Online Appendix

The Cost of Information: The Case of Constant Marginal Costs

Luciano Pomatto, Philipp Strack and Omer Tamuz

A. Discussion of the Continuity Axiom

Our continuity axiom may seem technical, and in a sense it is. However, there are some interesting technical subtleties involved with its choice. Indeed, it seems that a more natural choice of topology would be the topology of *weak convergence* of likelihood ratios. Under that topology, two experiments would be close if they had close expected utilities for decision problems with continuous bounded utilities. The disadvantage of this topology is that *no cost* that satisfies the rest of the axioms is continuous in this topology. To see this, consider the sequence of experiments in which a coin (whose bias depends on the state) is tossed n times with probability $1/n$, and otherwise is not tossed at all. Under our axioms these experiments all have the same cost—the cost of tossing the coin once. However, in the weak topology these experiments converge to the trivial experiment that yields no information and therefore has zero cost.

In fact, even the stronger *total variation* topology suffers from the same problem, which is demonstrated using the same sequence of experiments. Therefore, one must consider a *finer* topology (which makes for a weaker continuity assumption), which we do by also requiring the first N moments to converge. Note that increasing N makes for a finer topology and therefore a weaker continuity assumption, and that our results hold for all $N > 0$. An even stronger topology (which requires the convergence of all moments) is used by Mattner (1999, 2004) to characterize all continuous additive linear functionals on the space of all random variables on $\mathbb{R}$.

Nevertheless, the continuity axiom is technical. As we show in Theorem 5 it is not required when there are only two states, and we conjecture that it is not required in general.

B. Preliminaries

To simplify the notation, throughout the appendix we set $\Theta = \{0, 1, \ldots, n\}$.

B.1. Properties of the Kullback-Leibler Divergence

In this section we summarize some well known properties of the Kullback-Leibler divergence, and derive from them straightforward properties of the LLR cost.

Given a measurable space (X, Σ) we denote by $\mathcal{P}(X, \Sigma)$ the space of probability measures on (X, Σ). If $X = \mathbb{R}^d$ for some $d \in \mathbb{N}$ then Σ is implicitly assumed to be the corresponding Borel σ-algebra and we simply write $\mathcal{P}(\mathbb{R}^d)$.

For the next result, given two measurable spaces (Ω, Σ) and (Ω', Σ'), a measurable map $F\colon \Omega \to \Omega'$, and a measure $\eta \in \mathcal{P}(\Omega, \Sigma)$, we define the *push-forward* measure $F_*\eta \in \mathcal{P}(\Omega', \Sigma')$ by $[F_*\eta](A) = \eta(F^{-1}(A))$ for all $A \in \Sigma'$.

PROPOSITION 9: *Let $\nu_1, \nu_2, \eta_1, \eta_2$ be measures in $\mathcal{P}(\Omega, \Sigma)$, and let μ_1, μ_2 be probability measures in $\mathcal{P}(\Omega', \Sigma')$. Assume that $D_{\mathrm{KL}}(\nu_1\|\nu_2)$, $D_{\mathrm{KL}}(\eta_1\|\eta_2)$ and $D_{\mathrm{KL}}(\mu_1\|\mu_2)$ are all finite. Let $F\colon \Omega \to \Omega'$ be measurable. Then:*

1) *$D_{\mathrm{KL}}(\nu_1\|\nu_2) \geq 0$ with equality if and only if $\nu_1 = \nu_2$.*
2) *$D_{\mathrm{KL}}(\nu_1 \times \mu_1\|\nu_2 \times \mu_2) = D_{\mathrm{KL}}(\nu_1\|\nu_2) + D_{\mathrm{KL}}(\mu_1\|\mu_2)$.*
3) *For all $\alpha \in (0,1)$,*

$$D_{\mathrm{KL}}(\alpha\nu_1 + (1-\alpha)\eta_1\|\alpha\nu_2 + (1-\alpha)\eta_2) \leq \alpha D_{\mathrm{KL}}(\nu_1\|\nu_2) + (1-\alpha) D_{\mathrm{KL}}(\eta_1\|\eta_2).$$

4) *$D_{\mathrm{KL}}(F_*\nu_1\|F_*\mu_1) \leq D_{\mathrm{KL}}(\nu_1\|\mu_1)$.*

It is well known that KL-divergence satisfies the first three properties in the statement of the proposition. We refer the reader to (Austin, 2006, Proposition 2.4) for a proof of the last property.

LEMMA 1: *Two experiments $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$ that satisfy $\bar{\mu}_i = \bar{\nu}_i$ for every $i \in \Theta$ are equivalent in the Blackwell order.*

PROOF:

The result is standard, but we include a proof for completeness. Suppose $\bar{\mu}_i = \bar{\nu}_i$ for every $i \in \Theta$. Given the experiment μ and a uniform prior on Θ, the posterior probability of state i conditional on s is given almost surely by

$$p_i(s) = \frac{\mathrm{d}\mu_i}{\mathrm{d}\sum_{j\in\Theta}\mu_j}(s) = \frac{1}{\sum_{j\in\Theta}\frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_i}(s)} = \frac{1}{\sum_{j\in\Theta}\mathrm{e}^{\ell_{ji}(s)}} \tag{20}$$

and the corresponding expression applies to experiment ν. By assumption, conditional on each state the two experiments induce the same distribution of log-likelihood ratios (ℓ_{ij}). Hence, by (20) they must induce the same distribution over posteriors, hence be equivalent in the Blackwell order.

A consequence of Proposition 9 is that the LLR cost is monotone with respect to the Blackwell order:

PROOF OF PROPOSITION 1:

Let C be an LLR cost. It is immediate that if $\bar{\mu}_i = \bar{\nu}_i$ for every i then $C(\mu) = C(\nu)$. We can assume without loss of generality that $S = T = \mathcal{P}(\Theta)$, endowed with the Borel σ-algebra. This follows from the fact that we can define a new experiment $\rho = (\mathcal{P}(\Theta), (\rho_i))$ such that $\bar{\mu}_i = \bar{\rho}_i$ for every i (see, e.g. Le Cam (1996)), and apply the same result to ν . By Blackwell's Theorem there exists a probability space (R, λ) and a "garbling" map $G\colon S \times R \to T$ such that for each $i \in \Theta$ it holds that $\nu_i = G_*(\mu_i \times \lambda)$. Hence, by the first, second and fourth statements in Proposition 9,

$$\begin{aligned} D_{\mathrm{KL}}(\nu_i \| \nu_j) &= D_{\mathrm{KL}}(G_*(\mu_i \times \lambda) \| G_*(\mu_j \times \lambda)) \\ &\leq D_{\mathrm{KL}}(\mu_i \times \lambda \| \mu_j \times \lambda) \\ &= D_{\mathrm{KL}}(\mu_i \| \mu_j) + D_{\mathrm{KL}}(\lambda \| \lambda) \\ &= D_{\mathrm{KL}}(\mu_i \| \mu_j). \end{aligned}$$

Therefore, by Theorem 1, we have

$$C(\nu) = \sum_{i,j} \beta_{ij} D_{\mathrm{KL}}(\nu_i \| \nu_j) \leq \sum_{i,j} \beta_{ij} D_{\mathrm{KL}}(\mu_i \| \mu_j) = C(\mu).$$

We note that a similar argument shows that if all the coefficients β_{ij} are positive then $C(\mu) > C(\nu)$ whenever μ Blackwell dominates ν but ν does not dominate μ.

An additional direct consequence of Proposition 9 is that the LLR cost is convex:

PROPOSITION 10: *Let $\mu = (S, (\mu_i))$ and $\nu = (S, (\nu_i))$ be experiments in $\mathcal{E}$. Given $\alpha \in (0,1)$, define the experiment $\eta = (S, (\nu_i))$ as $\eta_i = \alpha\nu_i + (1-\alpha)\mu_i$ for each i. Then any LLR cost C satisfies*

$$C(\eta) \leq \alpha C(\nu) + (1-\alpha) C(\mu).$$

The result follows immediately from the third statement in Proposition 9. We now study the set

$$\mathcal{D} = \{(D_{\mathrm{KL}}(\mu_i \| \mu_j))_{i \neq j} : \mu \in \mathcal{E}\} \subseteq \mathbb{R}_+^{(n+1)n}$$

of all possible pairs of expected log-likelihood ratios induced by some experiment μ. The next result shows that $\mathcal{D}$ contains the strictly positive orthant.

LEMMA 2: $\mathbb{R}_{++}^{(n+1)n} \subseteq \mathcal{D}$

PROOF:

The set $\mathcal{D}$ is convex. To see this, let $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$ be two experiments. Without loss of generality, we can suppose that $S = T$, and $S = S_1 \cup S_2$, where S_1, S_2 are disjoint, and $\mu_i(S_1) = \nu_i(S_2) = 1$ for every i.

Fix $\alpha \in (0,1)$ and define the new experiment $\tau = (S, (\tau_i))$ where $\tau_i = \alpha\mu_i + (1-\alpha)\nu_i$ for every i. It can be verified that τ_i-almost surely, $\frac{\mathrm{d}\tau_i}{\mathrm{d}\tau_j}$ satisfies $\frac{\mathrm{d}\tau_i}{\mathrm{d}\tau_j}(s) = \frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)$ if $s \in S_1$ and $\frac{\mathrm{d}\tau_i}{\mathrm{d}\tau_j}(s) = \frac{\mathrm{d}\nu_i}{\mathrm{d}\nu_j}(s)$ if $s \in S_2$. It then follows that

$$D_{\mathrm{KL}}(\tau_i \| \tau_j) = \alpha D_{\mathrm{KL}}(\mu_i \| \mu_j) + (1-\alpha) D_{\mathrm{KL}}(\nu_i \| \nu_j).$$

Hence $\mathcal{D}$ is convex. We now show $\mathcal{D}$ is a convex cone. First notice that the zero vector belongs to $\mathcal{D}$, since it corresponds to the totally uninformative experiment.

In addition (see §B.B.1),

$$D_{\mathrm{KL}}((\mu\otimes\mu)_i\|(\mu\otimes\mu)_j) = D_{\mathrm{KL}}(\mu_i\times\mu_i\|\mu_j\times\mu_j) = 2D_{\mathrm{KL}}(\mu_i\|\mu_j)$$

Hence $\mathcal{D}$ is closed under addition. Because $\mathcal{D}$ is also convex and contains the zero vector, it is a convex cone.

Suppose, by way of contradiction, that the inclusion $\mathbb{R}^{(n+1)n}_{++} \subseteq \mathcal{D}$ does not hold. This implies we can find a vector $z \in \mathbb{R}^{(n+1)n}_{+}$ that does not belong to the closure of $\mathcal{D}$. Therefore, there exists a nonzero vector $w \in \mathbb{R}^{(n+1)n}$ and $t \in \mathbb{R}$ such that $w\cdot z > t \geq w\cdot y$ for all $y \in \mathcal{D}$. Because $\mathcal{D}$ is a cone, then $t \geq 0$ and $0 \geq w\cdot y$ for all $y \in \mathcal{D}$. Hence, there must exist a coordinate i_oj_o such that $w_{i_oj_o} > 0$. We now show this leads to a contradiction.

Consider the following three cumulative distribution functions on $[2,\infty)$:

$$F_1(x) = 1 - \frac{2}{x}$$
$$F_2(x) = 1 - \frac{\log^2 2}{\log^2 x}$$
$$F_3(x) = 1 - \frac{\log 2}{\log x},$$

and denote by π_1, π_2, π_3 the corresponding measures. A simple calculation shows that $D_{\mathrm{KL}}(\pi_3\|\pi_1) = \infty$, whereas $D_{\mathrm{KL}}(\pi_a\|\pi_b) < \infty$ for any other choice of $a, b \in \{1,2,3\}$.

Let $\pi_a^\varepsilon = (1-\varepsilon)\,\delta_2 + \varepsilon\pi_a$ for every $a \in \{1,2,3\}$, where δ_2 is the point mass at 2. Then still $D_{\mathrm{KL}}(\pi_3^\varepsilon\|\pi_1^\varepsilon) = \infty$, but, for any other choice of a and b in $\{1,2,3\}$, the divergence $D(\pi_a^\varepsilon\|\pi_b^\varepsilon)$ vanishes as ε goes to zero. Let $\pi_a^{\varepsilon,M}$ be the measure π_a^ε conditioned on the interval $[2,M]$. Then $D_{\mathrm{KL}}(\pi_a^{\varepsilon,M}\|\pi_b^{\varepsilon,M})$ tends to $D_{\mathrm{KL}}(\pi_a^\varepsilon\|\pi_b^\varepsilon)$ as M tends to infinity, for any a, b. It follows that for every $N \in \mathbb{N}$ there exist ε small enough and M large enough such that $D_{\mathrm{KL}}(\pi_3^{\varepsilon,M}\|\pi_1^{\varepsilon,M}) > N$ and, for any other choice of a, b, $D_{\mathrm{KL}}(\pi_a^{\varepsilon,M}\|\pi_b^{\varepsilon,M}) < 1/N$.

Consider the experiment $\mu = (\mathbb{R}, (\mu_i))$ where $\mu_{i_0} = \pi_3^{\varepsilon,M}$, $\mu_{j_0} = \pi_1^{\varepsilon,M}$ and $\mu_k = \pi_2^{\varepsilon,M}$ for all $k \notin \{i_0, j_0\}$ and with ε and M so that the inequalities above hold for N large enough. Then $\mu \in \mathcal{E}$ since all measures have bounded support. It satisfies $D_{\mathrm{KL}}(\mu_{i_o}\|\mu_{j_o}) > N$ and $D_{\mathrm{KL}}(\mu_i\|\mu_j) < 1/N$ for every other pair ij.

Now let $y \in \mathcal{D}$ be the vector defined by μ. Then $w\cdot y > 0$ for N large enough. A contradiction.

B.2. Experiments and Log-likelihood Ratios

It will be convenient to consider, for each experiment, the distribution over log-likelihood ratios with respect to the state $i = 0$ conditional on a state j. Given an experiment, we define $\ell_i = \ell_{i0}$ for every $i \in \Theta$. We say that a vector

$\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_n) \in \mathcal{P}(\mathbb{R}^n)^{n+1}$ of measures is *derived from the experiment* $(S, (\mu_i))$ if for every $i = 0, 1, \ldots, n$,

$$\sigma_i(E) = \mu_i\left(\{s : (\ell_1(s), \ldots, \ell_n(s)) \in E\}\right) \text{ for all measurable } E \subseteq \mathbb{R}^n.$$

That is, σ_i is the distribution of the vector $(\ell_1, \ldots, \ell_n)$ of log-likelihood ratios (with respect to state 0) conditional on state i. There is a one-to-one relation between the vector σ and the collection $(\bar{\mu}_i)$ of distributions defined in the main text: notice that $\ell_{ij} = \ell_{i0} - \ell_{j0}$ almost surely, hence knowing the distribution of $(\ell_{0i})_{i\in\Theta}$ is enough to recover the distribution of $(\ell_{ij})_{i,j\in\Theta}$. Nevertheless, working directly with σ (rather than $(\bar{\mu}_i)$) will simplify the notation considerably.

We call a vector $\sigma \in \mathcal{P}(\mathbb{R}^n)^{n+1}$ *admissible* if it is derived from some experiment. The next result provides a straightforward characterization of admissible vectors of measures.

LEMMA 3: *A vector of measures $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_n) \in \mathcal{P}(\mathbb{R}^n)^{n+1}$ is admissible if and only if the measures are mutually absolutely continuous and, for every i, satisfy $\frac{\mathrm{d}\sigma_i}{\mathrm{d}\sigma_0}(\xi) = e^{\xi_i}$ for σ_i-almost every $\xi \in \mathbb{R}^n$.*

PROOF:

If $(\sigma_0, \sigma_1, \ldots, \sigma_n)$ is admissible then there exists an experiment $\mu = (S, (\mu_i))$ such that for any measurable $E \subseteq \mathbb{R}^n$

$$\begin{aligned}\int_E e^{\xi_i}\,\mathrm{d}\sigma_0(\xi) &= \int 1_E\left((\ell_1(s), \ldots \ell_n(s))\right) e^{\ell_i(s)}\,\mathrm{d}\mu_0(s) \\ &= \int 1_E\left((\ell_1(s), \ldots \ell_n(s))\right) \mathrm{d}\mu_i(s)\end{aligned}$$

where 1_E is the indicator function of E. So, $\int_E e^{\xi_i}\,\mathrm{d}\sigma_0(\xi) = \sigma_i(E)$ for every $E \subseteq \mathbb{R}^n$. Hence e^{ξ_i} is a version of $\frac{\mathrm{d}\sigma_i}{\mathrm{d}\sigma_0}$.

Conversely, assume $\frac{\mathrm{d}\sigma_i}{\mathrm{d}\sigma_0}(\xi) = e^{\xi_i}$ for almost every $\xi \in \mathbb{R}^n$. Define an experiment $(\mathbb{R}^{n+1}, (\mu_i))$ where $\mu_i = \sigma_i$ for every i. The experiment $(\mathbb{R}^{n+1}, (\mu_i))$ is such that $\ell_i(\xi) = \xi_i$ for every $i > 0$. Hence, for $i > 0$, $\mu_i\left(\{\xi : (\ell_1(\xi), \ldots, \ell_n(\xi)) \in E\}\right)$ is equal to

$$\int 1_E\left((\ell_1(\xi), \ldots \ell_n(\xi))\right) e^{\xi_i}\,\mathrm{d}\sigma_0(\xi) = \int 1_E(\xi) e^{\xi_i}\,\mathrm{d}\sigma_0(\xi) = \sigma_i(E)$$

and similarly $\mu_0\left(\{\xi : (\ell_1(\xi), \ldots, \ell_n(\xi)) \in E\}\right) = \sigma_0(E)$. So $(\sigma_0, \sigma_1, \ldots, \sigma_n)$ is admissible.

B.3. Properties of Cumulants

The purpose of this section is to formally describe cumulants and their relation to moments. We follow Leonov and Shiryaev (1959) and Shiryaev (1996, p. 289).

Given a vector $\xi \in \mathbb{R}^n$ and an integral vector $\alpha \in \mathbb{N}^n$ we write $\xi^\alpha = \xi_1^{\alpha_1}\xi_2^{\alpha_2}\cdots\xi_n^{\alpha_n}$ and use the notational conventions $\alpha! = \alpha_1!\alpha_2!\cdots\alpha_n!$ and $|\alpha| = \alpha_1 + \cdots + \alpha_n$.

Let $A = \{0,\ldots,N\}^n \backslash \{0,\ldots,0\}$, for some constant $N \in \mathbb{N}$ greater or equal than 1. For every probability measure $\sigma_1 \in \mathcal{P}(\mathbb{R}^n)$ and $\xi \in \mathbb{R}^n$, let $\varphi_{\sigma_1}(\xi) = \int_{\mathbb{R}^n} e^{i\langle z,\xi\rangle}\,\mathrm{d}\sigma_1(z)$ denote the characteristic function of σ_1 evaluated at ξ. We denote by $\mathcal{P}_A \subseteq \mathcal{P}(\mathbb{R}^n)$ the subset of measures σ_1 such that $\int_{\mathbb{R}^n} |\xi^\alpha|\,\mathrm{d}\sigma_1(\xi) < \infty$ for every $\alpha \in A$. Every $\sigma_1 \in \mathcal{P}_A$ is such that in a neighborhood of $\mathbf{0} \in \mathbb{R}^n$ the cumulant generating function $\log\varphi_{\sigma_1}$ is well defined and the partial derivatives

$$\frac{\partial^{|\alpha|}}{\partial\xi_1^{\alpha_1}\partial\xi_2^{\alpha_2}\cdots\partial\xi_n^{\alpha_n}}\log\varphi_{\sigma_1}(\xi)$$

exist and are continuous for every $\alpha \in A$.

For every $\sigma_1 \in \mathcal{P}_A$ and $\alpha \in A$ let $\kappa_{\sigma_1}(\alpha)$ be defined as

$$\kappa_{\sigma_1}(\alpha) = i^{-|\alpha|}\frac{\partial^{|\alpha|}}{\partial\xi_1^{\alpha_1}\partial\xi_2^{\alpha_2}\cdots\partial\xi_n^{\alpha_n}}\log\varphi_{\sigma_1}(\mathbf{0})$$

With slight abuse of terminology, we refer to $\kappa_{\sigma_1} \in \mathbb{R}^A$ as the *vector of cumulants* of σ_1. In addition, for every $\sigma_1 \in \mathcal{P}_A$ and $\alpha \in A$ we denote by $m_{\sigma_1}(\alpha) = \int_{\mathbb{R}^n} \xi^\alpha\,\mathrm{d}\sigma_1(\xi)$ the mixed moment of σ_1 of order α and refer to $m_{\sigma_1} \in \mathbb{R}^A$ as the *vector of moments* of σ_1.

Given two measures $\sigma_1, \sigma_2 \in \mathcal{P}(\mathbb{R}^n)$ we denote by $\sigma_1 * \sigma_2 \in \mathcal{P}(\mathbb{R}^n)$ the corresponding convolution.

LEMMA 4: *For every $\sigma_1, \sigma_2 \in \mathcal{P}_A$, and $\alpha \in A$, $\kappa_{\sigma_1 * \sigma_2}(\alpha) = \kappa_{\sigma_1}(\alpha) + \kappa_{\sigma_2}(\alpha)$.*

PROOF:

The result follows from the well known fact that $\varphi_{\sigma_1 * \sigma_2}(\xi) = \varphi_{\sigma_1}(\xi)\varphi_{\sigma_2}(\xi)$ for every $\xi \in \mathbb{R}^n$.

The next result, due to Leonov and Shiryaev (1959) (see also Shiryaev, 1996, p. 290) establishes a one-to-one relation between the vector of moments m_{σ_1} and vector of cumulants κ_{σ_1} of a probability measure $\sigma_1 \in \mathcal{P}_A$. Given $\alpha \in A$, let $\Lambda(\alpha)$ be the set of all ordered collections $\left(\lambda^1,\ldots,\lambda^q\right)$ of non-zero vectors in $\mathbb{N}^n$ such that $\sum_{p=1}^q \lambda^p = \alpha$.

THEOREM 2: *For every $\sigma_1 \in \mathcal{P}_A$ and $\alpha \in A$,*

1) $m_{\sigma_1}(\alpha) = \sum_{(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)} \frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}\prod_{p=1}^q \kappa_{\sigma_1}(\lambda^p)$

2) $\kappa_{\sigma_1}(\alpha) = \sum_{(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)} \frac{(-1)^{q-1}}{q}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}\prod_{p=1}^q m_{\sigma_1}(\lambda^p)$

The result yields the following implication. Let $M_A = \{m_{\sigma_1} : \sigma_1 \in \mathcal{P}_A\} \subseteq \mathbb{R}^A$ and $K_A = \{\kappa_{\sigma_1} : \sigma_1 \in \mathcal{P}_A\} \subseteq \mathbb{R}^A$. Statement 2 in Theorem 2 shows the existence of a continuous function $h\colon M_A \to K_A$ such that $\kappa_{\sigma_1} = h(m_{\sigma_1})$ for every $\sigma_1 \in \mathcal{P}_A$. Moreover, statement 1 implies h is one-to-one.

B.4. Cumulants and Admissible Measures

We denote by $\mathcal{A}$ the set of vectors of measures $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_n)$ that are admissible and such that $\sigma_i \in \mathcal{P}_A$ for every i. To each $\sigma \in \mathcal{A}$ we associate the vector

$$m_\sigma = (m_{\sigma_0}, m_{\sigma_1}, \ldots, m_{\sigma_n}) \in \mathbb{R}^d$$

of dimension $d = (n+1)\,|A|$. Similarly, we define

$$\kappa_\sigma = (\kappa_{\sigma_0}, \kappa_{\sigma_1}, \ldots, \kappa_{\sigma_n}) \in \mathbb{R}^d.$$

In this section we study properties of the sets $\mathcal{M} = \{m_\sigma : \sigma \in \mathcal{A}\}$ and $\mathcal{K} = \{\kappa_\sigma : \sigma \in \mathcal{A}\}$.

LEMMA 5: *Let I and J be disjoint finite sets and let $(\phi_k)_{k \in I \cup J}$ be a collection of real valued functions defined on $\mathbb{R}^n$. Assume $\{\phi_k : k \in I \cup J\} \cup \{1_{\mathbb{R}^n}\}$ are linearly independent and the unit vector $(1, \ldots, 1) \in \mathbb{R}^J$ belongs to the interior of $\left\{(\phi_k(\xi))_{k \in J} : \xi \in \mathbb{R}^n\right\}$. Then*

$$C = \left\{ \left(\int_{\mathbb{R}^n} \phi_k \, \mathrm{d}\sigma_1 \right)_{k \in I} : \sigma_1 \in \mathcal{P}(\mathbb{R}^n) \text{ has finite support and } \int_{\mathbb{R}^n} \phi_k \, \mathrm{d}\sigma_1 = 1 \text{ for all } k \in J \right\}$$

is a convex subset of $\mathbb{R}^I$ with nonempty interior.

PROOF:
To ease the notation, let $Y = \mathbb{R}^n$ and denote by $\mathcal{P}_o$ be the set of probability measures on Y with finite support. Consider $F = \{\phi_k : k \in I \cup J\} \cup \{1_{\mathbb{R}^d}\}$ as a subset of the vector space $\mathbb{R}^Y$, where the latter is endowed with the topology of pointwise convergence. The topological dual of $\mathbb{R}^Y$ is the vector space of signed measures on Y with finite support. Let

$$D = \left\{ \left(\int_{\mathbb{R}^n} \phi_k \, \mathrm{d}\sigma_1 \right)_{k \in I \cup J} : \sigma_1 \in \mathcal{P}_o \right\} \subseteq \mathbb{R}^{I \cup J}.$$

Fix $k \in I \cup J$. Since ϕ_k does not belong to the linear space V generated by $F \backslash \{\phi_k\}$, then a standard application of the hyperplane separation theorem implies the existence of a signed measure

$$\rho = \alpha\sigma_1 - \beta\sigma_2$$

where $\alpha, \beta \geq 0$, $\alpha + \beta > 0$ and $\sigma_1, \sigma_2 \in \mathcal{P}_o$, such that ρ satisfies $\int \phi_k \, \mathrm{d}\rho > 0 \geq \int \phi \, \mathrm{d}\rho$ for every $\phi \in V$. This implies $\int \phi \, \mathrm{d}\rho = 0$ for every $\phi \in V$. By taking $\phi = 1_{\mathbb{R}^n}$, we obtain $\rho(\mathbb{R}^n) = 0$. Hence, $\alpha = \beta$. Therefore, $\int \phi_k \, \mathrm{d}\sigma_1 > \int \phi_k \, \mathrm{d}\sigma_2$ and $\int \phi_l \, \mathrm{d}\sigma_1 = \int \phi_l \, \mathrm{d}\sigma_2$ for every $l \neq k$. To summarize, we have shown that for every $k \in I \cup J$ there exist vectors $w^k, z^k \in D$ such that $w^k_k > z^k_k$ and $w^k_l = z^k_l$ for $l \neq k$.

Now let $\text{aff}(D)$ be the affine hull of D. As is well known, for every $d \in D$ we have the identity $\text{aff}(D) = d + \text{span}(D-d)$, where $\text{span}(D-d)$ is the vector space generated by $D-d$. Moreover, $\text{span}(D-d)$ is independent of the choice of $d \in D$ (see, for example, Jonathan M Borwein and Jon D Vanderwerff, 2010, Lemma 2.4.5).

Let $k \in I \cup J$ and let $1_k \in \mathbb{R}^{I \cup J}$ be the corresponding unit vector. By taking $d = z^k$ we obtain that $w^k - z^k \in \text{span}(D - z^k)$. Thus, $1_k \in \text{span}(D-d)$ for every k. Hence $\text{span}(D-d) = \mathbb{R}^{I \cup J}$. Therefore $\text{aff}(D) = \mathbb{R}^{I \cup J}$. Since D is convex, it has nonempty relative interior as a subset of $\text{aff}(D)$. We conclude that D has nonempty interior.

Now consider the hyperplane

$$H = \{z \in \mathbb{R}^{I \cup J} : z_k = 1 \text{ for all } k \in J\}$$

Let D^o be the interior of D. It remains to show that the hyperplane H satisfies $H \cap D^o \neq \emptyset$. This will imply that the projection of $H \cap D$ on $\mathbb{R}^I$, which equals C, has nonempty interior.

Let $w \in D^o$. By assumption, $(1, \ldots, 1) \in \mathbb{R}^J$ is in the interior of $\{(\phi_k(\xi))_{k \in J} : \xi \in Y\}$. Hence, there exists $\alpha \in (0,1)$ small enough and $\xi \in Y$ such that $\phi_k(\xi) = \frac{1}{1-\alpha} - \frac{\alpha}{1-\alpha} w_k$ for every $k \in J$. Define $z = \alpha w + (1-\alpha)(\phi_k(\xi))_{k \in I \cup J} \in D$. Then $z_k = 1$ for every $k \in J$. In addition, because $w \in D^o$ then $z \in D^o$ as well. Hence $z \in H \cap D^o$.

LEMMA 6: *The set $\mathcal{M} = \{m_\sigma : \sigma \in \mathcal{A}\}$ has nonempty interior.*

PROOF:

For every $\alpha \in A$ define the functions $(\phi_{i,\alpha})_{i \in \Theta}$ on $\mathbb{R}^n$ as

$$\phi_{0,\alpha}(\xi) = \xi^\alpha \text{ and } \phi_{i,\alpha}(\xi) = \xi^\alpha e^{\xi_i} \text{ for all } i > 0.$$

Define $\psi_0 = 1_{\mathbb{R}^n}$ and $\psi_i(\xi) = e^{\xi_i}$ for all $i > 0$. It is immediate to verify that

$$\{\phi_{i,\alpha} : i \in \Theta, \alpha \in A\} \cup \{\psi_i : i \in \Theta\}$$

is a linearly independent set of functions. In addition, $(1, \ldots, 1) \in \mathbb{R}^n$ is in the interior of $\{(e^{\xi_1}, \ldots, e^{\xi_n}) : \xi \in \mathbb{R}^n\}$. Lemma 5 implies that the set

$$C = \left\{ \left(\int_{\mathbb{R}^n} \phi_{i,\alpha} \, d\sigma_0 \right)_{\substack{i \in \Theta \\ \alpha \in A}} : \sigma_0 \in \mathcal{P}(\mathbb{R}^n) \text{ has finite support and } \int_{\mathbb{R}^n} e^{\xi_i} \, d\sigma_0(\xi) = 1 \text{ for all } i \right\}$$

has nonempty interior. Given σ_0 as in the definition of C, construct a vector $\sigma = (\sigma_0, \sigma_1, \ldots, \sigma_n)$ where for each $i > 0$ the measure σ_i is defined so that $(d\sigma_i / d\sigma_0)(\xi) = e^{\xi_i}$, σ_0-almost surely. Then, Lemma 3 implies σ is admissible.

Because each σ_i has finite support then $\sigma \in \mathcal{A}$. In addition,

$$m_\sigma = \left(\int_{\mathbb{R}^n} \phi_{i,\alpha} \, \mathrm{d}\sigma_0\right)_{\substack{i\in\Theta\\ \alpha\in A}}$$

hence $C \subseteq \mathcal{M}$. Thus, $\mathcal{M}$ has nonempty interior.

THEOREM 3: *The set $\mathcal{K} = \{\kappa_\sigma : \sigma \in \mathcal{A}\}$ has nonempty interior.*

PROOF:

Let $h : M_A \to K_A$ be the function defined in the discussion following Theorem 2, mapping vectors of moments to vectors of cumulants. Define $H : \mathcal{M} \to \mathcal{K}$ as

$$H(m_\sigma) = (h(m_{\sigma_0}), h(m_{\sigma_1}), \dots, h(m_{\sigma_n})) = \kappa_\sigma$$

for every $\sigma = (\sigma_0, \sigma_1, \dots, \sigma_n) \in \mathcal{A}$. Since h is continuous and one-to-one then so is H. Lemma 6 shows there exists an open set $U \subseteq \mathbb{R}^d$ included in $\mathcal{M}$. Let H_U be the restriction of H on U. Then H_U satisfies all the assumptions of Brouwer's Invariance of Domain Theorem,[35] which implies that $H_U(U)$ is an open subset of $\mathbb{R}^d$. Since $H(\mathcal{M}) \subseteq \mathcal{K}$, it follows that $\mathcal{K}$ has nonempty interior.

C. Automatic Continuity in the Cauchy Problem for Subsemigroups of $\mathbb{R}^d$.

A *subsemigroup* of $\mathbb{R}^d$ is a subset $\mathcal{S} \subseteq \mathbb{R}^d$ that is closed under addition, so that $x + y \in \mathcal{S}$ for all $x, y \in \mathcal{S}$. We say that a map $F\colon \mathcal{S} \to \mathbb{R}_+$ is *additive* if $F(x + y) = F(x) + F(y)$ for all $x, y, x + y \in \mathcal{S}$. We say that F is *linear* if there exists $(a_1, \dots, a_d) \in \mathbb{R}^d$ such that $F(x) = F(x_1, \dots, x_d) = a_1x_1 + \dots + a_dx_d$ for all $x \in \mathcal{S}$.

We can now state the main result of this section:

THEOREM 4: *Let $\mathcal{S}$ be a subsemigroup of $\mathbb{R}^d$ with a nonempty interior. Then every additive function $F\colon \mathcal{S} \to \mathbb{R}_+$ is linear.*

Before proving the theorem we will establish a number of claims.

CLAIM 1: *Let $\mathcal{S}$ be a subsemigroup of $\mathbb{R}^d$ with a nonempty interior. Then there exists an open ball $B \subset \mathbb{R}^d$ such that $aB \subset \mathcal{S}$ for all real $a \geq 1$.*

PROOF:

Let B_0 be an open ball contained in $\mathcal{S}$, with center x_0 and radius r. Given a positive integer k, note that kB_0 is the ball of radius kr centered at kr_0, and that it is contained in $\mathcal{S}$, since $\mathcal{S}$ is a semigroup. Choose a positive integer $M \geq 4$ such that $\frac{2}{3}Mr > \|x_0\|$, and let B be the open ball with center at Mx_0 and radius r (see Figure 5). Fix any $a \geq 1$, and write $a = \frac{1}{M}(n + \gamma)$ for some integer $n \geq M$ and $\gamma \in [0, 1)$. Then $\frac{n}{M}B$ is the ball of radius $\frac{n}{M}r$ centered at nx_0, which is

[35] Brouwer (1911). See also Theorem 2 in Tao (2011).

contained in nB_0, since nB_0 also has center nx_0, but has a larger radius nr. So $\frac{n}{M}B \subset nB_0$. We claim that furthermore $\frac{n+1}{M}B$ is also contained in nB_0. To see this, observe that the center of $\frac{n+1}{M}B$ is $(n+1)x_0$ and its radius is $\frac{n+1}{M}r$. Hence the center of $\frac{n+1}{M}B$ is at distance $\|x_0\|$ from the center of nB_0, and so the furthest point in $\frac{n+1}{M}B$ is at distance $\|x_0\| + \frac{n+1}{M}r$ from the center of nB_0. But the radius of nB_0 is

$$nr = \frac{2}{3}nr + \frac{1}{3}nr \geq \frac{2}{3}Mr + \frac{1}{3}nr > \|x_0\| + \frac{n+1}{M}r,$$

where the first inequality follows since $n \geq M$, and the second since $\frac{2}{3}Mr > \|x_0\|$ and $M \geq 4$. So nB_0 indeed contains both $\frac{n}{M}B$ and $\frac{n+1}{M}B$. Thus it also contains aB, and so $\mathcal{S}$ contains aB.

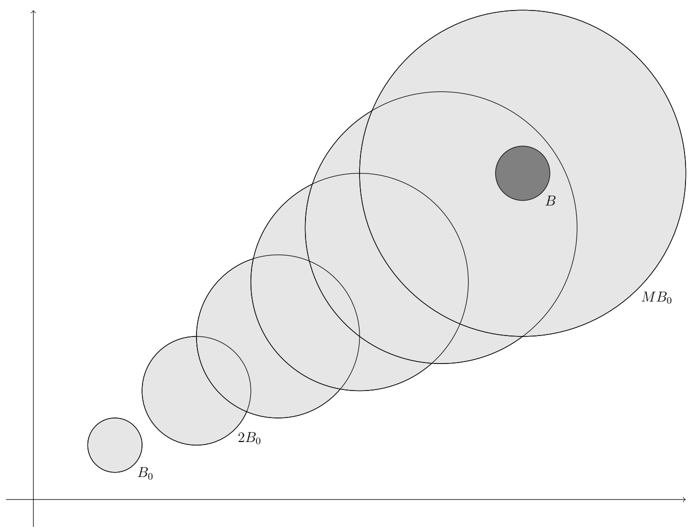

Figure 5. : Illustration of the proof of Claim 1. The dark ball B is contained in the light ones, and it is apparent from this image that so is any multiple of B by $a \geq 1$.

CLAIM 2: *Let $\mathcal{S}$ be a subsemigroup of $\mathbb{R}^d$ with a nonempty interior. Let $F\colon \mathcal{S} \to \mathbb{R}_+$ be additive and satisfy $F(ay) = aF(y)$ for every $y \in \mathcal{S}$ and $a \in \mathbb{R}_+$ such that $ay \in \mathcal{S}$. Then F is linear.*

PROOF:

If $\mathcal{S}$ does not include zero, then without loss of generality we add zero to it and set $F(0) = 0$. Let B be an open ball such that $aB \subset \mathcal{S}$ for all $a \geq 1$; the existence of such a ball is guaranteed by Claim 1. Choose a basis $\{b^1, \ldots, b^d\}$ of $\mathbb{R}^d$ that is a subset of B, and let $x = \beta_1 b^1 + \cdots + \beta_d b^d$ be an arbitrary element of $\mathcal{S}$. Let

$b = \max\{1/|\beta_i| : \beta_i \neq 0\}$, and let $a = \max\{1, b\}$. Then

$$F(ax) = F(a\beta_1 b^1 + \cdots + a\beta_d b^d).$$

Assume without loss of generality that for some $0 \leq k \leq d$ it holds that the first k coefficients β_i are non-negative, and the rest are negative. Then for $i \leq k$ it holds that $a\beta_i b^i \in \mathcal{S}$ and for $i > k$ it holds that $-a\beta_i b^i \in \mathcal{S}$; this follows from the defining property of the ball B, since each b^i is in B, and since $|a\beta_i| \geq 1$. Hence we can add $F(-a\beta_{k+1}b^{k+1} - \cdots - a\beta_d b^d)$ to both sides of the above displayed equation, and then by additivity,

$$\begin{aligned} &F(ax) + F(-a\beta_{k+1}b^{k+1} - \cdots - a\beta_d b^d) \\ &= F(a\beta_1 b^1 + \cdots + a\beta_d b^d) + F(-a\beta_{k+1}b^{k+1} - \cdots - a\beta_d b^d) \\ &= F(a\beta_1 b^1 + \cdots + a\beta_k b^k). \end{aligned}$$

Using additivity again yields

$$F(ax) + F(-a\beta_{k+1}b^{k+1}) + \cdots + F(-a\beta_d b^d) = F(a\beta_1 b^1) + \cdots + F(a\beta_k b^k).$$

Applying now the claim hypothesis that $F(ay) = aF(y)$ whenever $y, ay \in \mathcal{S}$ yields

$$aF(x) + (-a\beta_{k+1})F(b^{k+1}) + \cdots + (-a\beta_d)F(b^d) = a\beta_1 F(b^1) + \cdots + a\beta_k F(b^k).$$

Rearranging and dividing by a, we arrive at

$$F(x) = \beta_1 F(b^1) + \cdots + \beta_d F(b^d).$$

We can therefore extend F to a function that satisfies this on all of $\mathbb{R}^d$, which is then clearly linear.

CLAIM 3: *Let B be an open ball in $\mathbb{R}^d$, and let $\mathcal{B}$ be the semigroup given by $\cup_{a\geq 1} aB$. Then every additive $F\colon \mathcal{B} \to \mathbb{R}_+$ is linear.*

PROOF:

Fix any $x \in \mathcal{B}$, and assume $ax \in \mathcal{B}$ for some $a \in \mathbb{R}_+$. Since $\mathcal{B}$ is open, by Claim 2 it suffices to show that $F(ax) = aF(x)$. The defining property of $\mathcal{B}$ implies that the intersection of $\mathcal{B}$ and the ray $\{bx : b \geq 0\}$ is of the form $\{bx : b > a_0\}$ for some $a_0 \geq 0$. By the additive property of F, we have that $F(qx) = qF(x)$ for every rational $q > a_0$. Furthermore, if $b > b' > a_0$ then $n(b - b')x \in \mathcal{S}$ for n large enough. Hence

$$
\begin{aligned}
F(bx) &= \frac{1}{n}F(nbx) \\
&= \frac{1}{n}F\left(nb'x + (n(b-b')x)\right) \\
&= \frac{1}{n}F\left(nb'x\right) + \frac{1}{n}F\left(n(b-b')x\right) \\
&= F(b'x) + \frac{1}{n}F\left(n(b-b')x\right) \\
&\geq F(b'x).
\end{aligned}
$$

Thus the map $f\colon (a_0,\infty) \to \mathbb{R}^+$ given by $f(b) = F(bx)$ is monotone increasing, and its restriction to the rationals is linear. So f must be linear, and hence $F(ax) = aF(x)$.

Given these claims, we are ready to prove our theorem. PROOF OF THEOREM 4.:

Fix any $x \in \mathcal{S}$, and assume $ax \in \mathcal{S}$ for some $a \in \mathbb{R}_+$. By Claim 2 it suffices to show that $F(ax) = aF(x)$. Let B be a ball with the property described in Claim 1, and denote its center by x_0 and its radius by r. As in Claim 3, let $\mathcal{B}$ be the semigroup given by $\cup_{a\geq 1} aB$; note that $\mathcal{B} \subseteq \mathcal{S}$. Then there is some y such that $x+y, a(x+y), y, ay \in \mathcal{B}$; in fact, we can take $y = bx_0$ for $b = \max\{a, 1/a, |x|/r\}$ (see Figure 6). Then, on the one hand, by additivity,

$$F(ax+ay) = F(ax) + F(ay).$$

On the other hand, since $x+y, a(x+y), y, ay \in \mathcal{B}$,and since, by Claim 3, the restriction of F to $\mathcal{B}$ is linear, we have that

$$F(ax+ay) = F(a(x+y)) = aF(x+y) = aF(x) + aF(y) = aF(x) + F(ay),$$

thus

$$F(ax) + F(ay) = aF(x) + F(ay)$$

and so $F(ax) = aF(x)$.

D. Proof of Theorem 1

Throughout this section, we maintain the notation and terminology introduced in §B. It follows from the results in §B.B.1 that an LLR cost satisfies Axioms 1-4. For the rest of this section, we denote by C a cost function that satisfies the axioms. Let N be such that C is uniformly continuous with respect to the distance d_N. We use the same N to define the set $A = \{0,\ldots,N\}^n \backslash \{0,\ldots,0\}$ introduced in §B.B.3.

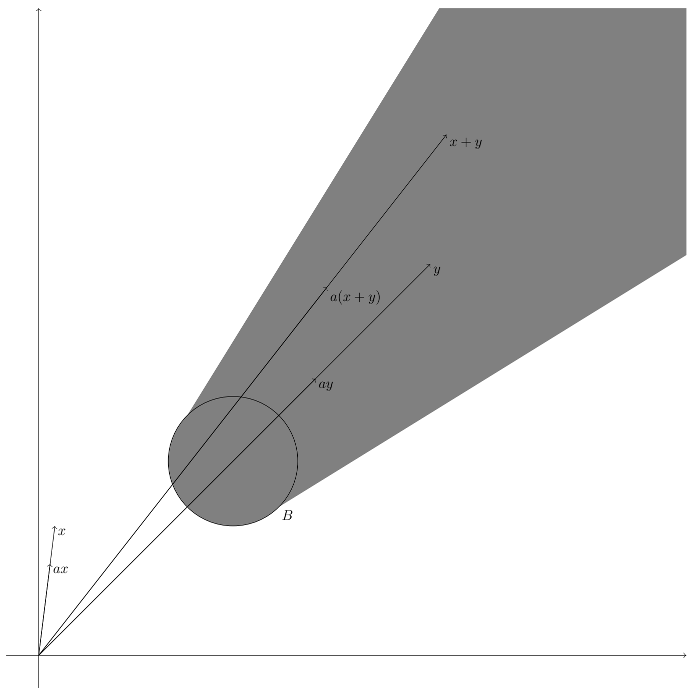

Figure 6. : An illustration of the proof of Theorem 4.

LEMMA 7: *Let μ and ν be two experiments that induce the same vector $\sigma \in \mathcal{A}$. Then $C(\mu) = C(\nu)$.*

PROOF:

Conditional on each $k \in \Theta$, the two experiments induce the same distribution for $(\ell_{i0})_{i\in\Theta}$. Because $\ell_{ij} = \ell_{i0} - \ell_{j0}$ almost surely, it follows that, conditional on each state, the two experiments induce the same distribution over the vector of all log-likelihood ratios $(\ell_{ij})_{i,j\in\Theta}$. Hence, $\bar{\mu}_i = \bar{\nu}_i$ for every i. Therefore, by Lemma 1 the two experiments are equivalent in the Blackwell order. The result now follows directly from Axiom 1.

Lemma 7 implies that we can define a function $c : \mathcal{A} \to \mathbb{R}_+$ as $c(\sigma) = C(\mu)$ where μ is an experiment inducing σ.

LEMMA 8: *Consider two experiments $\mu = (S, (\mu_i))$ and $\nu = (T, (\nu_i))$ that induce σ and τ in $\mathcal{A}$, respectively. Then*

1) *The experiment $\mu \otimes \nu$ induces the vector $(\sigma_0 * \tau_0, \dots, \sigma_n * \tau_n) \in \mathcal{A}$;*

2) *The experiment $\alpha \cdot \mu$ induces the measure $\alpha\sigma + (1-\alpha)\delta_{\mathbf{0}}$.*

PROOF:

(1) For every $E \subseteq \mathbb{R}^n$ and every state i,

$$\begin{aligned} & (\mu_i \times \nu_i)\left(\{(s,t) : (\ell_1(s,t), \ldots \ell_n(s,t)) \in E\}\right) \\ = \; & (\mu_i \times \nu_i)\left(\left\{(s,t) : \left(\log \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_0}(s) + \log \frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_0}(t), \ldots, \log \frac{\mathrm{d}\mu_n}{\mathrm{d}\mu_0}(s) + \log \frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_n}(t)\right) \in E\right\}\right) \\ = \; & (\sigma_i * \tau_i)(E) \end{aligned}$$

where the last equality follows from the definition of σ_i and τ_i. This concludes the proof of the claim.

(2) Immediate from the definition of $\alpha \cdot \mu$.

LEMMA 9: *The function $c : \mathcal{A} \to \mathbb{R}$ satisfies, for all $\sigma, \tau \in \mathcal{A}$ and $\alpha \in [0,1]$:*

1) $c(\sigma_0 * \tau_0, \ldots, \sigma_n * \tau_n) = c(\sigma) + c(\tau)$;

2) $c(\alpha\sigma + (1-\alpha)\delta_{\mathbf{0}}) = \alpha c(\sigma)$.

PROOF:

(1) Let $\mu \in \mathcal{E}$ induce σ and let $\nu \in \mathcal{E}$ induce τ. Then $C(\mu) = c(\sigma), C(\nu) = c(\tau)$ and, by Axiom 2 and Lemma 8, $c(\sigma_0 * \tau_0, \ldots, \sigma_n * \tau_n) = C(\mu \otimes \nu) = c(\sigma) + c(\tau)$. Claim (2) follows directly from Axiom 3 and Lemma 8.

LEMMA 10: *If $\sigma, \tau \in \mathcal{A}$ satisfy $m_\sigma = m_\tau$ then $c(\sigma) = c(\tau)$.*

PROOF:

Let μ be and ν be two experiments inducing σ and τ, respectively. Let $\mu^{\otimes r} = \mu \otimes \ldots \otimes \mu$ be the experiment obtained as the r-th fold independent product of μ. Axioms 2 and 3 imply

$$C((1/r) \cdot \mu^{\otimes r}) = C(\mu) \;\text{ and }\; C((1/r) \cdot \nu^{\otimes r}) = C(\nu)$$

In order to show that $C(\mu) = C(\nu)$ we now prove that $C((1/r) \cdot \mu^{\otimes r}) - C((1/r) \cdot \nu^{\otimes r}) \to 0$ as $r \to \infty$. To simplify the notation let, for every $r \in \mathbb{N}$,

$$\mu[r] = (1/r) \cdot \mu^{\otimes r} \;\text{ and }\; \nu[r] = (1/r) \cdot \nu^{\otimes r}$$

Let $\sigma[r] = (\sigma[r]_0, \ldots, \sigma[r]_n)$ and $\tau[r] = (\tau[r]_0, \ldots, \tau[r]_n)$ in $\mathcal{A}$ be the vectors of measures induced by $\mu[r]$ and $\nu[r]$.

We claim that $d_N(\mu[r], \nu[r]) \to 0$ as $r \to \infty$. First, notice that $\overline{\mu[r]}_i$ and $\overline{\nu[r]}_i$ assign probability $(r-1)/r$ to the zero vector $\mathbf{0} \in \mathbb{R}^{(n+1)^2}$. Hence

$$d_{tv}(\overline{\mu[r]}_i, \overline{\nu[r]}_i) = \sup_E \frac{1}{r} \left| \overline{\mu^{\otimes r}}_i(E) - \overline{\nu^{\otimes r}}_i(E) \right| \leq \frac{1}{r}.$$

For every $\alpha \in A$ we have

$$M_i^{\mu[r]}(\alpha) = \int \ell_{10}^{\alpha_1} \ldots \ell_{n0}^{\alpha_n} \,\mathrm{d}\mu[r]_i = \int_{\mathbb{R}^n} \xi_1^{\alpha_1} \cdots \xi_n^{\alpha_n} \,\mathrm{d}\sigma[r]_i(\xi) = m_{\sigma[r]_i}(\alpha) \tag{21}$$

We claim that $m_{\sigma[r]} = m_{\tau[r]}$. Theorem 2 shows the existence of a bijection $H : \mathcal{M} \to \mathcal{K}$ such that $H(m_\upsilon) = \kappa_\upsilon$ for every $\upsilon \in \mathcal{A}$. The experiment $\mu^{\otimes r}$ induces the vector $(\sigma_0^{*r}, \ldots, \sigma_n^{*r}) \in \mathcal{A}$, where σ_i^{*r} denotes the r-th fold convolution of σ_i with itself. Denote such a vector as σ^{*r}. Let $\tau^{*r} \in \mathcal{A}$ be the corresponding vector induced by $\nu^{\otimes r}$. Thus we have $\kappa_\sigma = H(m_\sigma) = H(m_\tau) = \kappa_\tau$, and

$$H(m_{\mu^{*r}}) = \kappa_{\sigma^{*r}} = (\kappa_{\sigma_0^{*r}}, \ldots, \kappa_{\sigma_n^{*r}}) = (r\kappa_{\sigma_0}, \ldots, r\kappa_{\sigma_n}) = r\kappa_\sigma = r\kappa_\tau = \kappa_{\tau^{*r}} = H(m_{\tau^{*r}})$$

Hence $m_{\sigma^{*r}} = m_{\tau^{*r}}$. It now follows from

$$m_{\sigma[r]_i}(\alpha) = \frac{1}{r} m_{\sigma_i^{*r}}(\alpha) + \frac{r-1}{r} 0$$

that $m_{\sigma[r]} = m_{\tau[r]}$, concluding the proof of the claim.

Equation (21) therefore implies that $M_i^{\mu[r]}(\alpha) = M_i^{\nu[r]}(\alpha)$. Thus

$$d_N(\mu[r], \nu[r]) = \max_i d_{tv}(\overline{\mu[r]}_i, \overline{\nu[r]}_i) \leq \frac{1}{r}.$$

Hence $d_N(\mu[r], \nu[r])$ converges to 0. Since C is uniformly continuous, then $C(\mu[r]) - C(\nu[r]) = 0$ must converge to 0 as well. This implies $C(\mu) = C(\nu)$.

LEMMA 11: *There exists an additive function $F\colon \mathcal{K} \to \mathbb{R}$ such that $c(\sigma) = F(\kappa_\sigma)$.*

PROOF:

It follows from Lemma 10 that we can define a map $G\colon \mathcal{M} \to \mathbb{R}$ such that $c(\sigma) = G(m_\sigma)$ for every $\sigma \in \mathcal{A}$. We can use Theorem 2 to define a bijection $H : \mathcal{M} \to \mathcal{K}$ such that $H(m_\sigma) = \kappa_\sigma$. Hence $F = G \circ H^{-1}$ satisfies $c(\sigma) = F(\kappa_\sigma)$ for every σ. For every $\sigma, \tau \in \mathcal{A}$, Lemmas 8 and 9 imply

$$F(\kappa_\sigma) + F(\kappa_\tau) = c(\sigma) + c(\tau) = c(\sigma_0 * \tau_0, \ldots, \sigma_n * \tau_n) = F(\kappa_{\sigma_0 * \tau_0}, \ldots, \kappa_{\sigma_n * \tau_n}) = F(\kappa_\sigma + \kappa_\tau)$$

where the last equality follows from the additivity of cumulants with respect to convolution.

LEMMA 12: *There exist $(\lambda_{i,\alpha})_{i \in \Theta, \alpha \in A}$ in $\mathbb{R}$ such that*

$$c(\sigma) = \sum_{i \in \Theta} \sum_{\alpha \in A} \lambda_{i,\alpha} \kappa_{\sigma_i}(\alpha) \quad \textit{for every} \quad \sigma \in \mathcal{A}.$$

PROOF:

As implied by Theorem 3, the set $\mathcal{K} \subseteq \mathbb{R}^d$ has nonempty interior. It is closed under addition, i.e. a subsemigroup. We can therefore apply Theorem 4 and conclude that the function F in Lemma 11 is linear.

LEMMA 13: *Let* $(\lambda_{i,\alpha})_{i\in\Theta,\alpha\in A}$ *be as in Lemma 12. Then*

$$c(\sigma)=\sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}m_{\sigma_i}(\alpha)\quad \textit{for every}\quad \sigma\in\mathcal{A}$$

PROOF:

Fix $\sigma\in\mathcal{A}$. Given $t\in(0,1)$, Lemma 12 and Theorem 2 imply

$$\begin{aligned}c\left(t\sigma+(1-t)\delta_{\mathbf{0}}\right) &= \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{\lambda=(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)}\frac{(-1)^{q-1}}{q}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}\prod_{p=1}^{q}m_{t\sigma_i+(1-t)\delta_0}\left(\lambda^p\right)\right)\\ &= \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{\lambda=(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)}\frac{(-1)^{q-1}}{q}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}t^q\prod_{p=1}^{q}m_{\sigma_i}\left(\lambda^p\right)\right)\\ &= \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{\lambda=(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)}\rho\left(\lambda\right)t^q\prod_{p=1}^{q}m_{\sigma_i}\left(\lambda^p\right)\right)\end{aligned}$$

where for every tuple $\lambda=\left(\lambda^1,\ldots,\lambda^q\right)\in\Lambda(\alpha)$ we let

$$\rho\left(\lambda\right)=\frac{(-1)^{q-1}}{q}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}$$

Lemma 9 implies $c(\sigma)=\frac{1}{t}c(t\sigma+(1-t)\,\delta_{\mathbf{0}})$ for every t. Hence

$$c(\sigma)=\sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{\lambda=(\lambda^1,\ldots,\lambda^q)\in\Lambda(\alpha)}\rho(\lambda)t^{q-1}\prod_{p=1}^{q}m_{\sigma_i}(\lambda^p)\right)\ \text{for all } t\in(0,1).$$

By considering the limit $t\downarrow 0$, we have $t^{q-1}\to 0$ whenever $q\neq 1$. Therefore

$$c(\sigma)=\sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}m_{\sigma_i}(\alpha)\ \text{ for all } \sigma\in\mathcal{A}.$$

LEMMA 14: *Let* $(\lambda_{i,\alpha})_{i\in\Theta,\alpha\in A}$ *be as in Lemmas 12 and 13. Then, for every* i*, if* $|\alpha|>1$ *then* $\lambda_{i,\alpha}=0$.

PROOF:

Let $\gamma=\max\left\{|\alpha|:\lambda_{i,\alpha}\neq 0 \text{ for some } i\right\}$. Assume, as a way of contradiction, that $\gamma > 1$. Fix $\sigma \in \mathcal{A}$. Theorem 2 implies

$$\begin{aligned} c(\sigma) &= \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}m_{\sigma_i}(\alpha) \\ &= \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha)}\frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}\prod_{p=1}^{q}\kappa_{\sigma_i}(\lambda^p)\right) \end{aligned}$$

For all $r \in \mathbb{N}$, let $\sigma^{*r} = (\sigma_0^{*r}, \dots, \sigma_0^{*r})$, where each σ_i^{*r} is the r-th fold convolution of σ_i with itself. Hence, using the fact that $\kappa_{\sigma_i^{*r}} = r\kappa_{\sigma_i}$, we obtain

$$c(\sigma^{*r}) = \sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha)}\frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}r^q\prod_{p=1}^{q}\kappa_{\sigma_i}(\lambda^p)\right) \tag{22}$$

By the additivity of c, $c(\sigma^{*r}) = rc(\sigma)$. Hence, because $\gamma > 1$, $c(\sigma^{*r})/r^\gamma \to 0$ as $r \to \infty$. Therefore, diving (22) by r^γ implies

$$\sum_{i\in\Theta}\sum_{\alpha\in A}\lambda_{i,\alpha}\left(\sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha)}\frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}r^{q-\gamma}\prod_{p=1}^{q}\kappa_{\sigma_i}(\lambda^p)\right) \to 0 \text{ as } r \to \infty. \tag{23}$$

We now show that (23) leads to a contradiction. By construction, if $\left(\lambda^1, \dots, \lambda^q\right) \in \Lambda(\alpha)$ then $q \le |\alpha|$. Hence $q \le \gamma$ whenever $\lambda_{i,\alpha} \neq 0$. So, in equation (23) we have $r^{q-\gamma} \to 0$ as $r \to \infty$ whenever $q < \gamma$. Hence in order for (23) to hold it must be that

$$\sum_{i\in\Theta}\sum_{\alpha\in A:|\alpha|=\gamma}\lambda_{i,\alpha}\left(\sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha),q=\gamma}\frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!}\prod_{p=1}^{q}\kappa_{\sigma_i}\left(\lambda^p\right)\right) = 0.$$

If $q = \gamma$ and $\lambda_{i,\alpha} \neq 0$ then $\gamma = |\alpha|$. In this case, in order for $\lambda = \left(\lambda^1, \dots, \lambda^q\right)$ to satisfy $\sum_{p=1}^q \lambda^p = \alpha$, it must be that each λ^p is a unit vector. Every such λ satisfies[36]

$$\prod_{p=1}^{q}\kappa_{\sigma_i}(\lambda^p) = \left(\int_{\mathbb{R}^n}\xi_1\,\mathrm{d}\sigma_i\left(\xi\right)\right)^{\alpha_1}\cdots\left(\int_{\mathbb{R}^n}\xi_n\,\mathrm{d}\sigma_i\left(\xi\right)\right)^{\alpha_n}$$

and

$$\sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha),q=|\alpha|}\frac{1}{q!}\frac{\alpha!}{\lambda^1!\cdots\lambda^q!} = \sum_{(\lambda^1,\dots,\lambda^q)\in\Lambda(\alpha),q=|\alpha|}\frac{\alpha!}{|\alpha|!} = L(\alpha)$$

[36]It follows from the definition of cumulant that for every unit vector $1_j \in \mathbb{R}^n$, $\kappa_{\sigma_i}(1_j) = \int_{\mathbb{R}^n}\xi_j\,\mathrm{d}\sigma_i(\xi)$.

where $L(\alpha)$ is the cardinality of the set of $\left(\lambda^1, \dots, \lambda^q\right) \in \Lambda(\alpha)$ such that $q = |\alpha|$. We obtain that

$$\sum_{i \in \Theta} \sum_{\alpha \in A : |\alpha| = \gamma} L(\alpha) \lambda_{i,\alpha} \left(\int_{\mathbb{R}^n} \xi_1 \, \mathrm{d}\sigma_i(\xi) \right)^{\alpha_1} \cdots \left(\int_{\mathbb{R}^n} \xi_n \, \mathrm{d}\sigma_i(\xi) \right)^{\alpha_n} = 0. \tag{24}$$

By replicating the argument in the proof of Lemma 6 we obtain that the set

$$\left\{ \left(\int_{\mathbb{R}^n} \xi_j \, \mathrm{d}\sigma_i(\xi) \right)_{i,j \in \Theta, j > 0} : \sigma \in \mathcal{A} \right\} \subseteq \mathbb{R}^{(n+1)n}$$

contains an open set U. Consider now the function $f : \mathbb{R}^{(n+1)n} \to \mathbb{R}$ defined as

$$f(z) = \sum_{i \in \Theta} \sum_{\alpha \in A : |\alpha| = \gamma} L(\alpha) \lambda_{i,\alpha} z_{i,1}^{\alpha_1} \cdots z_{i,n}^{\alpha_n}, \quad z \in \mathbb{R}^{(n+1)n}$$

Then (24) implies that f equals 0 on U. Hence, for every $z \in U, i \in \Theta$ and $\alpha \in A$ such that $|\alpha| = \gamma$,

$$L(\alpha) \lambda_{i,\alpha} = \frac{\partial^\gamma}{\partial^{\alpha_1} z_{i,1} \cdots \partial^{\alpha_n} z_{i,n}} f(z) = 0$$

hence $\lambda_{i,\alpha} = 0$. This contradicts the assumption that $\gamma > 1$ and concludes the proof.

For every $j \in \{1, \dots, n\}$ let $1_j \in A$ be the corresponding unit vector. We write λ_{ij} for $\lambda_{i,j}$. Lemma 14 implies that for every distribution $\sigma \in \mathcal{A}$ induced by an experiment $(S, (\mu_i))$, the function c satisfies

$$\begin{aligned} c(\sigma) &= \sum_{i \in \Theta} \sum_{j \in \{1, \dots, n\}} \lambda_{ij} \int_{\mathbb{R}^n} \xi_j \, \mathrm{d}\sigma_i(\xi) \\ &= \sum_{i \in \Theta} \sum_{j \in \{1, \dots, n\}} \lambda_{ij} \int_S \log \frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_0}(s) \, \mathrm{d}\mu_i(s) \\ &= \sum_{i \in \Theta} \sum_{j \in \{1, \dots, n\}} \lambda_{ij} \int_S \log \frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_0}(s) + \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_i}(s) - \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_i}(s) \, \mathrm{d}\mu_i(s) \end{aligned}$$

Hence, using the fact that $\frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_0}\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_i} = \frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_i}$, we obtain

$$\begin{aligned} c(\sigma) &= \sum_{i\in\Theta}\sum_{j\in\{1,\dots,n\}} \lambda_{ij}\int_S \log\frac{\mathrm{d}\mu_j}{\mathrm{d}\mu_i}\,\mathrm{d}\mu_i(s) + \sum_{i\in\Theta}\left(-\sum_{j\in\{1,\dots,n\}}\lambda_{ij}\right)\int_S \log\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_i}(s)\,\mathrm{d}\mu_i(s) \\ &= \sum_{i,j\in\Theta}\beta_{ij}\int_S \log\frac{\mathrm{d}\mu_i}{\mathrm{d}\mu_j}(s)\,\mathrm{d}\mu_i(s) \end{aligned}$$

where in the last step, for every i, we set $\beta_{ij} = -\lambda_{ij}$ if $j \neq 0$ and $\beta_{i0} = \sum_{j\neq 0}\lambda_{ij}$.

It remains to show that the coefficients (β_{ij}) are positive and unique. Because C takes positive values, Lemma 2 immediately implies $\beta_{ij} \geq 0$ for all i, j. The same Lemma easily implies that the coefficients are unique given C.

E. Proofs of the Results of Section V

PROOF OF PROPOSITION 3:

Let $\mu^\star \in \mathcal{P}(A)^n$ be an optimal experiment. Let $A^\star = \mathrm{supp}(\mu^\star)$ be the set of actions played in $\mu^\star$. It solves

$$\max_{\mu\in\mathbb{R}_+^{|\Theta|\times|A^\star|}}\left[\sum_{i\in\Theta} q_i\left(\sum_{a\in A}\mu_i(a)u(a,i)\right) - \sum_{i,j\in\Theta}\beta_{ij}\sum_{a\in A^\star}\mu_i(a)\log\frac{\mu_i(a)}{\mu_j(a)}\right] \tag{25}$$

$$\text{subject to} \quad \sum_{a\in A^\star}\mu_i(a) = 1 \text{ for all } i\in\Theta. \tag{26}$$

Reasoning as in Cover and Thomas (2012, Theorem 2.7.2) the Log-sum inequality implies that the function D_{KL} is convex when its domain is extended from pairs of probability distributions to pairs of vectors in $\mathbb{R}_+^{|A^\star|}$. Moreover, expected utility is linear in the choice probabilities. It then follows that the objective function in (25) is concave over $\mathbb{R}_+^{|\Theta|\times|A^\star|}$.

As (25) equals $-\infty$ whenever $\mu_i(a) = 0$ for some i and $\mu_j(a) > 0$ for some $j \neq i$ we have that $\mu_i^\star(a) > 0$ for all $i \in \Theta, a \in A^\star$. For every $\lambda \in \mathbb{R}^{|\Theta|}$ we define the Lagrangian $L_\lambda(\mu)$ as

$$L_\lambda(\mu) = \left[\sum_{i\in\Theta} q_i\left(\sum_{a\in A}\mu_i(a)u(a,i)\right) - \sum_{i,j\in\Theta}\beta_{ij}\sum_{a\in A}\mu_i(a)\log\frac{\mu_i(a)}{\mu_j(a)}\right] - \sum_{i\in\Theta}\lambda_i\sum_{a\in A}\mu_i(a)\,.$$

As $\mu^\star$ is an interior solution to (25), it follows from the Karush-Kuhn-Tucker theorem that there exists Lagrange multipliers $\lambda \in \mathbb{R}^{|\Theta|}$ such that $\mu^\star$ maximizes

$L_\lambda(\cdot)$ over $\mathbb{R}_+^{|\Theta|\times|A^\star|}$. As $\mu^\star$ is interior it satisfies the first order condition

$$\nabla L_\lambda(\mu^\star) = 0\,.$$

We thus have that for every state $i \in \Theta$ and every action $a \in A^\star$

$$0 = q_i u_i(a) - \lambda_i - \sum_{j\neq i}\left\{\beta_{ij}\left[\log\left(\frac{\mu_i^\star(a)}{\mu_j^\star(a)}\right) - 1\right] - \beta_{ji}\frac{\mu_j^\star(a)}{\mu_i^\star(a)}\right\}\,. \tag{27}$$

Subtracting (27) evaluated at a' from (27) evaluated at a yields the desired necessary conditions for the optimality of $\mu^\star$.

PROOF OF PROPOSITION 4:

We prove a slightly more general result. Assume the coefficients satisfy $\beta_{ij} \geq 1/f(d(i,j))^2$, where f is a strictly positive and increasing function f.

The cost of the optimal experiment $\mu^\star$ must satisfy $\|u\| \geq C(\mu^\star)$, otherwise the decision maker would be better off acquiring no information. Pinsker's inequality (see Borwein and Vanderwerff, 2010, p. 13) implies

$$C(\mu^\star) \geq \min\{\beta_{ij},\beta_{ji}\}(D_{\mathrm{KL}}(\mu_i^\star\|\mu_j^\star) + D_{\mathrm{KL}}(\mu_j^\star\|\mu_i^\star)) \geq \min\{\beta_{ij},\beta_{ji}\}\|\mu_i^\star - \mu_j^\star\|_1^2.$$

where $\|\mu_i^\star - \mu_j^\star\|_1 = \sum_{a\in A}|\mu_i^\star(a) - \mu_j^\star(a)|$ denotes the total-variation norm between the two distributions. We then obtain

$$\|\mu_i^\star - \mu_j^\star\|_1 \leq \sqrt{\|u\|\frac{1}{\min\{\beta_{ij},\beta_{ji}\}}} \leq \sqrt{\|u\|}f(d(i,j))\,.$$

In particular, if f is the identity function then $\|\mu_i^\star - \mu_j^\star \leq \sqrt{\|u\|}d(i,j)$.

PROOF OF PROPOSITION 6:

Given a vector $\mu \in \mathcal{P}(\{B,R\})^\Theta$, we use the shorthand μ_i to denote the probability $\mu_i(B)$ of guessing B in state i. For every μ, let

$$U(\mu) = \frac{1}{|\Theta|}\left(\sum_{i<n/2}(1-\mu_i) + \sum_{i>n/2}\mu_i\right) - C(\mu) \tag{28}$$

be the net expected payoff provided by μ, where C is an LLR cost function such that $\beta_{ij} = f(|i-j|)$ for some positive and strictly decreasing function f.

Let $\mathcal{P}_+$ be the set of probabilities μ such that each μ_i has support $\{B,R\}$. Let $\mu^\star$ be a solution to the problem $\max_{\mu\in\mathcal{P}_+} U(\mu)$. Such a solution exists and is unique. In fact, the problem $\max_{\mu\in\mathcal{P}(\{B,R\})^\Theta} U(\mu)$ has a solution. Now, if $\mu^\star$ is optimal and $\mu^\star \notin \mathcal{P}_+$, then either $\mu_i^\star = 0$ for every i or $\mu_i^\star = 1$ for every i. In either case $U(\mu^\star) = U(\mu)$, where $\mu \in \mathcal{P}_+$ is defined as $\mu_i = 1/2$ for every i. It follows that the problem $\max_{\mu\in\mathcal{P}_+} U(\mu)$ admits a solution $\mu^\star$. Over $\mathcal{P}_+$ the function C is strictly convex,[37] and thus U is strictly concave. Thus, the solution is unique.

We claim that $\mu^\star$ satisfies $\mu^\star_{n/2+r} = 1 - \mu^\star_{n/2-r}$ for every r. To see this, define $\mu \in \mathcal{P}_+$ as $\mu_{n/2+r} = 1 - \mu^\star_{n/2-r}$ for every r. Because $U(\mu^\star) = U(\mu)$ and U is strictly concave on $\mathcal{P}_+$, we conclude that $\mu = \mu^\star$.

Let $I \subseteq \mathcal{P}(\{B,R\})^\Theta$ be the set of vectors μ that are increasing, that is, satisfy $\mu_i \leq \mu_{i+1}$ for every $i < n$, and consider the optimization problem

$$\max_{\mu \in I \cap \mathcal{P}_+} U(\mu).$$

The set I is closed and U is upper semi-continuous. Thus, the problem $\max_{\mu \in I} U(\mu)$ has a solution. The same argument applied in the previous paragraph implies $\max_{\mu \in I \cap \mathcal{P}_+} U(\mu)$ admits a solution as well, and that such a solution is unique. We denote it by $\hat{\mu}$.

As we show in the next paragraph, the vector $\hat{\mu}$ is strictly increasing: it satisfies $\hat{\mu}_i < \hat{\mu}_{i+1}$ for every i. This implies $\mu^\star = \hat{\mu}$. Indeed, we have $U(\mu^\star) \geq U(\hat{\mu})$, since $\mu^\star$ is obtained by maximizing U over a larger domain. If $U(\mu^\star) > U(\hat{\mu})$ the concavity of U implies $U(\alpha\mu^\star + (1-\alpha)\hat{\mu}) > U(\hat{\mu})$ for all $\alpha \in [0,1]$. Because $\hat{\mu}$ is strictly increasing, then for α small enough the vector $\alpha\mu^\star + (1-\alpha)\hat{\mu}$ belongs to I, contradicting the optimality of $\hat{\mu}$. It follows that $U(\mu^\star) = U(\hat{\mu})$, and hence $\mu^\star = \hat{\mu}$, since the problem $\max_{\mu \in \mathcal{P}_+} U(\mu)$ has a unique solution.

We now show $\hat{\mu}$ is strictly increasing. Given $\nu, \rho \in (0,1)$ we denote by $D_1(\nu\|\rho)$ and $D_2(\nu\|\rho)$ the partial derivatives of the Kullback-Leibler divergence D_{KL} with respect to its the first and second arguments:

$$\begin{aligned} D_1(\rho\|\nu) &= \log\frac{\rho}{\nu} - \log\frac{1-\rho}{1-\nu} \\ D_2(\rho\|\nu) &= -\frac{\rho}{\nu} + \frac{1-\rho}{1-\nu}. \end{aligned}$$

Both derivatives are equal to zero if and only if $\nu = \rho$.

As a way of contradiction, suppose $\hat{\mu}$ is not strictly increasing. Let $[i,k]$ be a maximal interval of states over which $\hat{\mu}$ is constant. Let μ^ε be the vector obtained from $\hat{\mu}$ by increasing $\hat{\mu}_k$ by $\varepsilon > 0$ and decreasing $\hat{\mu}_i$ by ε (since $\hat{\mu} \in \mathcal{P}_+$, both operations are feasible for ε small enough). The function $\varepsilon \mapsto U(\mu^\varepsilon)$ is differentiable. Its derivative at $\varepsilon = 0$ is equal to

$$\frac{\mathrm{sgn}(k-n/2)}{|\Theta|} - \sum_{j\neq k}\beta_{jk}(D_2(\hat{\mu}_j\|\hat{\mu}_k) + D_1(\hat{\mu}_k\|\hat{\mu}_j)) - \frac{\mathrm{sgn}(i-n/2)}{|\Theta|} + \sum_{j\neq i}\beta_{ij}(D_2(\hat{\mu}_j\|\hat{\mu}_i) + D_1(\hat{\mu}_k\|\hat{\mu}_i)). \tag{29}$$

Since $\hat{\mu}$ is constant in the interval $[i,k]$, then $D_1(\hat{\mu}_j\|\hat{\mu}_m) = D_2(\hat{\mu}_j\|\hat{\mu}_m)$ whenever

[37] See Corollary 1.55 in Liese and Vajda (1987)

$i \leq j \leq m \leq k$. We can therefore rewrite (29) as

(30)
$$\frac{\operatorname{sgn}(k-n/2)}{|\Theta|} - \sum_{j>k} \beta_{jk}(D_2(\hat{\mu}_j \| \hat{\mu}_k) + D_1(\hat{\mu}_k \| \hat{\mu}_j)) - \sum_{j<i} \beta_{jk}(D_2(\hat{\mu}_j \| \hat{\mu}_k) + D_1(\hat{\mu}_k \| \hat{\mu}_j))$$
$$-\frac{\operatorname{sgn}(i-n/2)}{|\Theta|} + \sum_{j>k} \beta_{ij}(D_2(\hat{\mu}_j \| \hat{\mu}_i) + D_1(\hat{\mu}_k \| \hat{\mu}_i)) + \sum_{j<i} \beta_{ij}(D_2(\hat{\mu}_j \| \hat{\mu}_i) + D_1(\hat{\mu}_i \| \hat{\mu}_j)) \,.$$

The derivative (30) is strictly positive. Indeed, because $k \geq i$ then $\operatorname{sgn}(k-n/2) - \operatorname{sgn}(i-n/2) \geq 0$. Whenever $j > k$, since $\hat{\mu}_j > \hat{\mu}_k = \hat{\mu}_i$ and D is strictly convex over $\mathcal{P}_+$, we have

$$D_2(\hat{\mu}_j \| \hat{\mu}_k) = D_2(\hat{\mu}_j \| \hat{\mu}_i) < 0 \text{ and } D_1(\hat{\mu}_k \| \hat{\mu}_j) = D_1(\hat{\mu}_i \| \hat{\mu}_j) < 0$$

Moreover $\beta_{jk} > \beta_{ji}$ since $|j-k| < |i-k|$. It follows that

$$-\sum_{j>k} \beta_{jk}(D_2(\hat{\mu}_j \| \hat{\mu}_k) + D_1(\hat{\mu}_k \| \hat{\mu}_j)) + \sum_{j>k} \beta_{ij}(D_2(\hat{\mu}_j \| \hat{\mu}_i) + D_1(\hat{\mu}_i \| \hat{\mu}_j))$$

is strictly positive if $k < n$, and equal to 0 if $k = n$. An analogous argument shows that

$$-\sum_{j<i} \beta_{jk}(D_2(\hat{\mu}_j \| \hat{\mu}_k) + D_1(\hat{\mu}_k \| \hat{\mu}_j)) + \sum_{j<i} \beta_{ij}(D_2(\hat{\mu}_j \| \hat{\mu}_i) + D_1(\hat{\mu}_i \| \hat{\mu}_j))$$

is strictly positive if $i > 0$, and equal to 0 if $i = 0$. Because $\hat{\mu} \in \mathcal{P}_+$, then either $k < n/2 + r$, $i > n/2 - r$, or both. This implies that (30) is strictly positive. Hence, for small enough ε, the vector μ^ε satisfies $U(\mu^\varepsilon) > U(\hat{\mu})$, contradicting the hypothesis that $\hat{\mu}$ is optimal. We therefore conclude that $\hat{\mu}$ is strictly increasing, and thus $\mu^\star$ is strictly increasing as well.

Because $\mu^\star$ satisfies $\mu^\star_{n/2+r} = \mu^\star_{n/2-r}$ for every r, and $\mu^\star$ is strictly increasing, it follows that $m_i > m_j$ for every pair of states such that $|i - n/2| > |j - n/2|$.

PROOF OF PROPOSITION 7:

Denote by $\mathcal{P}_+$ be the set of probabilities $\mu \in \mathcal{P}(\{a_1, a_2\})^2$ such that $\operatorname{supp}(\mu) = \{a_1, a_2\}$. Let $\mu \in \mathcal{P}_+$ be an optimal experiment. We first show that μ satisfies $\mu_1(a_1) = \mu_2(a_2)$. To see this, define μ' as $\mu'_1(a_1) = \mu_2(a_2)$ and $\mu'_2(a_2) = \mu_1(a_1)$. Let $\mu'' = \frac{1}{2}\mu + \frac{1}{2}\mu'$. By the symmetry of the payoffs functions and of the prior, we have

$$\sum_{i\in\Theta} q_i \left(\sum_{a\in A} \mu_i(a) u(a,i) \right) = \sum_{i\in\Theta} q_i \left(\sum_{a\in A} \mu'_i(a) u(a,i) \right) = \sum_{i\in\Theta} q_i \left(\sum_{a\in A} \mu''_i(a) u(a,i) \right) .$$

Moreover, $C(\mu'') \leq \frac{1}{2}C(\mu) + \frac{1}{2}C(\mu')$ if $\mu \neq \mu'$, as C is strictly convex on $\mathcal{P}_+$. Since

μ is optimal, it must be that $\mu = \mu'$.

The optimality equation $\mathrm{MB}_1(a_1, a_2) = \mathrm{MC}_1(a_1, a_2)$ can now be rewritten as

$$\frac{1}{2}v = \beta\left[\xi\left(\log\left(\frac{\mu_1(a_1)}{\mu_2(a_1)}\right)\right) - \xi\left(\log\left(\frac{\mu_1(a_2)}{\mu_2(a_2)}\right)\right)\right].$$

with $\xi(x) = x + \mathrm{e}^x$. Simple calculations show the expression is in turn equal to

$$\frac{v}{2\beta} = \xi\left(\log\left(\frac{\mu[v]}{1-\mu[v]}\right)\right) - \xi\left(\log\left(-\frac{\mu[v]}{1-\mu[v]}\right)\right) = \zeta\left(\log\left(\frac{\mu[v]}{1-\mu[v]}\right)\right)$$

where $\zeta(x) = 2x + e^x - e^{-x}$. The result now follows by defining $\eta = \zeta^{-1}$.

PROOF OF PROPOSITION 5:

Consider a decision problem described by a payoff function u and a prior q. let μ and μ' be the optimal choice probabilities obtained under the coefficients (β_{ij}) and (β'_{ij}). The optimality of μ and μ' implies

$$\sum_{i,a} q_i u(i,a)\mu_i(a) - \sum_{i,j} \beta_{ij} D(\mu_i \| \mu_j) \geq \sum_{i,a} q_i u(i,a)\mu'_i(a) - \sum_{i,j} \beta_{ij} D(\mu'_i \| \mu'_j)$$
$$\sum_{i,a} q_i u(i,a)\mu'_i(a) - \sum_{i,j} \beta'_{ij} D(\mu'_i \| \mu'_j) \geq \sum_{i,a} q_i u(i,a)\mu_i(a) - \sum_{i,j} \beta'_{ij} D(\mu_i \| \mu_j)$$

Rearranging the two inequalities leads to

$$\sum_{i,j} \beta_{ij}(D(\mu'_i\|\mu'_j) - D(\mu_i\|\mu_j)) \geq \sum_{i,a} q_i u(i,a)(\mu'_i(a) - \mu_i(a)) \geq \sum_{i,j} \beta'_{ij}(D(\mu'_i\|\mu'_j) - D(\mu_i\|\mu_j)).$$

The result now follows.

F. Proof of Proposition 2 and Extensions

PROOF OF PROPOSITION 2:

Denote by $w > 0$ the length of W. Let $|\Theta| = n$. By Axiom a there exists a function $f\colon (0, w) \to \mathbb{R}_+$ such that $\beta^{\Theta}_{ij} = f(|i - j|)$ for $i \neq j$. Hence, if we translate W then β^{Θ}_{ij} remains unchanged. We can therefore assume without loss of generality that $W = (-\delta, w - \delta)$, for any $\delta \in (0, w)$.

Let $g\colon (0, w) \to \mathbb{R}_+$ be given by $g(t) = \frac{1}{2}f(t)t^2$. The Kullback-Leibler divergence between two normal distributions with unit variance and expectations i and j is $(i - j)^2/2$. Hence, by Axiom b there exists a constant $\kappa \geq 0$, independent of n, so that

$$(31) \qquad \frac{1}{2}\kappa = C^{\Theta}(\zeta^{\Theta}) = \sum_{i \neq j \in \Theta} \beta^{\Theta}_{ij}\frac{(i-j)^2}{2} = \sum_{i \neq j \in \Theta} g(|i - j|) \qquad \text{for any } \Theta \in \mathcal{T}$$

We show that (31) implies that

$$g(t) = \frac{\kappa}{2n(n-1)},$$

so that

$$\beta_{ij}^{\Theta} = 2g(|i-j|)\frac{1}{(i-j)^2} = \frac{\kappa}{n(n-1)}\frac{1}{(i-j)^2},$$

which will complete the proof. The case $n = 2$ is immediate, since then $\Theta = \{i, j\}$ and so (31) reduces to

$$\frac{1}{2}\kappa = 2g(|i-j|).$$

We now consider the case $n > 2$. Let $\Theta = \{i_1, i_2, \dots, i_{n-1}, x\}$ with $i_1 < i_2 < \dots < i_{n-1} < x$ and $x \in (0, w - \delta)$. Then (31) implies

$$\kappa = 2\sum_{\ell=1}^{n-1} g(x - i_\ell) + 2\sum_{k=1}^{n-1}\sum_{\ell=1}^{k-1} g(i_k - i_\ell).$$

Taking the difference between this equation and the analogous one corresponding to $\Theta' = \{i_1, i_2, \dots, i_{n-1}, y\}$ with $y \in (x, w - \delta)$ yields

$$0 = \sum_{\ell=1}^{n-1} g(x - i_\ell) - g(y - i_\ell).$$

Denoting $i_1 = -\varepsilon$, for some $\varepsilon \in (\delta, 0)$, we can write this as

$$0 = g(x + \varepsilon) - g(y + \varepsilon) + \sum_{\ell=2}^{n-1} g(x - i_\ell) - g(y - i_\ell).$$

Again taking a difference, this time of this equation with the analogous one obtained by setting $i_1 = 0$, we get

$$g(x) - g(y) = g(x + \varepsilon) - g(y + \varepsilon).$$

Rearranging yields

(32)
$$g(y + \varepsilon) - g(y) = g(x + \varepsilon) - g(x) \qquad \text{for all } x, y \in (0, w - \delta) \text{ and } \varepsilon \in (0, \delta).$$

Accordingly, for $\varepsilon \in (0, \delta)$ denote

(33)
$$h(\varepsilon) = g(x + \varepsilon) - g(x),$$

where by (32) the right hand side does not depend on the choice of $x \in (0, w - \delta - \varepsilon)$. It follows that

$$h(\varepsilon_1 + \varepsilon_2) = [g(x + \varepsilon_1 + \varepsilon_2) - g(x + \varepsilon_1)] + [g(x + \varepsilon_1) - g(x)] = h(\varepsilon_1) + h(\varepsilon_2) \tag{34}$$

for all $\varepsilon_1, \varepsilon_2 \in (0, \delta/2)$. That is, h satisfies the Cauchy functional equation on $(0, \delta/2)$.

Since g is non-negative, it follows from (31) that g is bounded by κ. Hence the absolute value of h is bounded by κ, by (33). It follows that $\lim_{\varepsilon \to 0} h(\varepsilon) = 0$. Otherwise, there is some n such that $|h(\varepsilon)| > \kappa/n$ for arbitrarily small ε, and then, by repeated application of (34),

$$h(n\varepsilon) = nh(\varepsilon) > \kappa,$$

where we choose ε small enough so that $n\varepsilon < \delta/2$.

From $\lim_{\varepsilon \to 0} h(\varepsilon) = 0$ and (34) it follows that h is continuous on $(0, \delta/2)$. As the Cauchy equation easily implies that h is linear when restricted to the rationals, continuity implies that h is linear on $(0, \delta/2)$. Thus, by (32) g is affine on $(0, w - \delta)$, and of the form $g(t) = at + b$ for some $a, b \in \mathbb{R}$. We claim that it must be that $a = 0$. Otherwise, for a given $\Theta = \{i_1, \ldots, i_{n-1}, x\}$, $\sum_{i \neq j \in \Theta} g(|i - j|)$ changes with x, in violation of (31). It follows that g is constant on $(0, w - \delta)$. And since we can take δ arbitrarily small, g is constant on its domain $(0, w)$. Finally, for (31) to be satisfied, this constant must be $\frac{\kappa}{2n(n-1)}$.

Axiom b calibrates the parameters (β_{ij}^{Θ}) using an experiment consisting of a measurement with Normally distributed noise. Different distributions for the noise would lead to different representations for the coefficients. For example, a natural alternative would be an experiment $(\mathbb{R}, (\xi_i)_{i \in \Theta})$ where each ξ_i is Laplace distribution with variance 1 and mean equal to the state i (the corresponding probability density function is $f(x) = \frac{1}{2}\mathrm{e}^{|x-i|}$). The divergence $D(\xi_i \| \xi_j)$ between any two such distribution is

$$\mathrm{e}^{-|i-j|} + |i - j| - 1.$$

As in the Normal case, this is a decreasing function of the distance between states. Even if the distribution used in axiom b is different, the proof of Proposition 2 can be applied with almost no modifications, and leads to a representation with parameters

$$\beta_{ij}^{\Theta} = \frac{\kappa}{n(n-1)} \frac{1}{\mathrm{e}^{-|i-j|} + |i - j| - 1}.$$

G. Identification

Consider the setup of §V.D, and given a pair of choice probabilities (μ_1, μ_2) define the quantities

$$\hat{\beta}_{12} = \frac{l_2 - l_1 + \log\frac{l_1}{l_2}}{\frac{(l_1-l_2)^2}{l_1 l_2} - (\log\frac{l_1}{l_2})^2} \qquad \text{and} \qquad \hat{\beta}_{21} = \frac{v}{2}\frac{\frac{l_2-l_1}{l_1 l_2} + \log\frac{l_1}{l_2}}{\frac{(l_1-l_2)^2}{l_1 l_2} - (\log\frac{l_1}{l_2})^2}$$

PROPOSITION 11: *The choice probabilities (μ_1, μ_2) are the optimal solution with respect to an LLR cost function if and only if $\hat{\beta}_{12}$ and $\hat{\beta}_{21}$ are non-negative and at least one is positive.*

PROOF:
As shown by Proposition 10, C is a convex function. We note that the condition (14) is equivalent to (13) which equals the first order condition for the optimization problem, which is sufficient because of the concavity of the optimization problem. If at least one of $\hat{\beta}_{1,2}, \hat{\beta}_{2,1}$ is positive, then the solution of the optimization problem is internal and the first order condition applies. Conversely, if both are zero then the optimization problem has no solution within its domain.

H. The cost of bounded experiments with binary state

In this section we restrict ourselves to the case of a binary state space $\Theta = \{0, 1\}$, and the class of *bounded* experiments $\mathcal{B}$: an experiment is said to be bounded if the beliefs that it induces are bounded away from 0 and 1. In terms of log-likelihood ratios, it is bounded if there is some M such that $\ell_{01}(s)$ is μ_0- and μ_1-almost surely in $[-M, M]$. The class of bounded experiments is contained in the class $\mathcal{E}$ of experiments considered in the rest of the paper. The bounded experiments contain all the experiments that have a finite set of possible realizations, and in which not state is ever conclusively excluded.

As we discuss above, a strengthening of Axiom 1 is Blackwell monotonicity: C is said to be Blackwell monotone if $C(\mu) \geq C(\nu)$ whenever If μ Blackwell dominates ν.

For the class of bounded experiments, we show that 2 and 3 are sufficient for proving that a Blackwell monotone cost is an LLR cost: the continuity axiom 4 is not needed. This proof heavily relies on a recent result of Mu et al. (2020), which characterizes the monotone and additive functions on the class of bounded Blackwell experiments with binary state. An extension of this result to large state spaces is currently out of reach, and so we do not have a more general proof. Nevertheless, we conjecture that the continuity axiom is generally redundant.

THEOREM 5: *Let $\Theta = \{0, 1\}$. A Blackwell monotone information cost function $C\colon \mathcal{B} \to \mathbb{R}_+$ satisfies Axioms 2 and 3 if and only if there exist $\beta_{01}, \beta_{10} \geq 0$ such* *that for every experiment $\mu \in \mathcal{B}$,*

$$C(\mu) = \beta_{01} D_{\mathrm{KL}}(\mu_0 \| \mu_1) + \beta_{10} D_{\mathrm{KL}}(\mu_1 \| \mu_0).$$

Before proving Theorem 5, we will introduce some definitions and results from Mu et al. (2020).

For $t \in (0, \infty]$, we denote by $R_t(\mu_0 \| \mu_1)$ the Rényi t-divergence between two probability μ_0, μ_1 defined on the same measurable space S. For $t \neq 1$, $t \neq \infty$,

$$R_t(\mu_0 \| \nu_1) = \frac{1}{t-1} \log \int_S \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(s) \right)^{t-1} \mathrm{d}\mu_0(s).$$

For $t = 1$

$$R_1(\mu_0 \| \mu_1) = \int_S \log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(s) \, \mathrm{d}\mu_0(s) = D_{\mathrm{KL}}(\mu_0 \| \mu_1).$$

For $t = \infty$, $R_\infty(\mu_0 \| \mu_1)$ is the essential maximum of the log-likelihood ratio $\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}$. Note that $R_t(\mu_0 \| \mu_1)$ is always non-negative, and positive whenever $\mu_0 \neq \mu_1$. Note also that if $\log \frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}$ is almost surely in $[-M, M]$ (as is always the case for bounded experiments, for some M) then $R_t \leq M$.

The following result is a reformulation of Theorem 2 in Mu et al. (2020) (see also Lemmas 5 and 6).[38]

THEOREM 6 (Mu et al. 2020): *An information cost function $C \colon \mathcal{B} \to \mathbb{R}_+$ satisfies Axioms 1 and 2 if and only if there exist two finite Borel measures m_0, m_1 on $[1/2, \infty]$ such that for every bounded experiment $\mu = (S, \mu_0, \mu_1)$ it holds that*

$$C(\mu) = \int_{[1/2,\infty]} R_t(\mu_0 \| \mu_1) \, \mathrm{d}m_0(t) + \int_{[1/2,\infty]} R_t(\mu_1 \| \mu_0) \, \mathrm{d}m_1(t).$$

Using this result, we can now prove Theorem 5.

PROOF OF THEOREM 5:

The argument that this representation satisfies the axioms is identical to the same argument in the proof of Theorem 1. It thus remains to be shown that the representation is implied by the axioms.

[38] The *data processing inequality* in that paper is monotonicity with respect to deterministic garblings, which is implied by Blackwell monotonicity. The additivity there translates immediately to additivity in the sense of Axiom 2.

By Theorem 6,

$$
\begin{aligned}
C(\mu) = {} & \beta_{01} D_{\mathrm{KL}}(\mu_0 \| \mu_1) + \beta_{10} D_{\mathrm{KL}}(\mu_1 \| \mu_0) \\
& + \int_{[1/2,1)} R_t(\mu_0 \| \mu_1)\, \mathrm{d}m_0(t) + \int_{[1/2,1)} R_t(\mu_1 \| \mu_0)\, \mathrm{d}m_1(t) \\
& + \int_{(1,\infty]} R_t(\mu_0 \| \mu_1)\, \mathrm{d}m_0(t) + \int_{(1,\infty]} R_t(\mu_1 \| \mu_0)\, \mathrm{d}m_1(t). \qquad (35)
\end{aligned}
$$

for some $\beta_{01}, \beta_{10} \geq 0$ and m_0, m_1 finite Borel measures on $[1/2, \infty]$ that assign measure 0 to the singleton $\{1\}$. To prove the claim, we show that m_0 and m_1 are the zero measures.

Let $\mu = (S, \mu_0, \mu_1)$ be a non-trivial bounded experiment, and let $\nu = (1/r) \cdot \mu^{\otimes r}$ for some r. It follows from the definition of Rényi t-divergences that for $t \neq 1$, $t \neq \infty$

$$
R_t(\nu_0 \| \nu_1) = \frac{1}{t-1} \log\left(\frac{r-1}{r} + \frac{1}{r} \left(\int_S \left(\frac{\mathrm{d}\mu_0}{\mathrm{d}\mu_1}(s) \right)^{t-1} \mathrm{d}\mu_0(s) \right)^r \right).
$$

Now, for $x > 1$,

$$
\lim_{r \to \infty} \log\left(\frac{r-1}{r} + \frac{1}{r} x^r \right) = \infty,
$$

and for $x < 1$ this same limit is 0. It thus follows that for $t > 1$ (including, trivially, $t = \infty$)

$$
\lim_{r \to \infty} R_t(\nu_0 \| \nu_1) = \infty, \qquad (36)
$$

since R_t is positive for non-trivial experiments, and so the integral in the expression for R_t is strictly greater than 1. For $t < 0$

$$
\lim_{r \to \infty} R_t(\nu_0 \| \nu_1) = 0, \qquad (37)
$$

since, again by the positivity of R_t, the integral in the expression for R_t is strictly less than 1.

It follows from (36) that both m_0 and m_1 must assign no mass to $(1, \infty]$, i.e. $m_0((1, \infty]) = m_1((1, \infty]) = 0$, since otherwise the integral $\int_{(1,\infty]} R_t(\mu_0 \| \mu_1)\, \mathrm{d}m_0(t)$ or $\int_{(1,\infty]} R_t(\mu_0 \| \mu_1)\, \mathrm{d}m_1(t)$ would diverge and by (35) the cost of the experiment $(1/r) \cdot \mu^{\otimes r}$ would diverge

$$
\lim_{r \to \infty} C((1/r) \cdot \mu^{\otimes r}) = \infty\,.
$$

This would contradict the axioms which imply that $C((1/r) \cdot \mu^{\otimes r}) = C(\mu)$. It then follows from (37) that $m_0((1/2,1)) = m_1((1/2,1)) = 0$, since otherwise

$$\lim_{r\to\infty} C((1/r)\cdot\mu^{\otimes r}) < C(\mu).$$

I. Uniform Separable Bayesian LLR Cost

PROOF OF PROPOSITION 8:

It is straightforward to verify that if the parameters satisfy $\beta_{ij}(q) = \gamma_{ij} q_i$, then C is uniformly posterior separable. We now prove the opposite implication.

Fix a prior q with full support, and consider an experiment μ where the set of signal realizations is a product $S_1 \times S_2$, with S_1 a finite set, and each μ_i satisfies $\mu_i(\{s\} \times S_2) > 0$ for every $s \in S_1$. We denote by μ_i^1 the marginal of μ_i on S_1, and by $\mu_i(\cdot|s)$ the measure on S_2 obtained by conditioning μ_i on $s \in S_1$.

The chain rule for the KL-divergence implies that the cost of such an experiment can be written as

$$C(\mu,q) = \sum_{ij} \beta_{ij}(q)\left[D_{\mathrm{KL}}(\mu_i^1\|\mu_j^1) + \sum_{s_1\in S_1}\mu_i^1(s_1) D_{\mathrm{KL}}(\mu_i(\cdot|s_1)\|\mu_j(\cdot|s_1))\right]. \tag{38}$$

Now assume C is uniformly posterior separable with respect to a function G. The cost of the experiment μ can then be written as follows. It will be convenient to denote posterior beliefs as random variables defined over the probability space $(\Theta \times S_1 \times S_2, \mathbb{P})$ where $\mathbb{P}$ is obtained from q and μ in the obvious way. Let p^2 be the posterior belief over Θ obtained by conditioning q on a realization (s_1, s_2), and let p^1 be the posterior belief obtained by conditioning q on a realization s_1. Then

$$\begin{aligned} C(\mu,q) &= \mathbb{E}\left[G(p^2) - G(p^1) + G(p^1) - G(q)\right] \\ &= \mathbb{E}\left[G(p^1) - G(q)\right] + \sum_{s_1\in S_1}\mathbb{P}(s_1)\mathbb{E}\left[G(p^2) - G(p^1)|p^1 = q(\cdot|s_1)\right]. \end{aligned}$$

Now consider the experiment $((\mu_i^1), S)$ which consists of observing the first realization s_1 but not the second. By uniform posterior separability, its cost, at the prior q, is given by

$$\mathbb{E}\left[G(p^1) - G(q)\right] = \sum_{ij}\beta_{ij}(q) D_{\mathrm{KL}}(\mu_i^1\|\mu_j^1).$$

Given a realization $s_1 \in S$, consider the experiment $((\mu_i(\cdot|s_1)), S_2)$. By considering now $p^1 = q(\cdot|s_1)$ as a prior, uniform separability implies that the cost of the experiment $((\mu_i(\cdot|s_1)), S_2)$ is equal to

$$\mathbb{E}\left[G(p^2) - G(p^1)|p^1 = q(\cdot|s^1)\right] = \sum_{ij}\beta_{ij}(q(\cdot|s_1)) D_{\mathrm{KL}}(\mu_i(\cdot|s_1)\|\mu_j(\cdot|s_1)).$$

The last two equations imply that the cost $C(\mu, q)$ can be rewritten as
(39)

$$\sum_{ij} \beta_{ij}(q) D_{\mathrm{KL}}(\mu_i^1 \| \mu_j^1) + \sum_{s_1 \in S_1} \mathbb{P}(s_1) \left(\sum_{ij} \beta_{ij}(q(\cdot|s_1)) D_{\mathrm{KL}}(\mu_i(\cdot|s_1) \| \mu_j(\cdot|s_1)) \right).$$

This equation can be interpreted as saying that the cost of running the experiment μ is equal to the cost of running the first experiment $((\mu_i^1), S_1)$ plus the expected cost of running the second experiment $((\mu_i(\cdot|s_1)), S_2)$, conditional on the signal realization s_1 from the first experiment. By equating (38) and (39) we obtain that

$$\sum_{s_1 \in S_1} \sum_{ij} \left[\beta_{ij}(q) \mu_i^1(s_1) - \mathbb{P}(s_1) \beta_{ij}(q(\cdot|s_1)) \right] D_{\mathrm{KL}}(\mu_i(\cdot|s_1) \| \mu_j(\cdot|s_1)) = 0. \tag{40}$$

Given a particular realization $s_1 \in S_1$, we are free to choose μ such that all the conditional experiments $((\mu_i(\cdot|s_1')), S_2)$, $s_1' \neq s_1$, are completely uninformative, and hence have cost 0. Thus, it must hold that for every $s_1 \in S_1$,

$$\sum_{ij} \left[\beta_{ij}(q) \mu_i^1(s_1) - \mathbb{P}(s_1) \beta_{ij}(q(\cdot|s_1)) \right] D_{\mathrm{KL}}(\mu_i(\cdot|s_1) \| \mu_j(\cdot|s_1)) = 0.$$

By Lemma 2, the latter can hold only if

$$\beta_{ij}(q) \mu_i^1(s_1) = \mathbb{P}(s_1) \beta_{ij}(q(\cdot|s_1)).$$

By dividing and multiplying the left-hand side by q_i and then applying Bayes' rule we obtain that

$$\frac{\beta_{ij}(q)}{q_i} = \frac{\beta_{ij}(q(\cdot|s_1))}{q(\cdot|s_1)}.$$

Given any $q' \in \mathcal{P}(\Theta)$ with full support, we can choose μ such that $q(\cdot|s_1) = q'$ for some s_1. The conclusion now follows by defining $\gamma_{ij} = \beta_{ij}(q)/q_i$.

Prior Dependence of Bayesian LLR Cost.

As we prove in Proposition 8, the only uniformly posterior separable LLR cost potentially assigns different cost to the same experiment at different prior beliefs. We next explore which experiments have prior dependent cost, through a simple example of binary experiments. Consider the standard setting of a binary state space $\Theta = \{1, 2\}$, and an experiment μ with a binary signal which equals the state with some probability $1/2 < r < 1$. For concreteness, imagine a coin whose probability of heads depends on the state and is either r or $1 - r$, and the experiment μ consists of tossing the coin. Consider a Bayesian LLR cost, with $b_{12} = b_{21} = b$. In this case, even though the effective (β_{ij})'s depend on the prior, a simple calculation shows that the cost of the experiment does not, and equals

$$C(\mu, q) = b(2r-1)\log\frac{r}{1-r}$$

for every prior q.[39]

Consider now the experiment ν in which the coin is tossed until a "heads" outcome. Under Bayesian LLR costs, the cost can be calculated to be

$$C(\nu, q) = \left(\frac{q_1}{r} + \frac{q_2}{1-r}\right) C(\mu, q).$$

This cost does depend on the prior: as the above display shows, it is equal to the cost of one toss of the coin, times the expected number of times that it is to be tossed. The latter quantity depends on the prior, in the obvious way. This cost is thus consistent with our additivity axiom, in the sense that this one-shot experiment ν—which is equivalent to a dynamic experiment in which μ is carried out a random number of times—has a cost that equals the expected number of repetition of μ, times the cost of each independent realization of μ.

We generalize the example of a biased coin toss to any experiment μ for which $D_{\mathrm{KL}}(\mu_1\|\mu_2) = D_{\mathrm{KL}}(\mu_2\|\mu_1)$. As the next proposition shows, this condition exactly captures prior independence of Bayesian LLR costs, in the symmetric case in which $b_{12} = b_{21}$.

PROPOSITION 12: *Let $\Theta = \{1, 2\}$. Let C be a uniformly posterior separable Bayesian LLR cost specified by $b_{12} = b_{21} = b > 0$. Let μ be a Blackwell experiment. Then the following are equivalent.*

1) $D_{\mathrm{KL}}(\mu_1\|\mu_2) = D_{\mathrm{KL}}(\mu_2\|\mu_1)$.

2) $C(\mu, q)$ is independent of the prior q.

PROOF:

Under the assumption that $b_{12} = b_{21} = b > 0$, the cost of an experiment μ at prior q is

$$C(\mu, q) = b\left[q_1 D_{\mathrm{KL}}(\mu_1\|\mu_2) + q_2 D_{\mathrm{KL}}(\mu_2\|\mu_1)\right].$$

Clearly, this quantity depends on q if and only if $D_{\mathrm{KL}}(\mu_1\|\mu_2) \neq D_{\mathrm{KL}}(\mu_2\|\mu_1)$.

[39]This contrasts with mutual information, where the prior affects the cost of this experiment: the cost is highest for the uniform prior, and vanishes as the prior tends towards certainty.