# Revealed Preference, Rational Inattention, and Costly Information Acquisition

Andrew Caplin and Mark Dean

Online Appendix: Not for Publication

## 1 Appendix 1: Proofs

### 1.1 Lemma 1

**Lemma 1** *If $\pi \in \Pi$ is consistent with $P_A \in P$, then it is sufficient for $\bar{\pi}_A$.*

**Proof.** Let $\pi \in \Pi$ be an information structure that is consistent with $P_A \in P$ . First, we list in order all distinct posteriors $\eta_i \in \Gamma(\pi)$ for $1 \leq i \leq I = |\Gamma(\pi)|$. Given that $\pi$ is consistent with $P_A$, there exists a corresponding optimal choice strategy $C_A : \Gamma(\pi) \to \Delta(A)$ such that the information structure and choice functions match the data,

$$P_A(a|\omega) = \sum_{i=1}^{I} \pi(\eta^i|\omega) C_A(a|\eta^i).$$

We also list in order all possible posteriors associated with the corresponding revealed information structure, $\gamma^j \in \bar{\Gamma} \equiv \Gamma(\bar{\pi}_A)$, $1 \leq j \leq |\bar{\Gamma}|$, and identify all chosen actions that are associated with posterior $\gamma^j$ as $\bar{F}^j$,

$$\bar{F}_j \equiv \{a \in A | \bar{\gamma}_A^a = \gamma^j\}.$$

The garbling matrix $b^{ij}$ sets the probability of $\gamma^j \in \bar{\Gamma}$ given $\eta^i \in \Gamma(\pi)$ as the probability of all choices associated with actions $a \in \bar{F}_j$

$$b^{ij} = \sum_{a \in \bar{F}_j} C_A(a|\eta^i).$$

Note that this is indeed a $|\Gamma(\pi)| \times |\bar{\Gamma}|$ stochastic matrix $B \geq 0$ with $\sum_{j=1}^{J} b^{ij} = 1$ all $i$. Given $\gamma^j \in \Gamma(\pi)$ and $\omega \in \Omega$, note that,

$$\sum_{i=1}^{I} b^{ij} \pi(\eta^i|\omega) = \sum_{i=1}^{I} \pi(\eta^i|\omega) \sum_{a \in \bar{F}_j} C_A(a|\eta^i) = \sum_{a \in \bar{F}_j} P_A(a|\omega),$$

by the data matching property. It is definitional that $\bar{\pi}_A(\gamma^j|\omega)$ is precisely equal to this, as the observed probability of all actions associated with posterior $\gamma^j \in \bar{\Gamma}$. Hence,

$$\bar{\pi}_A(\gamma^j|\omega) = \sum_{i=1}^{I} b^{ij} \pi(\eta^i|\omega),$$

as required for sufficiency. ■

## 1.2 Theorem 1 and 2

**Theorem 1** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set $(D, P)$ has a costly information representation if and only if it satisfies NIAS and NIAC.*

**Proof of Necessity.** Necessity of NIAS follows much as in CM15. Fix $A \in D$, $\pi_A \in \Pi$ and $C_A : \Gamma(\pi_A) \to \Delta(A)$ in a costly information representation, and $a \in Supp(P_A)$. By definition of a costly information representation,

$$\sum_{\gamma \in \Gamma(\pi_A)} C_A(a|\gamma) \left[ \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \right] \geq \sum_{\gamma \in \Gamma(\pi_A)} C_A(a|\gamma) \left[ \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \right] \text{ all } b \in A.$$

Substituting,

$$\gamma(\omega) = \frac{\mu(\omega)\pi_A(\gamma|\omega)}{\sum_{\upsilon \in \Omega} \mu(\nu)\pi_A(\gamma|\nu)},$$

cancelling the common denominator $\sum_{\upsilon \in \Omega} \mu(\nu)\pi_A(\gamma|\nu)$ in the inequality and substituting $P_A(a|\omega) = \sum_{\gamma \in \Gamma(\pi_A)} \pi_A(\gamma|\omega) C_A(a|\gamma)$, we derive,

$$\begin{aligned} \sum_{\omega \in \Omega} \mu(\omega) P_A(a|\omega) u(a(\omega)) &= \sum_{\gamma \in \Gamma(\pi_A)} C_A(a|\gamma) \left[ \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \right] \geq \\ \sum_{\gamma \in \Gamma(\pi_A)} C_A(a|\gamma) \left[ \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \right] &= \sum_{\omega \in \Omega} \mu(\omega) P_A(a|\omega) u(b(\omega)) \end{aligned}$$

establishing necessity of NIAS.

To confirm necessity of NIAC consider any sequence $A^1, A^2, .....A^J \in D$ with $A^J = A^1$ and (abusing notation slightly) corresponding information structure $\pi_j$ for $1 \leq j \leq J$. By optimality,

$$G(A^j, \pi_j) - K(\pi_j) \geq G(A^j, \pi_{j+1}) - K(\pi_{j+1}). \forall\ j \in \{1, .., J\},$$

so that,

$$\sum_{j=1}^{J-1} G(A^j, \pi_j) - K(\pi_j) \geq \sum_{j=1}^{J-1} G(A^j, \pi_{j+1}) - K(\pi_{j+1}).$$

Given that $K(\pi_1) = K(\pi_J)$, note that,

$$\sum_{j=1}^{J-1} G(A^j, \pi_j) - G(A^j, \pi_{j+1}) \geq \sum_{j=1}^{J-1} K(\pi_j) - K(\pi_{j+1}) = 0,$$

so that,

$$\sum_{j=1}^{J-1} G(A^j, \pi_j) \geq \sum_{j=1}^{J-1} G(A^j, \pi_{j+1}).$$

To establish that this is inherited by the revealed information structures $\bar{\pi}_j$ for $1 \leq j \leq J$, note from lemma 1 that $\pi_j$ is sufficient for $\bar{\pi}_j$, and so by Blackwell's Theorem (see Remark 1), $G(B, \pi_j) \geq G(B, \bar{\pi}_j)$ for all $B \in \mathcal{F}$. For $B = A^j$ this is an equality since both information structures give rise to the same state dependent stochastic demand,

$$G(A^j, \pi_j) = G(A^j, \bar{\pi}_j) = \sum_{a \in A^j} \sum_{\omega \in \Omega} \mu(\omega) P_{A^j}(a|\omega) u(a(\omega)).$$

Hence,

$$\sum_{j=1}^{J-1} G(A^j, \bar{\pi}_j) = \sum_{j=1}^{J-1} G(A^j, \pi_j) \geq \sum_{j=1}^{J-1} G(A^j, \pi_{j+1}) \geq \sum_{j=1}^{J-1} G(A^j, \bar{\pi}_{j+1}),$$

establishing NIAC. ■

**Proof of Sufficiency.** There are three steps in the proof that the NIAS and NIAC conditions are sufficient for $(D, P)$ to have a costly information representation. The first step is to establish that the NIAC conditions ensures that there is no global reassignment of the revealed information structures observed in the data to decision problems $A \in D$ that raises total gross surplus. The second step is use this observation to define a candidate cost function on information structures, $\bar{K} : \Pi \to \mathbb{R} \cup \infty$. The key is to note that, as the solution to the classical allocation problem of Koopmans and Beckmann [1957], this assignment is supported by "prices" set in expected utility units. It is these prices that define the proposed cost function. The final step is to apply the NIAS conditions to show that there exists a sequence of choice functions $\{C_A\}_{A \in D}$ such that $\left(\bar{K}, \{\bar{\pi}_A\}_{A \in D}, \{C_A\}_{A \in D}\right)$ form a costly information representation of $(D, P)$..

Enumerate decision problems in $D$ as $A^j$ for $1 \leq j \leq J$. Define the corresponding revealed information structures $\bar{\pi}_j$ for $1 \leq j \leq J$ as revealed in the corresponding data and let $\bar{\Pi} \equiv \cup_{j=1}^{J} \bar{\pi}_j$ be the set of all such structures across decision problems, with a slight enrichment to ensure that there are precisely as many structures as there are decision problems. If all revealed information structures are different, the set as just defined will have cardinality $J$. If there is repetition, then retain the decision problem index with which identical revealed information structures are associated so as to make them distinct. This ensures that the resulting set $\Pi$ has precisely $J$ elements. Index elements $\bar{\pi}_j \in \bar{\Pi}$ in order of the decision problem $A^j$ with which they are associated.

We now allow for arbitrary matchings of information structures to decision problems. First, let $\beta_{jl}$ denote the gross utility of decision problem $j$ combined with revealed information structure $l$,

$$\beta_{jl} = G(A^j, \bar{\pi}_l),$$

with $B$ the corresponding matrix. Define $\mathcal{M}$ to be the set of all matching functions $\theta : \{1, .., J\} \to \{1, .., J\}$ that are 1-1 and onto and identify the corresponding sum of gross payoffs,

$$S(\theta) = \sum_{j=1}^{J} \beta_{j\theta(j)}.$$

It is simple to see that the NIAC condition implies that the identify map $\theta^I(j) = j$ maximizes the sum over all matching functions $\theta \in \mathcal{M}$. Suppose to the contrary that there exists some alternative matching function that achieves a strictly higher sum, and denote this match $\theta^* \in \mathcal{M}$. In this case construct a first sub-cycle as follows: start with the lowest index $j_1$ such that $\theta^*(j_1) \neq j_1$. Define $j_2 = \theta^*(j_1)$ and now find $\theta^*(j_2)$, noting by construction that $\theta^*(j_2) \neq j_2$ as $\theta^*$ is one-to-one . Given that the domain is finite, this process will terminate after some $K \leq J$ steps with $\theta^*(j_K) = j_1$. If it is the case that $\theta^*(j) = j$ outside of the set $\cup_{k=1}^K j_k$, then we know the increase in the value of the sum is associated only with this cycle, hence,

$$\sum_{k=1}^{K-1} \beta_{j_k j_k} < \sum_{j=1}^{K-1} \beta_{j_k j_{k+1}},$$

directly in contradiction to NIAC. If this inequality does not hold strictly, then we know that there exists some $j'$ outside of the set $\cup_{k=1}^K j_k$ such that $\theta^*(j') \neq j'$. We can therefore iterate the process, knowing that the above strict inequality must be true for at least one such cycle to explain the strict increase in overall gross utility. Hence the identity map $\theta^I(j) = j$ indeed maximizes $S(\theta)$ amongst all matching functions $\theta \in \mathcal{M}$.

With this, we have established that the identity map solves an allocation problem of precisely the form analyzed by Koopmans and Beckmann [1957]. They characterize those matching functions $\theta : \{1, ..J\} \to \{1, ..J\}$ that maximize the sum of payoffs defined by a square payoff matrix such as $B$ that identifies the reward to matching objects of one set (decision problems in our case) to a corresponding number of objects in a second set (revealed information structures in our case). They show that the solution is the same as that of the linear program obtained by ignoring integer constraints,

$$\max_{x_{jl} \geq 0} \sum_{j,l} \beta_{jl} x_{jl} \text{ s.t. } \sum_{j=1}^{J} x_{jl} = \sum_{l=1}^{J} x_{jl} = 1.$$

Standard duality theory implies that the optimal assignment $\theta^I(j) = j$ is associated with a system of prices on revealed information structures such that the increment in net payoff from any move of any decision problem is not more than the increment in the cost of the information structure (see Koopmans and Beckmann [1957] pages 58-60).

Defining these prices as $\bar{K}_j$, their result implies that,

$$\beta_{jl} - \beta_{jj} = G(A^j, \bar{\pi}_l) - G(A^j, \bar{\pi}_j) \leq \bar{K}_l - \bar{K}_j;$$

or,

$$G(A_j, \bar{\pi}_j) - \bar{K}_j \geq G(A_j, \bar{\pi}_l) - \bar{K}_l.$$

The result of Koopmans and Beckmann therefore implies existence of a function $\bar{K} : \Pi \longrightarrow \mathbb{R}$ that decentralizes the problem from the viewpoint of the owner of the decision problems, seeking to identify surplus maximizing information structures to match to their particular problems. Note that if there are two distinct decision problems with the same revealed posterior, the result directly implies that they must have the same cost, so that one can in fact ignore the reference to the decision problem and retain only the posterior in the domain. Set $K(\pi) = \bar{K}_i$ if $\pi = \bar{\pi}_i$ for some $A^i \in D$, and $K(\pi) = \infty$ if $\pi \notin \bar{\Pi}$.

We have now completed construction of a qualifying cost function $\bar{K} : \Pi \to \mathbb{R} \cup \infty$ that satisfies $\bar{K}(\pi) \in \mathbb{R}$ for some $\pi \in \Pi$. By construction the observed information structure choices were always maximal: $\bar{\pi}^A \in \hat{\Pi}(K, A)$ for all $A \in D$. It remains to prove that $\bar{\pi}^A$ is consistent with $P_A$ for all $A \in D$. To do so, we must construct, for each $A \in D$, a function $C_A : \Gamma(\bar{\pi}_A) \to \Delta(A)$ such that (i) $\bar{\pi}_A$ and $C_A$ generate $P_A$, and (ii) for all $a \in A$ and $\gamma \in \Gamma(\pi_A)$ with $C_A(a|\gamma) \equiv \Pr(a|\gamma) > 0$,

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \text{ all } b \in A.$$

For each $\gamma \in \Gamma(\bar{\pi}_A)$, define:

$$C_A(a|\gamma) = \begin{cases} \frac{P_A(a)}{\sum_{\left\{b \in A | \bar{\gamma}_A^b = \gamma\right\}} P_A(b)} \text{ if } \bar{\gamma}_A^a = \gamma; \\ 0 \text{ otherwise;} \end{cases}$$

where $P_A(a)$ is the unconditional probability of choosing action $a$ in set $A$. Note that this defines a probability distribution over $A$ for each $\gamma \in \Gamma(\bar{\pi}_A)$. Note also that $C_A(a|\gamma) > 0$ only if $\bar{\gamma}_A^a = \gamma$. The NIAS condition implies that

$$\begin{aligned} \sum_{\omega \in \Omega} \mu(\omega) P_A(a|\omega) u(a(\omega)) &\geq \sum_{\omega \in \Omega} \mu(\omega) P_A(a|\omega) u(b(\omega)) \Rightarrow \\ \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(a(\omega)) &\geq \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(b(\omega)). \end{aligned}$$

The second line follows by dividing both sides by $P_A(a)$. Thus, NIAS ensures that the choice function is optimal.

It remains only to show that $\bar{\pi}_A$ and $C_A$ generate $P_A$. To see this, first note that, for any two $b, b' \in A$ such that $\bar{\gamma}_A^b = \bar{\gamma}_A^{b'}$, Bayes' rule implies,

$$\frac{P_A(b|\omega)}{P_A(b'|\omega)} = \frac{P_A(b)}{P_A(b')} \;\forall\; \omega \in \Omega \text{ s.t. } \bar{\gamma}_A^b(\omega) \neq 0. \tag{1}$$

Next note that, for every $\omega \in \Omega$ and $a \in A$ such that $P_A(a) > 0$

$$\begin{aligned} &\sum_{\gamma \in \Gamma(\bar{\pi}_A)} \bar{\pi}_A(\gamma|\omega) C_A(a|\gamma) = \bar{\pi}_A(\bar{\gamma}_A^a|\omega) C_A(a|\bar{\gamma}_A^a) = \sum_{\{c \in A | \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a)}{\sum_{\left\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\right\}} P_A(b)} \\ &= \sum_{\{c \in A | \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a|\omega)}{\sum_{\left\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\right\}} P_A(b|\omega)} = P_A(a|\omega), \end{aligned}$$

ensuring data matching. The penultimate equality follows from equation 1, while the final inequality comes from noting that the sums in the numerator and the denominator are identical. ■

## 1.3 Theorem 2

**Theorem 2** *Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, data set $(D, P)$ satisfies NIAS and NIAC if and only if it has a costly information representation with conditions K1 to K3 satisfied.*

**Proof.** The proof of necessity is immediate from theorem 1. The proof of sufficiency proceeds in four steps, starting with a costly information representation $\left(\bar{K}, \bar{\pi}\right)$ of $(D, P)$ of the form produced in theorem 1 based on satisfaction of the NIAS and NIAC conditions. A key feature of this function is that the function $\bar{K}$ is real-valued only on the revealed information structures $\bar{\Pi} \equiv \{\bar{\pi}_A | A \in D\}$ associated with all corresponding decision problems, otherwise being infinite. The first step is the proof is to construct a larger domain $\mathring{\Pi} \supset \bar{\Pi}$ to satisfy three additional properties: to include the inattentive strategy, $I \in \mathring{\Pi}$; to be closed under mixtures so that $\pi, \eta \in \mathring{\Pi}$ and $\alpha \in (0, 1)$ implies $\alpha \circ \pi + (1 - \alpha) \circ \eta \in \mathring{\Pi}$; and to be "closed under garbling," so that if $\pi \in \mathring{\Pi}$ is sufficient for information structure $\rho \in \Pi$, then $\rho \in \mathring{\Pi}$. The second step is to define a new function $\mathring{K}$ that preserves the essential elements of $\bar{K}$ while being real-valued on the larger domain $\mathring{\Pi} \supset \bar{\Pi}$, and thereby to construct the full candidate cost function $\mathring{K} : \Pi \to \mathbb{R} \cup \infty$. The third step is to confirm that $\mathring{K} \in \mathcal{K}$ and that $\mathring{K}$ satisfies the required conditions K1 through K3. The final step is to confirm that $\left(\mathring{K}, \{\bar{\pi}_A\}_{A \in D}\right)$ forms a costly information representation of $(D, P)$.

We construct the domain $\mathring{\Pi}$ in two stages. First, we define all information structures for which some revealed information structure $\bar{\pi} \in \Pi$ is sufficient;

$$\bar{\Pi}_S = \{\rho \in \Pi | \exists \pi \in \bar{\Pi} \text{ sufficient for } \rho\}.$$

Note that this is a superset of $\bar{\Pi}$ and that it contains $I$. The second step is to identify $\mathring{\Pi}$ as the smallest mixture set containing $\bar{\Pi}_S$: this is itself a mixture set since the arbitrary intersection of mixture sets is itself a mixture set.

By construction, $\mathring{\Pi}$ has three of the four desired properties: it is closed under mixing; it contains $\bar{\Pi}$, and it contains the inattentive strategy. The only condition that needs to be checked is that it retains the property of being closed under garbling:

$$\pi \in \mathring{\Pi} \text{ sufficient for } \rho \in \Pi \Longrightarrow \rho \in \mathring{\Pi} \ .$$

To establish this, it is useful first to establish certain properties of $\bar{\Pi}_S$ and of $\mathring{\Pi}$. The first is that $\bar{\Pi}_S$ is closed under garbling:

$$\pi \in \bar{\Pi}_S \text{ sufficient for } \rho \in \Pi \Longrightarrow \rho \in \bar{\Pi}_S.$$

Intuitively, this is because the garbling of a garbling is a garbling. In technical terms, the product of the corresponding garbling matrices is itself a garbling matrix. The second is that one can explicitly express $\mathring{\Pi}$ as the set of all finite mixtures of elements of $\bar{\Pi}_S$,

$$\mathring{\Pi} = \left\{ \pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j | J \in \mathbb{N}, (\lambda_1, ..\lambda_J) \in S^{J-1}, \pi_j \in \bar{\Pi}_S \right\},$$

where $S^{J-1}$ is the unit simplex in $\mathbb{R}^{J.}$ To make this identification, note that the set as defined on the RHS certainly contains $\bar{\Pi}_S$ and is a mixture set, hence is a superset of $\mathring{\Pi}$. Note moreover that all elements in the RHS set are necessarily contained in any mixture set containing $\bar{\Pi}_S$ by a process of iteration, making is also a subset of $\mathring{\Pi}$, hence finally one and the same set.

We now establish that if $\rho \in \Pi$ is a garbling of some $\pi \in \mathring{\Pi}$, then indeed $\rho \in \mathring{\Pi}$. The first step is to express $\pi \in \mathring{\Pi}$ as an appropriate convex combination of elements of $\bar{\Pi}_S$ as we now know we can,

$$\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j.$$

with all weights strictly positive, $\lambda_j > 0$ all $j$. Lemma 2 below establishes that in this case there exist garblings $\rho_j$ of $\pi_j \in \bar{\Pi}_S$ such that,

$$\rho = \sum_{j=1}^{J} \lambda_j \circ \rho_j,$$

establishing that indeed $\rho \in \mathring{\Pi}$ since, with $\bar{\Pi}_S$ closed under garbling, $\pi_j \in \bar{\Pi}_S$ and $\rho_j$ a garbling of $\pi_j$ implies $\rho_j \in \bar{\Pi}_S$.

We define the function $\mathring{K}$ on $\mathring{\Pi}$ in three stages. First we define the function $\bar{K}_S$ on the domain

$\bar{\Pi}_S$ by identifying for any $\rho \in \bar{\Pi}_S$ the corresponding set of revealed information structures $\bar{\pi} \in \bar{\Pi}$ of which $\rho$ is a garbling, and assigning to it the lowest such cost. Formally, given $\rho \in \bar{\Pi}_S$,

$$\bar{K}_S(\rho) \equiv \min_{\{\pi \in \bar{\Pi} | \pi \text{ sufficient for } \rho\}} \bar{K}(\pi).$$

Note that $\bar{K}_S(\pi) = \bar{K}(\pi)$ all $\pi \in \bar{\Pi}$. To see this, consider $A, A' \in D$ with $\bar{\pi}_{A'}$ sufficient for $\bar{\pi}_A$. By the Blackwell property, expected utility is at least as high using $\bar{\pi}_{A'}$ as using $\bar{\pi}_A$ for which it is sufficient,

$$G(A, \bar{\pi}_{A'}) \geq G(A, \bar{\pi}_A).$$

At the same time, since $(\bar{K}, \bar{\pi})$ forms part of a costly information representation of $(D, P)$, we know that $\bar{\pi}_A \in \hat{\Pi}(K, A)$, so that,

$$G(A, \bar{\pi}_A) - K(\bar{\pi}_A) \geq G(A, \bar{\pi}_{A'}) - K(\bar{\pi}_{A'}).$$

Together these imply that $K(\bar{\pi}_A) \leq K(\bar{\pi}_{A'})$, which in turn implies that $\bar{K}_S(\pi) = \bar{K}(\pi)$ all $\pi \in \bar{\Pi}$.

Note that $\bar{K}_S(\pi)$ also satisfies weak monotonicity on this domain, since if we are given $\rho, \eta \in \bar{\Pi}_S$ with $\rho$ sufficient for $\eta$, then we know that any information structure $\pi \in \bar{\Pi}$ that is sufficient for $\rho$ is also sufficient for $\eta$, so that the minimum defining $\bar{K}_S(\rho)$ can be no lower than that defining $\bar{K}_S(\eta)$.

The second stage in the construction is to extend the domain of the cost function from $\bar{\Pi}_S$ to $\mathring{\Pi}$. As noted above, this set comprises all finite mixtures of elements of $\bar{\Pi}_S$. Given $\pi \in \mathring{\Pi}$, we take the set of all such mixtures that generate it and define $\mathring{K}(\pi)$ to be the corresponding infimum,

$$\mathring{K}(\pi) = \inf_{\left\{ J \in \mathbb{N}, \lambda \in S^{J-1}, \{\pi^j\}_{j=1}^J \in \bar{\Pi}_S | \pi = \sum_{j=1}^J \lambda_j \circ \pi_j \right\}} \sum_{j=1}^J \lambda_j \bar{K}_S(\pi_j).$$

Note that this function is well defined since $\bar{K}_S$ is bounded below by the lowest cost in $\bar{K}$ and the feasible set is non-empty by definition of $\mathring{\Pi}$. We establish in Lemma 3 that the infimum is achieved. Hence, given $\pi \in \mathring{\Pi}$, there exists $J \in \mathbb{N}, \lambda \in S^{J-1}$, and elements $\pi_j \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^J \lambda_j \circ \pi_j$ such that,

$$\mathring{K}(\pi) = \sum_{j=1}^J \lambda_j \bar{K}_S(\pi_j).$$

We show now that $\mathring{K}$ satisfies K2, mixture feasibility. Consider distinct structures $\pi \neq \eta \in \mathring{\Pi}$. We know by Lemma 3 that we can find $J^{\pi,\eta} \in \mathbb{N}$, corresponding probability weights $\lambda^{\pi,\eta} \in S^{\pi,\eta}$ and elements $\eta_j, \pi_j \in \bar{\Pi}_S$ with $\eta = \sum_{j=1}^{J^\eta} \lambda_j^\eta \circ \eta_j$, $\pi = \sum_{j=1}^{J^\pi} \lambda_j^\pi \circ \pi_j$, and such that,

$$\begin{aligned} \mathring{K}(\eta) &= \sum_{j=1}^{J^\eta} \lambda_j^\eta \bar{K}_S(\eta_j); \\ \mathring{K}(\pi) &= \sum_{j=1}^{J^\pi} \lambda_j^\pi \bar{K}_S(\pi_j). \end{aligned}$$

Given $\alpha \in (0,1)$, consider now the mixture strategy defined by taking each strategy $\pi_j$ with probability $\alpha\lambda_j^\pi$ and each strategy $\eta_j$ with probability $(1-\alpha)\lambda_j^\eta$. By construction, this mixture strategy generates $\psi = [\alpha \circ \pi + (1-\alpha) \circ \eta] \in \Pi$ and hence we know by the infimum feature of $\mathring{K}(\psi)$ that,

$$\mathring{K}(\psi) \leq \sum_{j=1}^{J^\pi} \alpha\lambda_j^\pi \bar{K}_S(\pi_j) + \sum_{j=1}^{J^\eta}(1-\alpha)\lambda_j^\eta \bar{K}_S(\eta_j) = \alpha\mathring{K}(\pi) + (1-\alpha)\mathring{K}(\eta),$$

confirming mixture feasibility.

We show also that $\mathring{K}$ satisfies K1, weak monotonicity in information. Consider $\pi, \eta \in \mathring{\Pi}$ with $\pi$ sufficient for $\eta$. We know by Lemma 3 that we can find $J \in \mathbb{N}$, $\lambda \in S^{J-1}$, and corresponding elements $\{\pi_j\}_{j=1}^J \in \bar{\Pi}_S$ such that $\pi = \sum_{j=1}^J \lambda_j \circ \pi_j$ and such that,

$$\mathring{K}(\pi) = \sum_{j=1}^J \lambda_j \bar{K}_S(\pi_j).$$

We know also from Lemma 2 that we can construct $\{\eta_j\}_{j=1}^J \in \bar{\Pi}_S$ such that $\eta = \sum_{j=1}^J \lambda_j \circ \eta_j$ and such that each $\eta_j$ is a garbling of the corresponding $\pi_j$. Given that $\bar{K}_S$ satisfies weak monotonicity on its domain $\bar{\Pi}_S$, we conclude that,

$$\bar{K}_S(\pi_j) \geq \bar{K}_S(\eta_j) \ \forall\ j \in \{1, ..., J\}$$

By the infimum feature of $\mathring{K}(\eta)$ we therefore know that,

$$\mathring{K}(\eta) \leq \sum_{j=1}^J \lambda_j \bar{K}_S(\eta_j) \leq \sum_{j=1}^J \lambda_j \bar{K}_S(\pi_j) = \mathring{K}(\pi),$$

confirming weak monotonicity.

We show now that we have retained the properties that made $(\bar{K}, \bar{\pi})$ a costly information representation of $(D, P)$. Given $A \in D$, it is immediate that $\bar{\pi}$ and the choice function constructed in the proof of Theorem 1 is consistent with the data, since this was part of the initial definition. What needs to be confirmed is only that the revealed revealed information structures are optimal. Suppose to the contrary that there exists $A \in D$ such that,

$$G(A, \pi) - \mathring{K}(\pi) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A),$$

for some $\pi \in \mathring{\Pi}$. By Lemma 3 we can find $J \in \mathbb{N}$, a strictly positive vector $\lambda \in S^{J-1}$, and corresponding elements $\{\pi_j\}_{j=1}^J \in \bar{\Pi}_S$, such that $\pi = \sum_{j=1}^J \lambda_j \circ \pi_j$ and such that,

$$\mathring{K}(\pi) = \sum_{j=1}^J \lambda_j \bar{K}_S(\pi_j).$$

By the fact that $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ and by construction of the mixture strategy,

$$G(A, \pi) = \sum_{j=1}^{J} \lambda_j G(A, \pi_j),$$

so that,

$$\sum_{j=1}^{J} \lambda_j \left[ G(A, \pi_j) - \bar{K}_S(\pi_j) \right] > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

We conclude that there exists $j$ such that,

$$G(A, \pi_j) - \bar{K}_S(\pi_j) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

Note that each $\pi_j \in \bar{\Pi}_S$ inherits its cost $\bar{K}_S(\pi_j)$ from an element $\bar{\pi}_j \in \bar{\Pi}$ that is the lowest cost revealed information structure according to $\bar{K}$ on set $\bar{\Pi}$ that is sufficient for $\pi_j$,

$$\bar{K}_S(\pi_j) = \bar{K}(\bar{\pi}_j),$$

where the last equality stems from the fact (established above) that $\bar{K}_S(\pi) = \bar{K}(\pi)$ on $\bar{\pi} \in \bar{\Pi}$. Note by the Blackwell property that each strategy $\bar{\pi}_j \in \bar{\Pi}$ offers at least as high gross value as the strategy $\pi_j \in \bar{\Pi}_S$ for which it is sufficient, so that overall,

$$G(A, \bar{\pi}_j) - \bar{K}(\bar{\pi}_j) \geq G(A, \pi_j) - \bar{K}_S(\pi_j) > G(A, \bar{\pi}_A) - \mathring{K}(\bar{\pi}_A).$$

To complete the proof it is sufficient to show that,

$$\mathring{K}(\pi) = \bar{K}(\pi),$$

on $\pi \in \bar{\Pi}$. With this we derive the contradiction that,

$$G(A, \bar{\pi}_j) - \bar{K}(\bar{\pi}_j) > G(A, \bar{\pi}_A) - \bar{K}(\bar{\pi}_A),$$

in contradiction to the assumption that $\left(\bar{K}, \bar{\pi}\right)$ formed a costly information representation of $(D, P)$.

To establish that $\mathring{K}(\pi) = \bar{K}(\pi)$ on $\pi \in \bar{\Pi}$, note that we know already that $\bar{K}_S(\pi) = \bar{K}(\pi)$ on $\bar{\pi} \in \bar{\Pi}$. If this did not extend to $\mathring{K}(\pi)$, then we would be able to identify a mixture strategy $\psi \in \bar{\Pi}$ sufficient for $\bar{\pi}_A$ with strictly lower expected costs, $\mathring{K}(\psi) < \mathring{K}(\pi)$. To see that this is not possible, note first from Lemma 1 that all structures that are consistent with $A$ and $P_A$ are sufficient for $\bar{\pi}_A$. Weak monotonicity of $\mathring{K}$ on $\mathring{\Pi}$ then implies that the cost $\mathring{K}(\psi)$ of any mixture strategy sufficient for $\bar{\pi}_A$ satisfies $\mathring{K}(\psi) \geq \mathring{K}(\pi)$, as required.

The final and most trivial stage of the proof is to ensure that normalization (K3) holds. Note that $I \in \bar{\Pi}_S$, so that $\mathring{K}_S(I) \in \mathbb{R}$ according to the rule immediately above. If we renormalize this function by subtracting $\mathring{K}(I)$ from the cost function for all information structures then we impact on no margin of choice and do not interfere with mixture feasibility, weak monotonicity, or whether or not we have a costly information representation. Hence we can avoid pointless complication by assuming that $\mathring{K}(I) = 0$ from the outset so that this normalization is vacuous. In full, we define the candidate cost function $\overset{\circ}{K} : \overset{\circ}{\Pi} \to \mathbb{R} \cup \infty$ by,

$$\overset{\circ}{K}(\pi) = \begin{cases} \overset{\circ}{K}(\pi) \text{ if } \pi \in \overset{\circ}{\Pi} \\ \infty \text{ if } \pi \notin \overset{\circ}{\Pi}. \end{cases}$$

Note that weak monotonicity implies that the function is non-negative on its entire domain.

It is immediate that $\overset{\circ}{K} \in \mathcal{K}$, since $\overset{\circ}{K}(\pi) = \infty$ for $\pi \notin \overset{\circ}{\Pi}$ and the domain contains the corresponding inattentive strategy $I$ on which $\overset{\circ}{K}(\pi)$ is real-valued. It is also immediate that $\overset{\circ}{K}$ satisfies K3, since $\overset{\circ}{K}(I) = 0$ by construction. It also satisfies K1 and K2, and represents a costly information representation, completing the proof. ■

**Lemma 2** If $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ with $J \in \mathbb{N}$, $\lambda \in S^{J-1}$ with $\lambda_j > 0$ all $j$, and $\{\pi_j\}_{j=1}^{J} \in \Pi$, then for any garbling $\rho$ of $\pi$, there exist garblings $\rho_j$ of $\pi_j \in \Pi$ such that,

$$\rho = \sum_{j=1}^{J} \lambda_j \circ \rho_j,$$

**Proof.** By assumption, there exists a $|\Gamma(\pi)| \times |\Gamma(\rho)|$ matrix $B$ with $\sum_k b^{ik} = 1$ all $i$ and such that, for all $\gamma^k \in \Gamma(\rho)$,

$$\rho(\gamma^k|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \pi(\eta^i|\omega).$$

Since $\pi = \sum_{1}^{J} \lambda_j \circ \pi_j$, we know that $\Gamma(\pi^j) \subset \Gamma(\pi)$. Now define compressed matrix $B^j$ as the unique submatrix of $B$ obtained by first deleting all rows corresponding to posteriors $\eta^i \in \Gamma(\pi) \backslash \Gamma(\pi_j)$, and then deleting all columns corresponding to posteriors $\gamma^k$ such that $b^{ik} = 0$ all $\eta^i \in \Gamma(\pi_j)$. Define $\rho_j \in \Pi$ to be the strategy that has as its support the set of all posteriors that are possible given the garbling $\rho_j$ of $\pi_j$,

$$\Gamma(\rho_j) = \{\gamma^k \in \Gamma(\rho) | b^{ik} > 0 \text{ some } \eta^i \in \Gamma(\pi_j)\},$$

and in which state dependent probabilities of all posteriors are generated by the compressed matrix $B^j$,

$$\rho^j(\gamma^k|\omega) = \sum_{\eta^i \in \Gamma(\pi^j)} b^{ik} \pi_i(\eta^i|\omega),$$

for all $\gamma^k \in \Gamma(\rho_j)$.

Note by construction that each information structure $\rho_j$ is a garbling of the corresponding $\pi_j \in \Pi$, since each $B^j$ is itself a garbing matrix for which $\sum_k b^{ik} = 1$ for all $\eta^i \in \Gamma(\pi_j)$. It remains only to verify that $\rho = \sum_{j=1}^{J} \lambda_j \circ \rho_j$. This follows since,

$$\rho(\gamma^k|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \pi(\eta^i|\omega) = \sum_{\eta^i \in \Gamma(\pi)} b^{ik} \sum_{j=1}^{J} \lambda_j \pi_j(\eta^i|\omega) = \sum_{j=1}^{J} \lambda_j \sum_{\eta^i \in \Gamma(\pi^j)} b^{ik} \pi_j(\eta^i|\omega) = \sum_{j=1}^{J} \lambda_j \rho_j(\gamma^k|\omega).$$

■

**Lemma 3** Given $\pi \in \mathring{\Pi}$, there exists $J \in \mathbb{N}$, $\lambda \in S^{J-1}$, and elements $\pi_j \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ such that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j).$$

**Proof.** By definition $\mathring{K}(\pi)$ is the infimum of $\sum_{j=1}^{J} \lambda_j \bar{K}_S(\pi_j)$ over all lists $\{\pi_j\}_{j=1}^{J} \in \bar{\Pi}_S$ such that $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$. We now consider a sequence of such lists, indicating the order in this sequence in parentheses, $\{\pi_j(n)\}_{j=1}^{J(n)}$, such that in all cases there are corresponding weights $\lambda(n) \in S^{J(n)-1}$ with $\pi = \sum_{j=1}^{J(n)} \lambda_j(n) \circ \pi_j(n)$ and that achieve a value that is heading in the limit to the infimum,

$$\lim_{n \longrightarrow \infty} \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_j(n)) = \mathring{K}(\pi).$$

A first issue that we wish to avoid is limitless growth in the cardinality $J(n)$. The first key observation is that, by Charateodory's theorem, we can reduce the number of strictly positive weights in a convex combination $\pi = \sum_{j=1}^{J^*(n)} \lambda_j^*(n) \circ \pi_j(n)$ to have cardinality $J^*(n) \leq M+1$. We confirm now that we can do this without raising the corresponding costs, $\sum_{j=1}^{J^*(n)} \lambda_j^*(n) \bar{K}_S(\pi_j(n))$. Suppose that there is some integer $n$ such that the original set of information structures has strictly higher cardinality $J(n) > M+1$. Suppose further that the first selection of $J^1(n) \leq M+1$ such posteriors for which there exists a strictly positive probability weights $\delta_j^1(n)$ such that $\pi = \sum_{j=1}^{J^1(n)} \delta_j^1(n) \circ \pi_j(n)$ has higher such costs (note WLOG that we are treating these as the first $J^1(n)$ information structures in the original list). It is convenient to define $\delta_j^1(n) = 0$ for $J^1(n) + 1 \leq j \leq J(n)$ so that we can express this inequality in the simplest terms,

$$\sum_{j=1}^{J(n)} \delta_j^1(n) \bar{K}_S(\pi_j(n)) > \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_j(n)).$$

This inequality sets up an iteration. We first take the smallest scalar $\alpha^1 \in (0,1)$ such that,

$$\alpha^1 \delta_j^1(n) = \lambda_j(n).$$

That such a scalar exists follows from the fact that $\sum_{j=1}^{J^1(n)} \delta_j^1(n) = \sum_{j=1}^{J(n)} \lambda_j(n) = 1$, with all components in both sums strictly positive and with $J(n) > J^1(n)$. We now define a second set of probability weights $\lambda_j^2(n)$,

$$\lambda_j^2(n) = \frac{\lambda_j(n) - \alpha^1 \delta_j^1(n)}{1-\alpha^1},$$

for $1 \leq j \leq J(n)$. Note that these weights have the property that $\pi = \sum_{j=1}^{J(n)} \lambda_j^2(n) \circ \pi_j(n)$ and that,

$$\sum_{j=1}^{J(n)} \lambda_j^2(n) \bar{K}_S(\pi_j(n)) = \sum_{j=1}^{J(n)} \left[ \frac{\lambda_j(n) - \alpha^1 \delta_j^1(n)}{1-\alpha^1} \right] \bar{K}_S(\pi_j(n)) < \sum_{j=1}^{J(n)} \lambda_j(n) \bar{K}_S(\pi_k(n)).$$

By construction, note that we have reduced the number of strictly positive weights $\lambda_j^2(n)$ by at least one to $J(n)-1$ or less. Iterating the process establishes that indeed there exists a set of no more than $M+1$ posteriors such that a mixture produces that first strategy $\pi$ and in which this mixture has no higher weighted average costs than the original strategy. Given this, there is no loss of generality in assuming that $J(n) \leq M+1$ in our original sequence.

With this bound on cardinality, we know that we can find a subsequence of information structures $\pi_j(n)$ which all have precisely the same cardinality $J(n) = J \leq M+1$ all $n$. Going further, we can impose properties on all of the $J$ corresponding sequences $\{\pi_j(n)\}_{n=1}^{\infty}$. First, we can select subsequences in which the ranges of all corresponding information structures have the same cardinality independent of $n$,

$$|\Gamma(\pi_j(n))| = H^j,$$

for $1 \leq j \leq J$. Note we can do this because, for all $j$ and $n$, the number of posteriors in the information structure $\pi_j(n)$ is bounded above by the number of posteriors in the strategy $\pi$, which is finite.

With this, we can index the possible posteriors $\gamma^{jh}(n) \in \Gamma(\pi_j(n))$ in order, $1 \leq h \leq H^j$ and then select further subsequences in which these posteriors themselves converge to limit posteriors,

$$\gamma^{jh}(L) = \lim_{n \to \infty} \gamma^{jh}(n) \in \Gamma.$$

which is possible since posteriors lie in a compact set, and so have a convergent subsequence.

We ensure also that both the associated state dependent probabilities themselves and the weights $\lambda_j(n)$ in the expression $\pi = \sum_{j=1}^{J(n)} \lambda_j(n) \circ \pi_j(n)$ converge,

$$\begin{aligned} \lim_{n \to \infty} \pi \left( \gamma^{jh}(n) | \omega \right) &= \pi_{jh}(L|\omega); \\ \lim_{n \to \infty} \lambda_j(n) &= \lambda_j(L). \end{aligned}$$

Again, this is possible because the state dependent probabilities and weights lie in compact sets.

The final selection of a subsequence ensures that, given $1 \leq j \leq J$, each $\pi_j(n) \in \bar{\Pi}_S$ has its value defined by precisely the same revealed information structure $\bar{\pi}_j \in \bar{\Pi}$ as the least expensive among those that were sufficient for it and hence whose cost it was assigned in function $\bar{K}_S$. Technically, for each $1 \leq j \leq J$,

$$\bar{K}_S(\pi_j(n)) = \bar{K}(\bar{\pi}_j),$$

for $1 \leq n \leq \infty$: this is possible because the data set and hence the number of revealed information structures is finite.

We first use these limit properties to construct a list of limit information structures $\pi_j(L) \in \bar{\Pi}_S$ with $\pi = \sum_{j=1}^{J} \lambda_j \circ \pi_j$ for $1 \leq j \leq J$. Strategy $\pi_j(L)$ has range,

$$\Gamma(\pi_j(L)) = \cup_{h=1}^{H^j} \gamma^{jh}(L),$$

with state dependent probabilities,

$$[\pi_j(L)]_\omega \left(\gamma^{jh}(L)\right) = \pi_{ih}(L|\omega).$$

Note that the construction ensures that $\pi = \sum_{j=1}^{J} \lambda_j(L) \circ \pi_j(L)$. To complete the proof we must establish only that,

$$\mathring{K}(\pi) = \sum_{j=1}^{J} \lambda_j(L) \bar{K}_S(\pi_j(L)).$$

We know from the construction that, for each $n$,

$$\sum_{j=1}^{J} \lambda_j(n) \bar{K}_S(\pi_j(n)) = \sum_{j=1}^{J} \lambda_j(n) \bar{K}(\bar{\pi}_j).$$

Hence the result is established provided only,

$$\bar{K}_S(\pi_j(L)) \leq \bar{K}(\bar{\pi}_j),$$

which is true provided $\bar{\pi}_j$ being sufficient for all $\pi_j(n)$ implies that $\bar{\pi}_j$ is sufficient for the corresponding limit vector $\pi^j(L)$. That this is so follows by defining $B^j(L) = [b^{ih}(L)]^j$ to be the limit of any subsequence of the $|\Gamma(\bar{\pi}_j)| \times H^j$ stochastic matrices $B^j(n) = [b^{ih}(n)]^j$ which have the defining property of sufficiency,

$$[\pi_j(n)]_\omega (\gamma^{jh}(n)) = \sum_{\bar{\gamma}^i \in \Gamma(\bar{\pi}_j)} [b^{ih}(n)]^j \circ \bar{\pi}(\bar{\gamma}^i|\omega),$$

for all $\gamma^{jh}(n) \in \Gamma(\pi_j(n))$ and $\omega \in \Omega$. It is immediate that the equality holds up in the limit, establishing that indeed $\bar{\pi}_j$ is sufficient for each corresponding limit vector $\pi_j(L)$, confirming finally that $\bar{K}_S(\pi_j(L)) \leq \bar{K}(\bar{\pi}_j)$ and with it establishing the Lemma. ■

# 2 Appendix 2: No Strong Blackwell

A simple example with data on one decision problem with two equally likely states illustrates that one cannot further strengthen the result in this dimension. Suppose that there are three available actions $A = \{a, b, c\}$ with corresponding utilities,

$$(u(a(\omega_1)), u(a(\omega_2))) = (10, 0)\,; (u(b(\omega_1)), u(b(\omega_2))) = (0, 10)\,; (u(c(\omega_1)), u(c(\omega_2))) = (7.5, 7.5)\,.$$

Consider the following state dependent stochastic choice data in which the only two chosen actions are $a$ and $b$,

$$P_A(a|\omega_1) = P_A(b|\omega_2) = \frac{3}{4} = 1 - P_A(b|\omega_1) = 1 - P_A(a|\omega_1).$$

Note that this data satisfies NIAS; given posterior beliefs when $a$ is chosen, $a$ is superior to $b$ and indifferent to $c$, and when $b$ is chosen it is superior to $a$ and indifferent to $c$. It trivially satisfies NIAC since there is only one decision problem observed. We know from theorem 2 that is has a costly information representation with the cost of the revealed information structure $K(\bar{\pi}) \geq 0$ and that of the inattentive strategy being zero, $K(I) = 0$. Note that $\bar{\pi}$ is sufficient for $I$ but not vice versa, hence any strictly monotone cost function would have to satisfy $K(\bar{\pi}) > 0$. In fact it is not possible to find a representation with this property. To see this, note that both structures have the same gross utility,

$$G(A, \pi) = \frac{1}{2} * \frac{3}{4} * 10 + \frac{1}{2} * \frac{3}{4} * 10 = 1 * 7.5 = G(A, I),$$

where we use the fact that the inattentive strategy involves picking action $c$ for sure. In order to rationalize selection of the inattentive strategy, it must therefore be that $\bar{\pi}$ is no more expensive than $I$, contradicting strict monotonicity.

## References

Tjalling C. Koopmans and Martin Beckmann. Assignment problems and the location of economic activities. *Econometrica*, 25(1):53–76, 1957.