aligned}$$

for $\chi = \gamma, \eta$.

Note that since $\gamma \neq \eta$ and (4.18) holds, we know that there exists $j \in \bar{\Omega}\backslash\{k, l\}$ with $\gamma(j) \neq \eta(j)$, so that $\gamma_1 \neq \eta_1$. Since γ,η $\in$int$\Delta\bar{\Omega}$, γ_1,η_1 $\in$int$\Delta(\bar{\Omega}\backslash l)$. Corollary 1.2, implies the existence of a problem with $A = \{a, b\}$ such that there is an optimal strategy

$$(Q_1, q_1) \in \hat{\Lambda}(\mu_1, \{a, b\}|K)$$

in which the only chosen posteriors are γ_1 and η_1 and $Q_1(\gamma_1) = Q_1(\eta_1) = 0.5$. Feasibility Implies Optimality (Lemma 1.2) implies also that the deterministic strategy involving each action being chosen deterministically at its corresponding posterior is optimal. Without loss of generality, we set

$$q_1(a|\gamma_1) = q_1(b|\eta_1) = 1.$$

Given that $\mu_1(l) = \gamma_1(l) = \eta_1(l) = 0$, the payoffs to a and b in state l do not affect choice. We are therefore free to choose them as we wish. We set $a(l) = a(k)$ and $b(l) = b(k)$ so that states k and l are redundant.

We now consider decision problem $(\mu_t, \{a, b\})$. Consider the partition $\{\bar{\Omega}_z\}$ which combines states k and l and leaves all other states as singleton sets. Because states k and l are redundant, expected utility is independent of the probability μ_t. Because K is invariant

$$K(\mu_t, Q) \geq K(\mu_t, \nu_{\{\bar{\Omega}_z\},\mu_t} Q).$$

Since the policy (Q_1, q_1) is optimal for the problem $(\mu_1, \{a, b\})$, it follows that $(\nu_{\{\bar{\Omega}_z\},\mu_t} Q_1, q')$ is optimal for $(\mu_t, \{a, b\})$ where $q'(c|\nu_{\{\bar{\Omega}_z\},\mu_t}\gamma) = q_1(c|\gamma)$ for $c = a, b$. It remains to show that $\nu_{\{\bar{\Omega}_z\},\mu_t} Q_1 = Q_t$. But this follows from the fact that

$$\begin{aligned} \chi_t(k) &= t[\chi(k) + \chi(l)] \\ &= \frac{\mu_t(k)}{\bar{\mu}(k) + \bar{\mu}(l)}[\chi(k) + \chi(l)] \\ &= \nu_{\{\bar{\Omega}_z\},\mu_t} \chi_1(k) \end{aligned}$$

for $\chi = \{\gamma, \eta\}$.

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,

$$\begin{aligned} \chi_{\bar{t}}(j) &= t[\chi(k) + \chi(l)] \\ &= \frac{\chi(k) + \chi(l)}{\bar{\mu}(k) + \bar{\mu}(l)} \bar{\mu}(k) \\ &= \frac{\chi(k)}{\bar{\mu}(k)} \bar{\mu}(k) \\ &= \chi(j) \end{aligned}$$

for $\chi = \gamma, \eta$. The third equality follows form the assumption $\gamma, \eta \in \Gamma_{kl}(\alpha)$. ■

The Trapezoid Condition We are building towards using the rectangle condition for additive separability. The rectangle condition states that a function $f(a, b)$ on $X_A \times X_B$ is additively separable if

$$f(a_1, b_1) - f(a_2, b_1) = f(a_1, b_2) - f(a_2, b_2).$$

for all $a_1, a_2 \in X_A$ and $b_1, b_2 \in X_B$. Instead of constructing rectangles, we use the ratio sets to construct trapezoids which we then transform into rectangles. To construct these trapezoids we match sets of posteriors that lie on different ratio sets and differ in two dimensions. We now define the set of posteriors that differ only on states k and l.

Definition 6 *Given $\hat{\gamma} \in int\Delta\bar{\Omega}$ 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 $\Delta\bar{\Omega}$ 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 A4 reproduces Figure A3 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 $\gamma(1) = \gamma(2) = 1$.

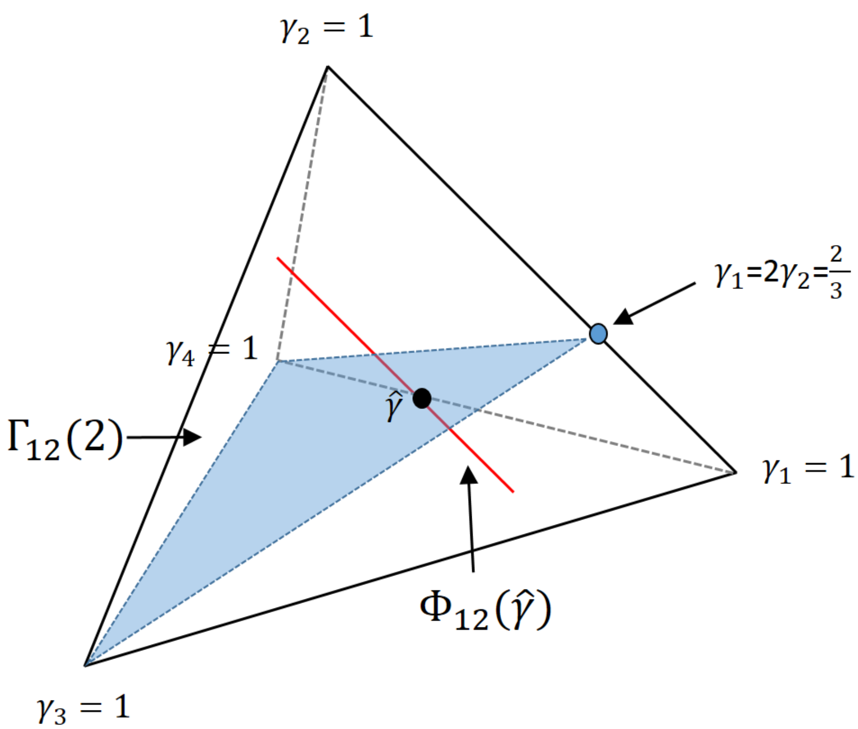

Figure A4

With these definitions we introduce the key sets of four posteriors.

Definition 7 *A set of four distinct posteriors* $\gamma_1, \gamma_2, \eta_1, \eta_2 \in int\Delta\bar{\Omega}$ *satisfy the* ***trapezoid condition*** *if there exist distinct* $\alpha_1 \neq \alpha_2 > 0$ *and* $1 \leq k \neq l \leq J$ *such that*:

1. γ_1,$\eta_1 \in \Gamma_{kl}(\alpha_1)$,
$$\frac{\gamma_1(k)}{\gamma_1(l)} = \frac{\eta_1(k)}{\eta_1(l)} = \alpha_1.$$
2. γ_2,$\eta_2 \in \Gamma_{kl}(\alpha_2)$,
$$\frac{\gamma_2(k)}{\gamma_2(l)} = \frac{\eta_2(k)}{\eta_2(l)} = \alpha_2.$$
3. $\gamma_2 \in \Phi_{kl}(\gamma_1)$, *so that* $\gamma_2(j) = \gamma_1(j)$ *for* $j \neq k, l$.
4. $\eta_2 \in \Phi_{kl}(\eta_1)$, *so that* $\eta_2(j) = \eta_1(j)$ *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 $\gamma_1 = \gamma_2$ and $\eta_1 = \eta_2$, contrary to the defining feature that these are distinct posteriors.

Figure A5 illustrates the Trapezoid Condition. γ_1 and η_1 both lie in $\Gamma_{12}(\alpha_1)$ and γ_2 and η_3 both lie in $\Gamma_{12}(\alpha_2)$. γ_1 and γ_2 both on the line segment $\Phi_{12}(\gamma_1)$ and η_1 and η_2 both on the line segment $\Phi_{12}(\eta_1)$. The key observation is that, since the points in $\Phi_{12}(\gamma_1)$ and $\Phi_{12}(\eta_1)$ only differ in their first and second coordinate, the two line segments are parallel, so that the points $\gamma_1, \gamma_2, \eta_1$ and η_2 form a trapezoid. Below we will use this observation to establish the additive separability of $T_{(ji)}(\gamma_1)$.

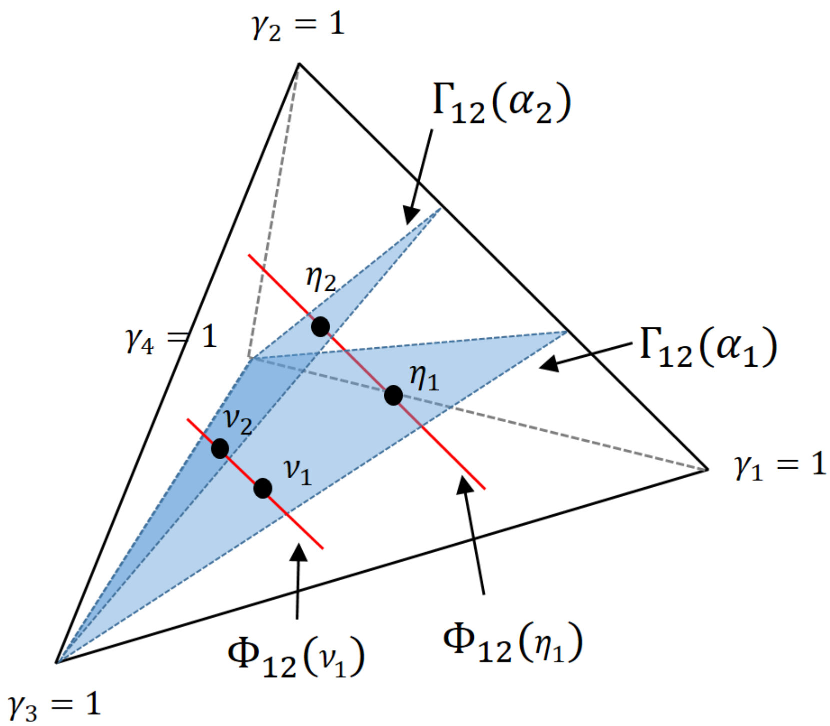

Figure A5

Equalization of Differences For Two-Sided Directional Derivatives Our next result relates the Trapezoid Condition, invariance, and the Lagrangian lemma. If four posteriors γ_1, γ_2, η_1, and η_2 satisfy the Trapezoid Condition and if T is differentiable at each of these points, then invariance and the Lagrangian Lemma relate the change in $T_{(ji)}$ between γ_1 and γ_2 to that between η_1 and η_2.

Lemma 4.8: Suppose K is invariant and uniformly posterior separable and suppose that T is differentiable at $\gamma_1, \gamma_2, \eta_1, \eta_2 \in \text{int}\Delta\bar{\Omega}$ that satisfy the Trapezoid Condition for some pair of distinct states $1 \leq k \neq l \leq J$, then

$$T_{(ji)}(\gamma_1) - T_{(ji)}(\gamma_2) = T_{(ji)}(\eta_1) - T_{(ji)}(\eta_2) \tag{4.20}$$

for all pairs of distinct states $1 \leq i \neq j \leq J$

Proof. Consider γ_1,η_1, γ_2 and η_2 satisfying the Trapezoid Condition such that $T_{(ji)}$ exists at all four points.

By Lemma 4.7, defining,

$$\bar{\mu} = \frac{\gamma_1 + \eta_1}{2},$$

there exists $a, b \in \mathcal{A}$ with $a(k) = a(l)$ and $b(k) = b(l)$ such that, for $t \in [0,1]$, $(Q_t, q_t) \in \hat{\Lambda}(\mu_t, A)$ where $Q_t(\gamma_t) = Q_t(\eta_t) = \frac{1}{2}$, $q_t(a|\gamma_t) = 1$ and $q_t(b|\eta_t) = 1$,and 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

$$\chi(t,j) = \begin{cases} t[\chi(k) + \chi(l)] & \text{for } j = k; \\ (1-t)[\chi(k) + \chi(l)] & \text{for } j = l; \\ \chi(j) & \text{otherwise.} \end{cases} \quad \text{for } 1 \leq j \leq J;$$

for $\chi = \{\gamma_1, \eta_1\}$. Where we have placed t as an argument in brackets and avoided the subscript so as to avoid confusion with γ_1 and η_1. In particular, for $\bar{t}_1 = \frac{\alpha_1}{\alpha_1+1} \in (0,1)$,

$$\frac{\gamma(\bar{t}_1, k)}{\gamma(\bar{t}_1, l)} = \frac{t[\chi(k) + \chi(l)]}{(1-t)[\chi(k) + \chi(l)]} = \alpha_2$$

so that $\gamma(\bar{t}_1) = \gamma_1$. Similarly

$$\begin{aligned} \eta(\bar{t}_1) &= \eta_1. \\ \mu(\bar{t}_1) &= \bar{\mu}, \end{aligned}$$

Moreover, since T is differentiable at both γ_1 and η_1, Lemma 4.6, the Optimality and Directional Derivatives Lemma, then implies,

$$N^a_{(ji)}(\gamma_1) = N^b_{(ji)}(\eta_1), \tag{4.21}$$

for all $i, j \in J$.

Now consider

$$\bar{t}_2 = \frac{\alpha_2}{\alpha_2 + 1}.$$

We have

$$\frac{\gamma(\bar{t}_2, k)}{\gamma(\bar{t}_2, l)} = \frac{t[\chi(k) + \chi(l)]}{(1-t)[\chi(k) + \chi(l)]} = \alpha_2$$

As $\gamma_2 \in \Phi_{kl}(\gamma_1)$, $\gamma(\bar{t}_2) = \gamma_2$. Similarly $\eta(\bar{t}_2) = \eta_2$. Therefore given the problem $(\mu(\bar{t}_2), A)$, γ_2 is the revealed posterior associated with a and η_2 is the revealed posterior associated with b. Since T is differentiable at both γ_2 and η_2, Lemma 4.6 then implies that,

$$N^a_{(ji)}(\gamma_2) = N^b_{(ji)}(\eta_2) \tag{4.22}$$

for all $i, j \in J$. Since $N^a(\gamma) = \sum_j a(j)\gamma(j) - T(\gamma)$, equations (4.21) and (4.22) imply

$$T_{(ji)}(\gamma_1) = b(i) - b(j) - a(i) + a(j) + T_{(ji)}(\eta_1)$$

and

$$T_{(ji)}(\gamma_2) = b(i) - b(j) - a(i) + a(j) + T_{(ji)}(\eta_2)$$

Subtracting these two equations yields the desired result. ■

Monotonicity and Limits Condition (4.20) 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 $\Delta\bar{\Omega}$ that satisfy the Trapezoid Condition. 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 $\gamma \in$int$\Delta\bar{\Omega}$ 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 4.9: Given $\gamma \in$int$\Delta\bar{\Omega}$ and $\epsilon > 0$ such that $\gamma + \epsilon(e_i - e_j) \in$int$\Delta\bar{\Omega}$,

$$T_{\overrightarrow{ji}}(\gamma) \leq -T_{\overrightarrow{ij}}(\gamma + \epsilon(e_i - e_j)) \leq T_{\overrightarrow{ji}}(\gamma + \epsilon(e_i - e_j)) \tag{4.23}$$

Proof. The result follows directly from monotonicity properties of the subdifferential maps of convex functions (Rockafellar [1970], 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,

$$Y \equiv (\gamma - \delta(e_i - e_j), \gamma + \delta(e_i - e_j)) \subset \gamma \in \text{int}\Delta\bar{\Omega},$$

which is possible since int$\Delta\bar{\Omega}$ is relatively open on the line. We then define convex function $G : \mathbb{R} \to \mathbb{R}$ to have value

$$G(\alpha) = T(\gamma + \alpha(e_i - e_j)),$$

on $\alpha \in (-\delta, \delta)$, with its value being infinite elsewhere. It is direct from the definitions that $T_{\overrightarrow{ji}}(\gamma + \alpha(e_i - e_j))$ and the right derivatives of $G(\alpha)$ are equivalent at corresponding points,

$$\begin{aligned} T_{\overrightarrow{ji}}(\gamma + \alpha(e_i - e_j)) &= \lim_{\varsigma \downarrow 0} \frac{T(\gamma + \alpha(e_i - e_j) + \varsigma(e_i - e_j)) - T(\gamma + \alpha(e_i - e_j))}{\varsigma} \\ &= \lim_{\varsigma \downarrow 0} \frac{G(\alpha + \varsigma) - G(\alpha)}{\varsigma} \equiv G'_+(\alpha). \end{aligned}$$

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 [1970] 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 (4.23). ■

Lemma 4.10: Given any sequence $\{\epsilon_n\}_{n=1}^{\infty}$ with $\epsilon_n > 0$ and $\lim_{n\to\infty} \epsilon_n = 0$,

$$\lim_{n\to\infty} T_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) = T_{\overrightarrow{ji}}(\gamma) \tag{4.24}$$

Proof. By the directional derivative monotonicity Lemma 4.9, given $\{\epsilon_n\}_{n=1}^{\infty}$ with $\epsilon_n > 0$, and $\lim_{n\to\infty} \epsilon_n = 0$, such that $\gamma + \epsilon_n(e_i - e_j) \in$int$\Delta\bar{\Omega}$, (4.23) implies that the limit exists and that the inequality survives,

$$\lim_{n\to\infty} T_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) \geq T_{\overrightarrow{ji}}(\gamma). \tag{4.25}$$

Conversely, Rockafellar [1970] Theorem 24.5 shows with full generality that given convex function $T : \mathbb{R}^J \to \bar{\mathbb{R}}$, any $\gamma \in \mathbb{R}^J$ at which $T(\gamma)$ is finite, and any sequence $\{\gamma^n\}_{n=1}^{\infty} \to \gamma$ and $y \in \mathbb{R}^J$,

$$\limsup T'(\gamma^n|y) \leq T'(\gamma|y). \tag{4.26}$$

Defining $\gamma^n = (\gamma + \epsilon_n(e_i - e_j)$ and $y = e_i - e_j$ we get,

$$\lim_{n\to\infty} T_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) \leq T_{\overrightarrow{ji}}(\gamma). \tag{4.27}$$

Combining (4.25) with (4.27) establishes (4.24), and completes the proof of the Lemma. ■

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 int$\Delta\bar{\Omega}$ whose affine hull is of lower dimension, several of which have appeared already in the proof. When thinking about T even on its full domain int$\Delta\bar{\Omega}$, there is a subtlety in the statement. Since int$\Delta\bar{\Omega}$ 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" int$\Delta\bar{\Omega}$. 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 8 *Given a non-empty convex set $Y \subset \mathbb{R}^J$, define $\Delta\Omega Y =$int$\Delta\bar{\Omega}(Y)$ to be the corresponding subdomain of int$\Delta\bar{\Omega}$,*

$$\Delta\Omega Y \equiv Y \cap int\Delta\bar{\Omega};$$

and define $T^Y : \Delta\Omega Y \to \mathbb{R}$ with $T^Y(\gamma) \equiv T(\gamma)$ to be the restriction of T to this domain. Finally, given $\gamma \in \Delta\Omega Y$ and $1 \leq i \neq j \leq J$ such that there exists $\delta > 0$ such that,

$$[\gamma - \delta(e_i - e_j), \gamma + \delta(e_i - e_j)] \subset int\Delta\bar{\Omega}(Y), \tag{4.28}$$

we define one and two-sided directional derivative $T^Y_{\overrightarrow{ji}}$ and $T^Y_{(ji)}$ in the standard manner.

We now state the key result about convergence of one-sided directional derivatives for appropriately selected sequences of posteriors.

Lemma 4.11: For any non-empty convex set $Y \subset \mathbb{R}^J$, T^Y is almost everywhere differentiable in the relative interior of $\Delta\Omega Y$ and whenever $T^{Y}_{\overrightarrow{ji}}(\gamma)$ is well-defined,

$$T^{Y}_{\overrightarrow{ji}}(\gamma) = T_{\overrightarrow{ji}}(\gamma). \tag{4.29}$$

Proof. Given that Y is non-empty and convex set, T^Y is a proper convex function. Rockafellar [1970] Theorem 25.5 translates precisely to the fact that the set $\{\gamma \in \Delta\Omega Y | T^Y \text{ is differentiable}\}$ is dense in $\Delta\Omega Y$ and its complement, the points of non-differentiability, is of measure zero in the relative interior of $\Delta\Omega Y$. The equality $T^{Y}_{\overrightarrow{ji}} = T_{\overrightarrow{ji}}$ is definitional given the existence of appropriate convergent sequences in the shared domain and equality of the underlying function. ■

Lemma 4.12: Suppose K is invariant and uniformly posterior separable. Given $\{\gamma_1, \gamma_2, \eta_1, \eta_2\} \subset \text{int}\Delta\bar{\Omega}$ satisfying the Trapezoid Condition for some pair of distinct states $1 \leq k \neq l \leq J$, then

$$T_{\overrightarrow{ji}}(\gamma_1) - T_{\overrightarrow{ji}}(\eta_1) = T_{\overrightarrow{ji}}(\gamma_2) - T_{\overrightarrow{ji}}(\eta_2), \tag{4.30}$$

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 $\{\xi_m\} \subset \text{int}\Delta\bar{\Omega}$ for $\xi = \gamma, \eta$ and $m = 1, 2$ satisfying the Trapezoid Condition. We construct four corresponding sequences of posteriors $\{\xi^n_m\}_{n=1}^{\infty}$ that converge to ξ_m as $n \to \infty$. We ensure that at each n (4.20) holds and that (4.20) converges to (4.30). Specifically, let $\Theta(\xi_m)$ be the set containing posteriors that lie within $\frac{1}{n}$ of ξ_m in the direction $\overrightarrow{ji}$:

$$\Theta(\xi_m) = \{\gamma \in \text{int}\Delta\bar{\Omega} | \gamma = \xi_m + \lambda(e_i - e_j) \text{ and } \lambda \in (0, \frac{1}{n})\}.$$

for all $\gamma \in \Theta(\xi_m)$.

Note that $\Theta(\xi_m)$ is a convex subset of $\text{int}\Delta\bar{\Omega}$, hence satisfies the conditions of the Lemma 4.11, so that $T^{\Theta(\xi_m))}_{(ji)} = T_{(ji)}$ 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 $T_{(ji)}$ exists,

$$\Lambda_n(\xi_m) = \{\lambda \in (0, \frac{1}{n}) | T_{(ij)}(\gamma) \text{ exists at } \gamma = \xi_m + \lambda(e_i - e_j)\}.$$

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,

$$\xi^n_m = \xi_m + \bar{\lambda}(n)(e_i - e_j),$$

for $\xi = \gamma, \eta$ and $m = 1, 2$.

By construction, for each n, $\xi^n_m \in \text{int}\Delta\bar{\Omega}$ for $\xi = \gamma, \eta$ and $m = 1, 2$ satisfy the Trapezoid Condition. To confirm, note that for $\xi = \gamma, \eta$ and $m = 1, 2$, $\xi^n_m(k)$ and $\xi_m(k)$ differ only in coordinates i and j. Hence we know that $\xi^n_m(k) = \xi_m(k)$ and $\xi^n_m(l) = \xi_m(l)$. Given that the ξ_m satisfy the Trapezoid Condition, it follows that the ξ^n_m satisfy the Trapezoid Condition as well. Since $T_{(ji)}$ exists at each of the ξ^n_m, and since the ξ^n_m satisfy the Trapezoid Condition, Lemma 4.8

states

$$T_{(ji)}(\gamma_1^n) - T_{(ji)}(\eta_1^n) = T_{(ji)}(\gamma_2^n) - T_{(ji)}(\eta_2^n) \tag{4.31}$$

Now consider each element $T_{(ji)}(\xi_m^n)$ 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 4.9 directly to conclude that (4.23) holds,

$$T_{\overrightarrow{ji}}(\xi_m^n) \geq T_{\overrightarrow{ji}}(\xi_m) \tag{4.32}$$

Since $\bar{\lambda}(n) \to 0$ we know in addition that $\lim_{n \longrightarrow \infty} \xi_m^n = \xi_m$, so that Lemma 4.10 applies to show that,

$$\lim_{n\to\infty} T_{\overrightarrow{ji}}(\xi_m^n) = \lim_{n\to\infty} T_{(ji)}(\xi_m^n) = T_{\overrightarrow{ji}}(\xi_m). \tag{4.33}$$

Substituting (4.33) in (4.31) establishes (4.30), and completes the proof. ■

Propagating Existence of Directional Derivatives A key result shows how to propagate existence of directional derivatives. If the two-sided directional derivative $T_{(ji)}(\gamma)$ exists at a point γ, then $T_{(ji)}(\eta)$ exists for all $\eta \in \Gamma_{ji}^{\alpha}$ where $\alpha = \frac{\gamma(j)}{\gamma(i)}$. The intuition for the result is that this set of η can be linked to γ by a problem in which the states i and j are redundant. Invariance 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. If $T(\gamma)$ is smooth in the direction (ji), $T(\eta)$ 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 [1970].

Lemma 4.13: Given $\gamma \in \text{int}\Delta\bar{\Omega}$ and $1 \leq i \neq j \leq J$ such that $T_{(ji)}(\gamma)$ exists, then $T_{(ji)}(\eta)$ exists for all $\eta \in \text{int}\Delta\bar{\Omega}$ such that:

$$\frac{\eta(j)}{\eta(i)} = \frac{\gamma(j)}{\gamma(i)}. \tag{4.34}$$

Proof. The proof is by contradiction. Choose γ such that $T_{(ji)}(\gamma)$ exists and suppose that there exists η satisfying (4.34) such that $T_{(ji)}(\eta)$ does not exist. By Lemma 4.4 above, this means that $-T_{\overrightarrow{jl}}(\eta) \neq T_{\overrightarrow{lj}}(\eta)$.

By Lemma 4.7 there exists (μ, A) with $A = \{a, b\}$ such that γ is the revealed posterior related to a and η is the revealed posterior related to b and the states i and j are redundant. Lemma 4.7 also 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 $t \in (0, 1)$, and γ_t, the revealed posterior for action a, and η_t,the revealed posterior for action b, are defined analogously. 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

$$N^a(\gamma_t) - \sum_{k=1}^{J-1} \theta_t(k)\gamma_t(k) = N^b(\eta_t) - \sum_{k\neq j} \theta_t(k)\eta_t(k) \geq N^c(\gamma') - \sum_{k=1}^{J-1} \theta(k)\gamma'(k)$$

for all $\gamma' \in \Delta\bar{\Omega}$ and $c = \{a, b\}$.

Since $T_{(ji)}(\gamma)$ exists, the Optimal Directional Derivative Lemma (Lemma 4.6) tells us that

$$T_{(ji)}(\gamma) = a(i) - a(j) - \theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) \tag{4.35}$$

and since $T_{(ji)}(\eta)$ does not exist, Lemma 4.6 implies that,

$$-T_{\overrightarrow{ij}}(\eta) - b(i) + b(J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) \leq 0 \leq T_{\overrightarrow{ji}}(\eta) - b(i) + b(J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j),$$

with one of these two inequalities strict. Without loss of generality suppose

$$T_{\overrightarrow{ji}}(\eta) - b(i) + b(J) + \theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) = \Delta > 0. \tag{4.36}$$

Define now $Y \subset \mathbb{R}^J$ as all vectors γ_t,

$$Y = \{\gamma_t \in \text{int}\Delta\bar{\Omega} | t \in [0, 1]\},$$

noting that, since $\Delta\bar{\Omega}Y = Y$ since $Y \subset$int$\Delta\bar{\Omega}$. Lemma 4.11 implies that T^Y is differentiable for almost all γ_t, so that Lemma 4.4 implies that the two-sided directional derivative,

$$T^Y_{(ji)}(\gamma_t) = T_{(ji)}(\gamma_t),$$

also exists for almost all $\gamma_t \in Y$.

Now consider a sequence $\gamma_{t(n)} \to \gamma$ such that $t(n) > \bar{t}$ and $T_{(ji)}(\gamma_{t(n)})$ exists. Lemma 4.10 implies that

$$\lim_{t(n)\to\bar{t}} T_{(ji)}(\gamma_{t(n)}) = T_{(ji)}(\gamma).$$

Therefore there exists $t(m) \neq \bar{t}$ such that $T_{(ji)}(\gamma_{t(m)}) \in (T_{(ji)}(\gamma), T_{(ji)}(\gamma) + \Delta)$. Given that $T_{(ji)}(\gamma_{t(m)})$ exists,

$$T_{(ji)}(\gamma_{t(m)}) = a(i) - a(j) - \theta_{t(m)}(i) - \theta_{t(m)}(j). \tag{4.37}$$

Hence, with $T_{(ji)}(\gamma_{t(m)}) - T_{(ji)}(\gamma) \in (0, \Delta)$, we can subtract the right-hand sides of (4.35) from (4.37) to conclude that,

$$\theta_{\bar{t}}(i) + \theta_{\bar{t}}(j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) \in (0, \Delta),$$

so that,

$$-\theta_{t(m)}(i) - \theta_{t(m)}(j) < -\theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) + \Delta. \tag{4.38}$$

Applying now the Optimal Directional Derivative Lemma (Lemma 4.6) to $\eta_{t(m)}$ we conclude that,

$$-T_{\overrightarrow{ij}}(\eta_{t(m)}) \leq b(i) - b(j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) \leq T_{\overrightarrow{ji}}(\eta_{t(m)}).$$

Substitution of (4.38) thereupon yields,

$$-T_{\overrightarrow{ij}}(\eta_{t(m)}) \leq b(i) - b(j) - \theta_{t(m)}(i) - \theta_{t(m)}(j) < b(i) - b(j) - \theta_{\bar{t}}(i) - \theta_{\bar{t}}(j) + \Delta = T_{\overrightarrow{ji}}(\eta)$$

But $t(m) > \bar{t}$, so that the Lemma 4.9 implies directly that $-T_{\overrightarrow{ij}}(\eta_{t(m)}) > T_{\overrightarrow{ji}}(\eta)$. This contradiction establishes the result. ■

Weak Form Additive Separability In this section we establish a weak form of additive separability. The basic observation is that (4.20) is very close to the rectangle condition for this form of additive separability. The difference is that $\gamma_1, \gamma_2, \eta_1, \eta_2 \in$int$\Delta\bar{\Omega}$ satisfying the Trapezoid Condition form a trapezoid, not a rectangle. We rectify this problem by deforming int$\Delta\bar{\Omega}$ and applying additive separability to the new space.

Our deformation is the combination of two mappings. The first maps int$\Delta\bar{\Omega}$ into the set of vectors $Z^{J-1}$of length $J-1$ whose elements z_j are strictly positive and sum to a number that is strictly less than one. Let

$$Z^M = \{(z_1, \ldots z_{J-1}) \in \mathbb{R}^{J-1} | z_j > 0 \text{ all } m \text{ and } \sum_m z_j < 1\},$$

We define the mapping Ψ_1 :int$\Delta\bar{\Omega} \to Z^{J-1}$ by dropping the coordinate $\gamma(J)$

$$\Psi_1(\gamma(j)) = \gamma(j) \text{ for } 1 \leq j \leq J-1.$$

The second mapping maps Z^{J-1} into $X = (0,1) \times Z^{J-2}$. Define $\Psi_2(z) : Z^{J-1} \to X$ by,

$$\Psi_2(z(j)) = \begin{cases} \frac{z(j)}{1-\sum_{m=2}^{J-1} z(m)} & \text{for } j = 1; \\ z(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases}$$

Now X is a rectangle. We define the mapping Ψ :int$\Delta\bar{\Omega} \to X$ as the composition of Ψ_1 and Ψ_2 :

$$\begin{aligned} \Psi &= \Psi_2(\Psi_1(\gamma)) = \begin{cases} \frac{\Psi_1(\gamma(j))}{1-\sum_{m=2}^{J-1} \Psi_1(\gamma(m))} & \text{for } j = 1; \\ \Psi_1(\gamma(j)) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases} \\ &= \begin{cases} \frac{\gamma(j)}{1-\sum_{m=2}^{J-1} \gamma(m)} & \text{for } j = 1; \\ \gamma(j) & \text{for } 2 \leq i \neq j \leq J-1. \end{cases} \end{aligned}$$

The next lemma shows that Ψ is bijective.

Lemma 4.14: $\Psi = \Psi_2 \circ \Psi_1$ is bijective.

Proof. Clearly Ψ_1 :int$\Delta\bar{\Omega} \to Z^{J-1}$ 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. The composition of two bijective mappings is bijective. This completes the proof. ■

The next lemma shows that, in this space, the Trapezoid Condition transforms into a rectangle condition on X.

Lemma 4.15: Given $\gamma_1, \gamma_2, \eta_1, \eta_2 \in \text{int}\Delta\bar{\Omega}$ that satisfy the Trapezoid Condition for states $k = 1$ and $j = J$, the elements $x_1, x_2, y_1, y_2 \in X$ such that $x_m = \Psi(\gamma_m)$ and $y_m = \Psi(\eta_m)$ 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 $x_m = \Psi(\gamma_m)$ and $y_m = \Psi(\eta_m)$ for $m = 1, 2$. By the Trapezoid Condition and the definition of Ψ, for $2 \leq j \leq J-1$ and $m = 1, 2$,

$$x_m(j) = \Psi(\gamma_m(j)) = \gamma_m(j) = \eta_m(j) = \Psi(\eta_m(j)) = \eta_m(j).$$

Note also that,

$$\begin{aligned} x_1(1) - x_2(1) &= \frac{\gamma_1(1)}{\gamma_1(1)+\gamma_1(J)} - \frac{\gamma_2(1)}{\gamma_2(1)+\gamma_2(J)} \\ &= \frac{\frac{\gamma_1(1)}{\gamma_1(J)}}{\frac{\gamma_1(1)}{\gamma_1(J)}+1} - \frac{\frac{\gamma_2(1)}{\gamma_2(J)}}{\frac{\gamma_2(1)}{\gamma_2(J)}+1} = \frac{\alpha}{\alpha+1} - \frac{\alpha}{\alpha+1} = 0 \end{aligned}$$

Similarly,

$$y_1(1) - y_2(1) = \frac{\frac{\eta_1(1)}{\eta_1(J)}}{\frac{\eta_1(1)}{\eta_1(J)}+1} - \frac{\frac{\eta_1(1)}{\eta_1(J)}}{\frac{\eta_1(1)}{\eta_1(J)}+1} = 0,$$

completing the proof. ■

With this we are in position to establish our first version of additive separability.

Lemma 4.16: Suppose that K is invariant and uniformly posterior separable. Then, given $2 \leq i \neq j \leq J-1$, $T_{\overrightarrow{ji}}(\gamma)$ 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,

$$T_{\overrightarrow{ji}}(\gamma) = \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2), ..., \gamma(J-1)\right)$$

Proof. We consider any four posteriors $\gamma_1, \gamma_2, \eta_1, \eta_2 \in$int$\Delta\bar{\Omega}$ that satisfy the Trapezoid Condition for states $k = 1$ and $l = J$. We now define $x_1, x_2, y_1, y_2 \in X$ by $x_m = \Psi(\gamma_m)$ and $y_m = \Psi(\eta_m)$ for $m = 1, 2$. We transfer the directional derivatives to this space by defining the function $\mathcal{T} : X \to \mathbb{R}$ by,

$$\mathcal{T}(x) \equiv T_{\overrightarrow{ji}}(\Psi^{-1}(x)),$$

using the bijective function Ψ :int$\Delta\bar{\Omega} \to X$ introduced in Lemma 4.14 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 $\gamma_1, \gamma_2, \eta_1$, and η_2 satisfy the Trapezoid Condition, Lemma 4.8, states,

$$T_{\overrightarrow{ji}}(\gamma_1) - T_{\overrightarrow{ji}}(\eta_1) = T_{\overrightarrow{ji}}(\gamma_2) - T_{\overrightarrow{ji}}(\eta_2)$$

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 4.15, $x_m = \Psi(\gamma_m)$ 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 $T_{\overrightarrow{ji}}$. Given $x = \Psi(\gamma)$,

$$\begin{aligned} T_{\overrightarrow{ji}}(\gamma) &= \mathcal{T}(\Psi^{-1}(\gamma)) = \mathcal{T}(x) = \boldsymbol{A}(x(1)) + \boldsymbol{B}(x(2),..,x(J-1)) \\ &= \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2),..,\gamma(J-1)\right), \end{aligned}$$

completing the proof. ■

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 T function.

Lemma 4.17: Suppose that K is invariant and uniformly posterior separable. If T is differentiable at $\gamma \in \text{int}\Delta\bar{\Omega}$, then, given $1 < i \neq j < J$ there exists $\boldsymbol{B} : \mathbb{R}^{J-2} \longrightarrow \mathbb{R}$ such that

$$T_{(ji)}(\gamma) = \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right) \tag{4.39}$$

Proof. We arbitrarily order states, fix states 1 and J, and consider distinct states $2 \leq i \neq j \leq J-1$. By Lemma 4.4, if T is differentiable at γ then $T_{(ji)}(\gamma)$ exists. We set $i = 2$ and $j = 3$. Given the symmetry of T (Lemma 4.1), this is without loss of generality.

Applying Lemma 4.16,

$$T_{(32)}(\gamma) = \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right).$$

By Lemma 4.4, we also know that,

$$T_{(23)}(\gamma) = -T_{(32)}(\gamma) = -\boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) - \boldsymbol{B}\left(\gamma(2), \gamma(3), ..., \gamma(J-2), \gamma(J-1)\right). \tag{4.40}$$

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 $\gamma^\sigma \in \text{int}\Delta\bar{\Omega}$ as the correspondingly permuted posterior, $\gamma^\sigma(j) = \gamma(\sigma^{-1}(j))$. Lemma 4.5 then states that, since $T_{(ji)}(\gamma)$ exists,

$$T_{(23)}(\gamma) = T_{(32)}(\gamma^\sigma)$$

Directly by Lemma 4.16,

$$\begin{aligned} T_{(32)}(\gamma^\sigma) &= \boldsymbol{A}\left(\frac{\gamma^\sigma(1)}{\gamma^\sigma(1)+\gamma^\sigma(J)}\right) + \boldsymbol{B}\left(\gamma^\sigma(2), \gamma^\sigma(3), ..., \gamma^\sigma(J-2), \gamma^\sigma(J-1)\right) \\ &= \boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right) + \boldsymbol{B}\left(\gamma(3), \gamma(2), ..., \gamma(J-2), \gamma(J-1)\right). \end{aligned} \tag{4.41}$$

Since both equal $T_{(23)}(\gamma)$ we know that the right-hand sides of (4.40) and (4.41) are equal,

$$2\boldsymbol{A}\left(\frac{\gamma(1)}{\gamma(1)+\gamma(J)}\right)=-\boldsymbol{B}\left(\gamma(2),\gamma(3),...,\gamma(J-2),\gamma(J-1)\right)-\boldsymbol{B}\left(\gamma(3),\gamma(2),...,\gamma(J-2),\gamma(J-1)\right) \tag{4.42}$$

By assumption $T_{(32)}(\gamma)$ exists so that by Lemma 4.13 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 (4.42) 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, there 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 4.18: Suppose that K is invariant and uniformly posterior separable. Then there exists $\bar{\boldsymbol{B}} : (0,1) \times (0,1) \to \mathbb{R}$ such that, given $\gamma \in \text{int}\Delta\bar{\Omega}$, and states $1 \leq i \neq j \leq J$,

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)). \tag{4.43}$$

Proof. Given arbitrarily fixed states 1 and J with $J \geq 4$, Lemma 4.17 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

$$T_{(32)}(\gamma) = \boldsymbol{B}\left(\gamma(2),\gamma(3),...,\gamma(J-2),\gamma(J-1)\right)$$

if T is differentiable at γ. If $J = 4$, then $\boldsymbol{B}$ has only two arguments and is of the desired form,

$$T_{(ji)}(\bar{\gamma}) = \boldsymbol{B}\left(\bar{\gamma}(i),\bar{\gamma}(j)\right) \equiv \boldsymbol{B}\left(\bar{\gamma}(2),\bar{\gamma}(3)\right) \equiv \bar{\boldsymbol{B}}\left(\bar{\gamma}(2),\bar{\gamma}(3)\right).$$

By the symmetry Lemma 4.5, this same function applies regardless of how we label states, completing the proof for $J = 4$.

$$T_{(ji)}(\bar{\gamma}) = \bar{\boldsymbol{B}}(\bar{\gamma}(i),\bar{\gamma}(j))$$

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 4.17 and Lemma 4.5, we can transpose posteriors 1 and s without changing the form of the function, so that,

$$T_{(ji)}(\gamma) = \boldsymbol{B}(\gamma(2),..,\gamma(s-1),\gamma(1),\gamma(s+1),...,\gamma(J-2),\gamma(J-1)). \tag{4.44}$$

Raising $\gamma(s)$ and reducing $\gamma(J)$ has no effect on the right hand side of (4.44), hence no effect on the RHS of (4.39) 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

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)),$$

where $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ is the common value. To complete the proof, note again that by the symmetry of T (Lemma 4.1), 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 γ at which T is differentiable 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 4.19: There exists $\bar{\boldsymbol{B}} : (0, 1) \times (0, 1) \to \mathbb{R}$ such that, given $\gamma \in$int$\Delta\bar{\Omega}$,

$$T_{\overrightarrow{ji}}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(j)). \tag{4.45}$$

Proof. Where $T_{(ji)}(\gamma)$, exists, Lemma 4.4 shows that it is equal to $T_{\overrightarrow{ji}}(\gamma)$. Hence the function defined in (4.43) is of the appropriate form for all γ at which T is differentiable. What is left is to establish that we can define $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ that equals $T_{\overrightarrow{ji}}(\gamma)$ on γ at which T is not differentiable.

Consider γ at which T is not differentiable, and consider any sequence $\{\gamma_n\}_{n=1}^{\infty}$ with $\gamma_n = \gamma + \epsilon_n(e_i - e_j)$ such that $T_{(ji)}(\gamma_n)$ 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 $T^{Y(\gamma,i,j)}$:int$\Delta\Omega Y(\gamma, i, j) \to \mathbb{R}$ is the restriction of T to $Y(\gamma, i, j) \cap$int$\Delta\bar{\Omega}$. Lemma 4.11 states that $T^{Y(\gamma,i,j)}$ is almost everywhere differentiable in the relative interior of int$\Delta\Omega Y(\gamma, i, j)$, and that $T^{Y(\gamma,i,j)}_{(ij)} = T_{(ji)}$. We can therefore select the sequence $\{\gamma_n\}_{n=1}^{\infty}$ from int$\Delta\Omega Y(\gamma, i, j)$.

As $T_{(ji)}(\gamma_n)$ exists, Lemma 4.18 implies,

$$T_{(ji)}(\gamma_n) = \bar{\boldsymbol{B}}(\gamma_n(i), \gamma_n(j)).$$

Lemma 4.10 then ensures that,

$$\lim_{n\to\infty} T_{\overrightarrow{ji}}(\gamma + \epsilon_n(e_i - e_j)) = T_{\overrightarrow{ji}}(\gamma).$$

We therefore define $\bar{\boldsymbol{B}}(\gamma)$ for γ at which T is not differentiable as,

$$\bar{\boldsymbol{B}}(\gamma) \equiv \lim_{n\to\infty} \bar{\boldsymbol{B}}(\gamma_n(i), \gamma_n(j)) = T_{\overrightarrow{ji}}(\gamma), \tag{4.46}$$

By construction we know that $T_{\overrightarrow{ji}}(\gamma) = \bar{\boldsymbol{B}}(\gamma)$ on the full domain, and that it is of the form $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ on γ at which T is differentiable. Equation (4.46) implies that $\bar{\boldsymbol{B}}(\gamma)$ takes the form $\bar{\boldsymbol{B}}(\gamma(i), \gamma(j))$ at points of non-differentiability, completing the proof of (4.45) 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 Invariance under Compression, such cases will be ruled out.

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. If $T_{(ji)}(\gamma), T_{(ik)}(\gamma), T_{(jk)}(\gamma)$ all exist then they are interdependent,

$$T_{(ji)}(\gamma) = T_{(jk)}(\gamma) + T_{(ki)}(\gamma). \tag{4.47}$$

In the next lemma 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 4.20 Given $\gamma \in \text{int}\Delta\bar{\Omega}$, suppose that there exist three distinct indices $1 \leq i, j, k \leq J$ such that $T_{(ki)}(\gamma)$ and $T_{(kj)}(\gamma)$ both exist. Then $T_{(ji)}(\gamma)$ exists and,

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(k)) - \bar{\boldsymbol{B}}(\gamma(j), \gamma(k)). \tag{4.48}$$

Proof. Given Lemma 4.1, we may take $i = 1$, $j = 2$ and $k = 3$ without loss of generality.

Given $\gamma \in \text{int}\Delta\bar{\Omega}$, define the set X, as the set of positive pairs (x_1, x_1) that sum to something less than the sum of $\gamma(1)$, $\gamma(2)$ and $\gamma(3)$,

$$X \equiv \left\{ x \in \mathbb{R}^2 | x_1, x_2 > 0 \text{ and } x_1 + x_2 < \sum_{l=\{1,2,3\}} \gamma(l) \right\}. \tag{4.49}$$

Define $\eta(x) \in \text{int}\Delta\bar{\Omega}$, as the vector that has x_1 and x_2 as its first arguments and agrees with γ for all arguments greater than three:

$$[\eta(x)](l) = \begin{cases} x_1 \text{ if } l = 1; \\ x_2 \text{ if } l = 2; \\ \sum_{l=\{1,2,3\}} \gamma(l) - x_1 - x_2 \text{ if } l = 3; \\ \gamma(l) \text{ otherwise} \end{cases} \tag{4.50}$$

Finally, define $H : X \to \mathbb{R}$ by

$$H(x) = T(\eta(x)) \tag{4.51}$$

Note that

$$\eta(\gamma(1), \gamma(2)) = \gamma$$

Note also that, given $T_{(32)}(\gamma)$ exists,

$$\begin{aligned} T_{(32)}(\gamma) &= \lim_{\epsilon \to 0} \frac{T(\gamma + \epsilon(e_2 - e_3)) - T(\gamma)}{\epsilon} = \lim_{\epsilon \to 0} \frac{T(\eta(\gamma(1), \gamma(2) + \epsilon)) - T(\eta(\gamma(1), \gamma(2)))}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{H(\gamma(1), \gamma(2) + \epsilon) - H(\gamma(1), \gamma(2))}{\epsilon} \end{aligned}$$

Hence,

$$T_{(32)}(\gamma) = \frac{\partial}{\partial\gamma(2)} H(\gamma(1), \gamma(2)). \tag{4.52}$$

Analogously,

$$T_{(31)}(\gamma) = \frac{\partial}{\partial\gamma(1)} H_1(\gamma(1), \gamma(2)).$$

Since both partials exist, note from Rockafellar [1970] Theorem 25.1 that H is differentiable at $(\gamma(1), \gamma(2))$ and from Theorem 25.2 that the directional derivative function $H'(\gamma(1), \gamma(2)|y)$ 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,

$$H'(\gamma(1), \gamma(2)|e_1 - e_2) = \frac{\partial}{\partial\gamma(1)} H_1(\gamma(1), \gamma(2)) - \frac{\partial}{\partial\gamma(2)} H(\gamma(1), \gamma(2)) = T_{(31)}(\gamma) - T_{(32)}(\gamma). \tag{4.53}$$

To complete the proof of (4.48), note directly from the definitions that $H'(\gamma|e_2 - e_1)$ is equal to $T_{(ji)}(\gamma)$,

$$\begin{aligned} H'(\gamma(1), \gamma(2)|e_1 - e_2) &= \lim_{\epsilon \to 0} \frac{H(\gamma(1) + \epsilon, \gamma(2) - \epsilon) - H(\gamma(1), \gamma(2))}{\epsilon} \\ &= \lim_{\epsilon \to 0} \frac{T(\gamma + \epsilon(e_1 - e_2)) - T(\gamma)}{\epsilon} \equiv T_{(21)}(\gamma) \end{aligned} \tag{4.54}$$

Setting the right-hand sides of (4.53) and (4.54) to equality establishes that

$$T_{(ji)}(\gamma) = T_{(ki)}(\gamma) - T_{(kj)}(\gamma).$$

In light of Lemma 4.19 this completes the proof of (4.48). ■

Lemma 4.20 points the way to a possible method for expressing $T_{(ji)}(\gamma)$ in a fully additively separable manner. Since $T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \gamma(k)) - \bar{\boldsymbol{B}}(\gamma(j), \gamma(k))$ and $T_{(ji)}(\gamma)$ only depends on $\gamma(i)$ and $\gamma(j)$, it should be possible to fix $\gamma(k)$ and write

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), \bar{x}) - \bar{\boldsymbol{B}}(\gamma(j), \bar{x})$$

so that $T_{(ji)}(\gamma)$ 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 $\gamma \in \text{int}\Delta\bar{\Omega}$. In the following, we establish additive separability on a subset of $\text{int}\Delta\bar{\Omega}$ and then drive the value of $\bar{x}$ down to zero to establish additive separability on the whole of $\text{int}\Delta\bar{\Omega}$.

Lemma 4.21: 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 $\gamma \in \text{int}\Delta\bar{\Omega}$ with $\gamma(k) = x(\epsilon)$, $T_{(ik)}(\gamma)$ 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,

$$\text{int}\Delta\bar{\Omega}Y(k, \epsilon) = \left\{\gamma \in \text{int}\Delta\bar{\Omega} | \gamma(k) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)\right\}. \tag{4.55}$$

Per the general prescription, define the restricted function $T^{Y(k,\epsilon)}$:int$\Delta\Omega Y(k,\epsilon) \to \mathbb{R}$, and note for arbitrary indices $1 \leq j \neq l \leq J$,

$$T^{Y(k,\epsilon)}_{\overrightarrow{lj}}(\gamma) = T_{\overrightarrow{lj}}(\gamma),$$

given that there is suitable variation of the posterior in all directions.

By Lemma 4.11 we know that $T^{Y(k,\epsilon)}$ is almost every differentiable in the relative interior of $\Delta\Omega Y(k,\epsilon)$. At any point of differentiability of $T^{Y(k,\epsilon)}$, we know by Lemma 4.4 that all two-sided directional derivatives exists. Hence, we know that $T_{(ik)}(\gamma)$ exists for almost all $\gamma \in$int$\Delta\Omega Y(k,\epsilon)$. But we already know from Lemma 4.19 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

$$I(x) = \{y \in (0, 1-x) | T_{(ik)}(\gamma) \text{ exists when } \gamma(k) = x \text{ and } \gamma(i) = y\} \subset (0, 1-x). \tag{4.56}$$

Note that the union of these sets across $x \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ is precisely the set of $\gamma \in$int$\Delta\Omega Y(k,\epsilon)$ on which $T_{(ik)}(\gamma)$ exists, which we know to have the same measure as the relative interior of int$\Delta\Omega Y(k,\epsilon)$. 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 4.22: 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,

$$T_{(ji)}(\gamma) = f^\epsilon(\gamma(i)) - f^\epsilon(\gamma(j)), \tag{4.57}$$

for all $\gamma \in$int$\Delta\bar{\Omega}$ such that $\gamma(i), \gamma(j) \in I(\epsilon)$ and $\gamma(i) + \gamma(j) < 1 - \epsilon$.

Proof. Given $\epsilon \in (0, 0.5)$, fix $x(\epsilon) \in \left(\frac{\epsilon}{8}, \frac{\epsilon}{4}\right)$ and the dense subset $I(\epsilon)$ of $(0, 1-\epsilon)$ so that the conditions of the Lemma 4.21 are satisfied. Now consider int$\Delta\bar{\Omega}(\epsilon)$, the set of posteriors for which Lemma 4.21 tells us that both $T_{(ik)}(\gamma)$ and $T_{(jk)}(\gamma)$ are well-defined,

$$\text{int}\Delta\bar{\Omega}(\epsilon) = \left\{\gamma \in \text{int}\Delta\bar{\Omega} | \gamma(i), \gamma(j) \in I(\epsilon), \gamma(k) = x(\epsilon)\right\},$$

Since both $T_{(ik)}(\gamma)$ and $T_{(jk)}(\gamma)$ exist on int$\Delta\bar{\Omega}(\epsilon)$, by Lemma 4.21,

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), x(k)) - \bar{\boldsymbol{B}}(\gamma(j), x(k)). \tag{4.58}$$

for $\gamma \in$int$\Delta\bar{\Omega}(\epsilon)$. 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 according to Lemma 4.18,

$T_{(ji)}(\gamma)$ depends only on $\gamma(i)$ and $\gamma(j)$. Hence given γ' such that $\gamma'(i) = \gamma(i)$ and $\gamma'(j) = \gamma(i)$,

$$T_{(ji)}(\gamma') = T_{(ji)}(\gamma) = f^{\varepsilon}(\gamma(i)) - f^{\varepsilon}(\gamma(j)).$$

Hence the characterization applies to all $\gamma \in$int$\Delta\bar{\Omega}$ such that $\gamma(i), \gamma(j) \in I(\epsilon)$ and $\gamma(i)+\gamma(j) < 1-\epsilon$ as required. ■

Lemma 4.23: There exists $f : \bar{I} \to \mathbb{R}$ with $\bar{I} \subset (0,1)$ of full measure such that for all $\gamma \in$int$\Delta\bar{\Omega}$ with $\gamma(i), \gamma(j) \in \bar{I}$, $T_{(ji)}(\gamma)$ exists and,

$$T_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)) \tag{4.59}$$

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 4.21 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 $T_{(ji)}$ 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 $\gamma(i) \in I(n)$, $T_{(ji)}$ exists, and since $\gamma(i) \in I(n+1)$, $T_{(ki)}$ exists. It follows from Lemma 4.20, that $T_{(kj)}$ 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 $\gamma(i) \in I(n)$, $T_{(ji)}$ exists, and since $T_{(kj)}$ exists, it follows from Lemma 4.20, that $T_{(ki)}$ exists. Hence $y \in I(n+1)$.

By Lemma 4.22,

$$T_{(ji)}(\gamma) = \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n))$$

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

$$\begin{aligned} T_{(ji)}(\gamma) &= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n)) \\ &= \bar{\boldsymbol{B}}(\gamma(i), x(n)) - \bar{\boldsymbol{B}}(z, x(n)) - \bar{\boldsymbol{B}}(\gamma(j), x(n)) + \bar{\boldsymbol{B}}(z, x(n)) \\ &= \boldsymbol{G}^n(\gamma(i)) - \boldsymbol{G}^n(\gamma(j)) \end{aligned}$$

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 4.20 applied to $\tilde{\gamma}$ with $\tilde{\gamma}(i) = x(n+1)$, $\tilde{\gamma}(j) = x(n)$ and $\tilde{\gamma}(k) = \gamma(i)$:

$$T_{(ji)}(\tilde{\gamma}) = \bar{\boldsymbol{B}}(x(n+1), x(n)) = \bar{\boldsymbol{B}}(x(n+1), \gamma(i)) - \bar{\boldsymbol{B}}(x(n), \gamma(i)) = -\bar{\boldsymbol{B}}(\gamma(i), x(n+1)) + \bar{\boldsymbol{B}}(\gamma(i), x(n))$$

and similarly for z in place of $\gamma(i)$.

Hence we can define a limit function $f : \cup_{n=1}^{\infty} I(n) \rightarrow \mathbb{R}$ 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 4.22, we know that with $\gamma(i), \gamma(j) \in I(n)$, a dense subset of $(0, 1 - \epsilon(n))$, (4.59) holds,

$$T_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j). \tag{4.60}$$

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 4.24: The function $f : \bar{I} \longrightarrow \mathbb{R}$ defined in Lemma 4.23 for which (4.59) 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 $\gamma \in \text{int}\Delta\bar{\Omega}$ with $\gamma(i) = x$ and $\gamma(j) = x + \epsilon$, by Lemma 4.23,

$$T_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j)) = f(x) - f(x + \epsilon).$$

If we now define $\gamma' \in \text{int}\Delta\bar{\Omega}$ as,

$$\gamma' = \gamma + \epsilon(e_i - e_j),$$

Lemma 4.23 implies that,

$$T_{(ji)}(\gamma') = f(\gamma'(i)) - f(\gamma'(j)) = f(x + \epsilon) - f(x).$$

By the monotonicity lemma, Lemma 4.9, $\gamma' = \gamma + \epsilon(e_i - e_j)$ for $\epsilon > 0$ implies $T_{(ji)}(\gamma') \geq T_{(ji)}(\gamma)$, 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 4.25: Given $\eta \in$int$\Delta\bar{\Omega}$, $T_{(ji)}(\eta)$ exists if and only if $f(\gamma(i)) - f(\gamma(j))$ is continuous at η.

Proof. Consider $\eta \in$int$\Delta\bar{\Omega}$. 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 $T^{\ Y(\eta,i,j)}$. By Lemma 4.11 this function is differentiable almost everywhere in the relative interior of the restricted domain int$\Delta\Omega Y(\eta, i, j)$. Hence the directional derivative in the only relevant direction,

$$T_{(ji)}^{Y(k,\epsilon)}(\gamma) = T_{(ji)}(\gamma),$$

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 $\gamma_n = \eta + \epsilon(n)(e_i - e_j)$, $T_{(ji)}(\gamma_n)$ exists. We select also $\{\epsilon'(n)\}_{n=1}^{\infty} < 0$ with $\lim_{n\to\infty} \epsilon'(n) = 0$ such that, defining $\gamma'_n = \eta + \epsilon'(n)(e_i - e_j)$, $T_{(ji)}(\gamma'_n)$ exists.

Since T is convex we know that the one-sided directional derivatives are monotonically increasing,

$$\lim_{n\to\infty} T_{(ji)}(\gamma'_n) \leq -T_{\overrightarrow{ji}}(\eta) \leq T_{\overrightarrow{ji}}(\eta) \leq \lim_{n\to\infty} T_{(ji)}(\gamma_n). \tag{4.61}$$

We now use Lemma 4.23 to substitute in (4.61) at all γ_n and γ'_n since $T_{(ji)}(\circ)$ is well-defined at these points, to arrive at,

$$\begin{aligned} T_{(ji)}(\gamma_n) &= f(\eta(i) + \epsilon_n) - f(\eta(j) - \epsilon_n); \\ T_{(ji)}(\gamma'_n) &= f(\eta(i) + \epsilon'_n) - f(\eta(j) - \epsilon'_n); \end{aligned}$$

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,

$$\lim_{n\to\infty} T_{(ji)}(\gamma'_n) = \lim_{n\to\infty} T_{(ji)}(\gamma_n),$$

hence by (4.61),

$$-T_{\overrightarrow{ji}}(\eta) = T_{\overrightarrow{ji}}(\eta),$$

establishing through Lemma 4.4 that $T_{(ji)}(\eta)$ exists.

Suppose conversely that $T_{(ji)}(\eta)$ does not exist. In this case we know by Lemma 4.4 that $T_{\overrightarrow{ji}}(\eta) < T_{\overrightarrow{ji}}(\eta)$, so that by (4.61),

$$\lim_{n\to\infty} T_{(ji)}(\gamma_n) = \lim_{n\to\infty} f(\eta(i)+\epsilon'_n) - f(\eta(j)-\epsilon'_n) < \lim_{n\to\infty} T_{(ji)}(\gamma_n) = \lim_{n\to\infty} f(\eta(i)+\epsilon_n) - f(\eta(j)-\epsilon_n),$$

establishing that $f(\gamma(i)) - f(\gamma(j))$ is discontinuous at η, and completing the proof. ■

Existence of Directional Derivatives

Lemma 4.26: Given η and given $\alpha = \frac{\eta(k)}{\eta(l)}$, if $T_{(kl)}(\eta)$ exists then T is differentiable for almost all $\gamma \in \Gamma_{kl}(\alpha)$.

Proof. Consider $\eta \in$int$\Delta\bar{\Omega}$ such that $T_{(kl)}(\eta)$ exists and set $\alpha = \frac{\eta(k)}{\eta(l)}$. Since $\Gamma_{kl}(\alpha)$ is convex, $T^{\Gamma_{kl}(\alpha)}$ almost everywhere differentiable on the relative interior of $\Gamma_{kl}(\alpha)$ by Lemma 4.11. At points of differentiability, we know from Lemma 4.4 that $T_{(ji)}(\gamma) = T^{\Gamma_{kl}(\alpha)}_{(ji)}(\gamma)$ 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 $T^{\Gamma_{kl}(\alpha)}$ is differentiable. As $T_{(kl)}(\eta)$ exists, Lemma 4.13 implies that $T_{(kl)}(\gamma)$ also exists. Hence at all such γ, we know that all 2-sided directional derivatives exist. Following precisely the steps in Lemma 4.20, we can remove an arbitrary state $k \neq i, j$ from the domain and construct set X as in (4.49), then define $\eta(x) \in$int$\Delta\bar{\Omega}$ on $x \in X$ as in (4.50) and function $H(x)$ on X by (4.51), whose partial derivatives are precisely the directional derivatives $T_{(km)}(\gamma)$,

$$T_{(km)}(\gamma) = H_1(\gamma(m), \gamma(i)),$$

all $m \neq k$.

Since all partials of this function therefore exist, we note from Rockafellar [1970] 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 [1970] Theorem 25.2 implies that T is differentiable at γ, completing the proof. ■

Lemma 4.27: $T_{(ji)}(\gamma)$ exists for all i, j and $\gamma \in$int$\Delta\bar{\Omega}$.

Proof. The proof is by contradiction. Consider a posterior η at which $T_{(ji)}(\eta)$ does not exist. It follows from Lemma 4.25 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 4.25, $T_{(ji)}(\gamma)$ 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 $\gamma \in \Gamma_{ji}^{\alpha}$ such that $T_{(ji)}(\gamma)$ does not exist. But by Lemma 4.13, if $T_{(ji)}(\gamma)$ exists for any $\eta \in \Gamma_{ji}^{\alpha}$, then $T_{(ji)}(\gamma)$ exists for all $\gamma \in \Gamma_{ji}^{\alpha}$. Hence for almost all $\alpha \in \left(\frac{\eta(i)}{1-\eta(i)}, \infty\right)$, $T_{(ji)}(\gamma)$ does not exist for any $\gamma \in \Gamma_{ji}^{\alpha}$. 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 T is differentiable almost everywhere. This contradiction establishes the result. ■

Lemma 4.28: T is continuously differentiable on $\gamma \in \text{int}\Delta\bar{\Omega}$ and $f(\gamma(j))$ is continuous on $\text{int}\Delta\bar{\Omega}$

Proof. Lemma 4.27 establishes that the directional derivatives $T_{(ji)}(\gamma)$ exist for all (ji) and all $\gamma \in \text{int}\Delta\bar{\Omega}$. It follows from Rockafellar [1970] Theorem 25.2 that T is differentiable and from Rockafellar [1970] Corollary 25.5.1 that T is continuously differentiable on $\gamma \in \text{int}\Delta\bar{\Omega}$. Since $T_{(ji)}(\gamma) = f(\gamma(i)) - f(\gamma(j))$, the continuity of f follows. ■

4.2.4 Invariance

We can use the arguments in Lemma 3 of Hebert and La'O (2020) to show that if the cost function K is differentiable invariant and posterior-separable, then $T_{\mu}(\gamma)$ is invariant where

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma) T_{\mu}(\gamma)$$

First, given $\bar{\Omega}$, $\mu \in \text{int(supp } \bar{\Omega})$ and $\{\bar{\Omega}_z\}_{z=1,\ldots Z}$, we define the operator. $\nu_{\bar{\Omega}_z,\mu} : \Delta\bar{\Omega} \to \Delta\bar{\Omega}$ as follows. Given $\gamma \in \Delta\bar{\Omega}$ construct $\gamma' = \nu_{\bar{\Omega}_z,\mu}(\gamma)$ as the probability distribution that satisfies the following two conditions:

1. For all $\bar{\Omega}_z$, $\gamma(\bar{\Omega}_z) = \gamma'(\bar{\Omega}_z)$
2. For all $\omega \in \bar{\Omega}_z$, $\gamma'(\omega|\bar{\Omega}) = \mu(\omega|\bar{\Omega}_z)$

Given $\nu_{\bar{\Omega}_z,\mu}$, we define the invariance of $T_{\mu}(\gamma)$.

Definition 9 $T_{\mu}(\gamma)$ *is* ***invariant*** *if for all finite sets of states, $\bar{\Omega} \subset \Omega$, all partitions of $\bar{\Omega}$, $\{\bar{\Omega}_z\}_{z=1,\ldots Z}$, all pairs of priors μ and μ' that place equal probability on each partition subset, and all posteriors $\gamma \in \Delta\bar{\Omega}$:*

$$T_{\mu}(\gamma) \geq T_{\mu}(\nu_{\bar{\Omega}_z,\mu}(\gamma))$$

and

$$T_{\mu}(\nu_{\bar{\Omega}_z,\mu}(\gamma)) = T_{\mu'}(\nu_{\bar{\Omega}_z,\mu'}(\gamma)).$$

We now have the following lemma.

Lemma 4.29: Suppose that K is invariant posterior-separable such that

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma) T_{\mu}(\gamma)$$

and that $T_\mu(\gamma)$ is differentiable. Then $T_\mu(\gamma)$ is invariant.

Proof. Fix a finite set of states $\bar{\Omega} \subset \Omega$, a partition $\{\bar{\Omega}_z\}$ of $\bar{\Omega}$, a pair of priors μ and μ' that place equal probability on each partition subset, and a posterior $\gamma \in \Delta\bar{\Omega}$. We wish to show that:

$$T_\mu(\gamma) \geq T_\mu(\nu_{\{\bar{\Omega}_z\},\mu}(\gamma)) \tag{4.62}$$

and

$$T_\mu(\nu_{\{\bar{\Omega}_z\},\mu}(\gamma)) = T_{\mu'}(\nu_{\{\bar{\Omega}_z\},\mu'}(\gamma)). \tag{4.63}$$

We begin by establishing monotonicity (4.62). Suppose, for now, that $\mu \in$int$\Delta\bar{\Omega}$. Consider two information structures:

$$\begin{aligned} Q_\varepsilon(\gamma) &= \varepsilon \\ Q_\varepsilon\left(\mu - \frac{\varepsilon}{1-\varepsilon}(\gamma - \mu)\right) &= 1 - \varepsilon \end{aligned}$$

and

$$\begin{aligned} \hat{Q}_\varepsilon(\nu_{\{\bar{\Omega}_z\},\mu}\gamma) &= \varepsilon \\ \hat{Q}_\varepsilon\left(\nu_{\{\bar{\Omega}_z\},\mu}\gamma - \frac{\varepsilon}{1-\varepsilon}(\mu - \nu_{\{\bar{\Omega}_z\},\mu}\gamma)\right) &= 1 - \varepsilon \end{aligned}$$

Note that $\hat{Q} = \nu_{\{\bar{\Omega}_z\},\mu}Q$.

At $\varepsilon = 0$, neither Q_0 and $\hat{Q}_0$ involves any learning and both sets of posteriors are equal to the prior, so that

$$K(Q_0, \mu) = K(\hat{Q}_0, \mu) = 0$$

Since K is invariant

$$K(Q_\varepsilon, \mu) \geq K(\hat{Q}_\varepsilon, \mu)$$

It follows that

$$\frac{d}{d\varepsilon}\left[K(Q_\varepsilon, \mu) - K(\hat{Q}_\varepsilon, \mu)\right]_{\varepsilon=0} \geq 0$$

Note that

$$\begin{aligned} \frac{d}{d\varepsilon}\left[K(Q_\varepsilon, \mu)\right]_{\varepsilon=0} &= \frac{d}{d\varepsilon}\left[\varepsilon T_\mu(\gamma) + (1-\varepsilon)T_\mu\left(\mu - \frac{\varepsilon}{1-\varepsilon}(\gamma - \mu)\right)\right]_{\varepsilon=0} \\ &= \left[T_\mu(\gamma) - T_\mu(\mu) + (1-\varepsilon)\frac{d}{d\varepsilon}T_\mu\left(\mu - \frac{\varepsilon}{1-\varepsilon}(\gamma - \mu)\right)\right]_{\varepsilon=0} \\ &= T_\mu(\gamma) \end{aligned}$$

where the last equality follows from the observations that $T_\mu(\mu) = 0$ and since $T_\mu(\gamma)$ is positive, the directional derivatives are all zero at $T(\mu)$. Similarly $\frac{d}{d\varepsilon}\left[K(\mu, \hat{Q}_\varepsilon)\right]_{\varepsilon=0} = T_\mu(\nu_{\{\bar{\Omega}_z\},\mu}\gamma)$. Together we have,

$$T_\mu(\gamma) - T_\mu(\nu_{\{\bar{\Omega}_z\},\mu}\gamma) \geq 0$$

as desired.

If $\mu \in \partial\bar{\Omega}$, then we repeat the argument with $\bar{\Omega}'$ =supp μ.

Now consider (4.62). Consider μ_1 and μ_2 which place equal probabilities on each subset $\bar{\Omega}_z$, $\mu_1(\bar{\Omega}_z) = \mu_2(\bar{\Omega}_z)$. Consider the information structures.

$$\begin{aligned} \hat{Q}_{\varepsilon,x}(\nu_{\{\bar{\Omega}_z\},\mu_x}\gamma) &= \varepsilon \\ \hat{Q}_{\varepsilon,x}\left(\mu_x - \frac{\varepsilon}{1-\varepsilon}(\nu_{\{\bar{\Omega}_z\},\mu_x}\gamma - \nu_{\{\bar{\Omega}_z\},\mu_x}\mu)\right) &= 1-\varepsilon \end{aligned}$$

for $x = 1, 2$. Since K is invariant

$$K(\mu_1, \hat{Q}_{\varepsilon,1}) = K(\mu_2, \hat{Q}_{\varepsilon,2})$$

for all ε. Hence

$$\frac{d}{d\varepsilon}\left[K(\mu_1, \hat{Q}_{\varepsilon,1}) - K(\mu_2, \hat{Q}_{\varepsilon,2})\right]_{\varepsilon=0} = 0$$

Following the above logic

$$\frac{d}{d\varepsilon}\left[K(\mu_1, \hat{Q}_{\varepsilon,1})\right]_{\varepsilon=0} = T_{\mu_1}(\nu_{\{\bar{\Omega}_z\},\mu_1}\gamma)$$

and

$$\frac{d}{d\varepsilon}\left[K(\mu_2, \hat{Q}_{\varepsilon,2})\right]_{\varepsilon=0} = T_{\mu_2}(\nu_{\{\bar{\Omega}_z\},\mu_2}\gamma)$$

It follows that

$$T_{\mu_1}(\nu_{\{\bar{\Omega}_z\},\mu_1}\gamma) = T_{\mu_2}(\nu_{\{\bar{\Omega}_z\},\mu_2}\gamma)$$

as required. ■

4.2.5 A Bregman divergence

There is a relationship between posterior-separable attention costs functions and divergences. A divergence is a weak notion of the distance between probability distributions. Given an arbitrary state space S, a divergence $D(p||q)$ is a function from int$\Delta S\times$intΔS to $\bar{\mathbb{R}}$ satisfying $D(p||q) \geq 0$ and $D(q||q) = 0$. As $T_\mu(\gamma) \geq 0$ and $T_\mu(\mu)$, $T_\mu(\gamma)$ defines a divergence on supp μ.

Fix $\bar{\Omega}$ and consider $\mu \in$int$\Delta\bar{\Omega}$. Let $D_B(\gamma||\mu)$ map int$\Delta\bar{\Omega}\times$int$\Delta\bar{\Omega}$ into $\bar{\mathbb{R}}$ be defined as follows. Given $\mu, \gamma \in \Delta$int$\bar{\Omega}$

$$D_B(\gamma||\mu) = T_\mu(\gamma) = T_{\bar{\Omega}}(\gamma) - T_{\bar{\Omega}}(\mu) + \nabla T_{\bar{\Omega}}(\mu)\cdot(\gamma - \mu)$$

Note that at this point we may apply Lemma 4.28 and assume that $T_{\bar{\Omega}}$ is differentiable so that $\nabla T_{\bar{\Omega}}(\mu)$ is the gradient vector. $D_B(\gamma||\mu)$ is a Bregman divergence. Since $D_B(\gamma||\mu)$ is an invariant and differentiable Bregman divergence. The result of Jiao, *et al.* [2014] implies that $D_B(\gamma||\mu)$ is the Kullback-Leibler divergence.[3]

[3] In the working paper Caplin, Dean and Leahy [2017], we present an alternative proof that the Shannon cost function is the only posterior-separable cost function that satisfies the two Axioms, Locally Invariant Posteriors and Invariance under Compression.

4.2.6 A single cost function

To this point we have fixed the state space $\bar{\Omega}$ and shown that for decision problems (μ, A) with supp $\mu = \bar{\Omega}$, the cost function is the expected Kullback-Leibler divergence with a parameter κ that may depend on $\bar{\Omega}$. We complete the proof by by showing that κ is independent of $\bar{\Omega}$.

First, note that the Symmetric Cost Lemma (Lemma 3.2) implies that the cost function has the same form for all state spaces with equal cardinality, so that κ is equalized for all $\bar{\Omega}$ with cardinality J.

Now consider $\bar{\Omega}_1$ and $\bar{\Omega}_2$ with $\bar{\Omega}_2 = \bar{\Omega}_1 \cup \bar{\omega}$. If $\bar{\Omega}_1$ has cardinality J, then $\bar{\Omega}_2$ has cardinality $J+1$. Suppose $J \geq 2$. Consider an arbitrary decision problem (μ_1, A) such that supp $\mu_1 = \bar{\Omega}_1$. Suppose that for all $\omega_1, \omega_2 \in \bar{\Omega}_1$, the payoff profiles $A(\omega_1) \neq A(\omega_2)$. So that (μ_1, A) is a basic form of itself, $(\mu_1, A) \in \mathcal{B}(\mu_1, A)$. Suppose further that there exists $\omega' \in \bar{\Omega}_1$ such that $A(\bar{\omega}) = A(\omega')$ and $\mu_2(\bar{\omega}) + \mu_2(\omega') = \mu_1(\omega')$ so that (μ_1, A) is a basic form of (μ_1, A), $(\mu_1, A) \in \mathcal{B}(\mu_2, A)$.

Since K is invariant, the data generated by K satisfies Invariance under Compression.

Consider $(Q_1, q_1) \in \hat{\Lambda}(\mu_1, A)$. Invariance under Compression implies the existence of $(Q_2, q_2) \in \hat{\Lambda}(\mu_2, A)$ such that $P_{(Q_2,q_2|\mu_2,A)}(a|\omega) = P_{(Q_1,q_1|\mu_1,A)}(a|\omega)$ for all $\omega \in \bar{\Omega}_1$ where $P_{(Q_1,q_1|\mu_1,A)}(a|\omega)$ is the state dependent stochastic choice data generated by the policy (Q_1, q_1) in the problem (μ_1, A). As $\mu_1(\omega) = \mu_2(\omega)$ for all $\omega \in \bar{\Omega}_1 \backslash \omega'$, Bayes rule implies

$$\gamma_1^a(\omega) = \frac{P_{(Q_1,q_1|\mu_1,A)}(a|\omega)\mu_1(\omega)}{P(a)} = \frac{P_{(Q_2,q_2|\mu_2,A)}(a|\omega)\mu_2(\omega)}{P(a)} = \gamma_2^a(\omega) \tag{4.64}$$

for all $\omega \in \bar{\Omega}_1 \backslash \omega'$ where γ_1^a is the posterior such that $q_1(a|\gamma) > 0$ and γ_2^a is the posterior such that $q_2(a|\gamma) > 0$. Note that since $J \geq 2$, $\bar{\Omega}_1 \backslash \omega'$ is not empty.

Proposition 2 in Caplin, Dean and Leahy [2019] shows that if $P_{(Q_1,q_1|\mu_1,A)}(a) > 0$ and $P_{(Q_1,q_1|\mu_1,A)}(b) > 0$, then

$$\frac{\exp(a(\omega)/\kappa^J)}{\exp(b(\omega)/\kappa^J)} = \frac{\gamma_1^a(\omega)}{\gamma_1^b(\omega)} \tag{4.65}$$

for all $\omega \in \bar{\Omega}_1$. Here κ^J is the parameter associated with decision problems of cardinality J. Similarly,

$$\frac{\exp(a(\omega)/\kappa^{J+1})}{\exp(b(\omega)/\kappa^{J+1})} = \frac{\gamma_2^a(\omega)}{\gamma_2^b(\omega)} \tag{4.66}$$

for all $\omega \in \bar{\Omega}_2$.

Combining (4.64), (4.65), and (4.66) yields:

$$\frac{\exp(a(\omega)/\kappa^J)}{\exp(b(\omega)/\kappa^{J_1})} = \frac{\exp(a(\omega)/\kappa^{J+1})}{\exp(b(\omega)/\kappa^{J+1})} \tag{4.67}$$

for all $\omega \in \bar{\Omega}_1 \backslash \omega'$. As the original decision problem (μ_1, A) was arbitrary, this equation holds for arbitrary $a(\omega)$ and $b(\omega)$ and, in particular, for $a(\omega) \neq b(\omega)$. Rearrainging, (4.67) when $a(\omega) \neq b(\omega)$, yields:

$$\kappa^{J_1} = \kappa^{J+1}$$

Hence κ is equal for all decision problems with cardinality $J \geq 2$. This completes the proof.

References

[1] Amari, Shun-ichi (2016), *Information Geometry and Its Applications*," Springer Japan.

[2] Caplin, Andrew, Mark Dean, and John Leahy (2017), "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," NBER Working Paper No. 23652.

[3] Caplin, Andrew, Mark Dean, and John Leahy (2019), "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies* 86, 1061-1094.

[4] Hébert, Benjamin, and Jennifer La'O (2020), "Information acquisition, Efficiency, and Non-Fundamental Volatility," NBER Working Paper No. 26771.

[5] Jiao, Jiantao, Thomas Courtade, Albert No, Kartik Venkat, and Tsachy Weissman (2014), "Information Measures: the Curious Case of the Binary Alphabet," *IEEE Transactions on Information Theory* 60, 7616–7626.

[6] Painsky, Amichai, and Gregory Wornell (2020), "Bregman Divergence Bounds and Universality Properties of the Logarithmic Loss," *IEEE Transactions on Information Theory* 66, 1658-1673.

[7] Rockafellar, R. Tyrrell (1970), *Convex Analysis,* Princeton: Princeton University Press.

[8] Rudin, Walter (1976), *Principles of Mathematical Analysis*, 3rd Edition, New York: McGraw-Hill.