ECON0106: **Microeconomics**

17. An Introduction to Information Orders*

Duarte Gonçalves[†]

University College London

1. Overview

Information is a key input in the decision-making process.

- News outlets **gather and process information** about current events and report them to the public.
- The public in turn **uses this information** to decide which political party to vote for, which stocks to invest in, and which products to by.
- Firms **produce information** through their R&D departments so as to develop new, cheaper methods of production — detecting absenteeism, machine failure, or automate decision-processes —- or innovate on goods and services they produce — smarter phones, self-driving cars, ride-sharing apps, life-saving medical drugs, etc. Researchers also generate information, having to decide which research questions to pursue, which methods to use, and which results to report.
- The **design of information** is also often relevant: think about the decision of which macro variables to measure, of deciding on which questions to use in an exam, the criteria underlying credit score ratings, etc.
- As the is no scarcity of information, it is fundamental to understand how agents **choose between sources of information**: think about the decision of which newspaper to read (and its implications on political polarisation), which financial consultant or wealth manager to hire (and the consequences for wealth inequality), which advisor to choose, etc.
- The **transmission of information** is an increasingly important concern, not only because of strategic aspects and effective communication (strategic disclosure of financial

*Last updated: 23 September 2025.

[†] Department of Economics, University College London; duarte.goncalves@ucl.ac.uk. Please do not share these notes with people outside of this class.

reports or public statistics, commercial advertising and political campaigning) — especially when attention is scarce — but also because of privacy concerns (encryption, data protection).

- Relatedly, the **diffusion of information** has always been a key point of interest, rising to prominence with social networks and the spread of misinformation, or the spread of market rumours and its impact on market volatility.
- Finally, since information is bought and sold, how should information be **priced**?

There is a wealth of application of information economics covering all fields and topics in economics — and the list above is by no means exhaustive. So these are just a few of the many different questions that may arise when thinking about information and that these lecture notes will not attempt to answer (but you are very welcome to engage with in the 2nd year course in topics in economic theory).

Instead, we will focus on a simpler question: how do we model *more informative than*? This is a fundamental question, as it will allow us to compare information sources, and to understand information choice, a preliminary step to addressing the topics above.

2. Modelling Information

We consider an unknown state of the world $\theta \in \Theta$ and a prior probability distribution $\mu \in \Delta(\Theta)$.

Information is then something that changes our beliefs about the unknown state of the world θ. We model this as a **statistical experiment or information structure**, a function $\pi : \Theta \to \Delta(X)$, where X is a set of signals. Without loss of generality (wlog), we assume that $\pi(x) := \sum_\theta \pi(x|\theta)\mu(\theta) > 0$ for all x, where $\pi(x)$ denotes the ex-ante probability of signal x; for ease of notation, we'll write $x \sim \pi$ to denote this. Unless stated otherwise, we assume Θ and X are finite, where label elements as $\Theta = \{\theta_1, \ldots, \theta_{|\Theta|}\}$ and $X = \{x_1, \ldots, x_{|X|}\}$.

We can then represent our information structure as a $|X| \times |\Theta|$ matrix, $\pi := (\pi_{ij})_{i=1,\ldots,|X|, j=1,\ldots,|\Theta|}$, where $\pi(x|\theta)$ denotes the probability of observing signal x given state θ. The matrix π is column-stochastic, with $\pi_{ij} \in [0,1]$ and $\sum_{i=1,\ldots,|X|} \pi_{ij} = 1$. The signal labels have no content.

The matrix π can be written as:

$$\pi = \begin{pmatrix} \pi(x_1 \mid \theta_1) & \cdots & \pi(x_1 \mid \theta_{|\Theta|}) \\ \vdots & \pi(x \mid \theta) & \vdots \\ \pi(x_{|X|} \mid \theta_1) & \cdots & \pi(x_{|X|} \mid \theta_{|\Theta|}) \end{pmatrix}$$

An information structure is **fully informative** if $\forall x \in X$ such that $\pi(x|\theta) > 0$ for some $\theta \in \Theta$, then $\pi(x|\theta') = 0$ for any $\theta' \neq \theta$. In other words, if we observe signal x, we know that the state is

θ with probability one, as in no other state of the world would we ever get the chance to observe signal x. When $|X| = |\Theta|$, this amounts to π being the identity matrix (up to permutation of rows/columns). In contrast, π is **fully uninformative** if $\pi(x|\theta) = \pi(x|\theta')$ for any $\theta, \theta' \in \Theta$; i.e. each row of $[\pi(x|\theta)]_{\theta,x}$ is a constant vector. This means that the signal x does not provide any information about the state θ.

Example 1. Suppose that $\theta \in \{G, N\}$ denotes the fact that an equipment is working properly (G) or not (N). There is a quality control in place that checks the equipment and produces a signal $x \in \{g, n\}$. This quality control procedure is good at detecting malfunctions, i.e. it will only report there is a malfunction if indeed there is one $\pi(n|G) = 0$, but might not run all the tests necessary to detect a malfunction in that it might report that the equipment is working properly even if it is not $\pi(n|N) \in [0,1]$. π is then an information structure. If $\pi(n|N) = 1$, then π is fully informative (and equivalent to the identity matrix). If $\pi(n|N) \in (0,1)$, then π is not fully informative; in particular, if $\pi(n|N) = \pi(n|G)$, then π is fully uninformative.

3. Blackwell Order

In the context of information, we say that "more information is better." But what does it mean for something to be more informative?

3.1. Value of Information

Consider a **decision problem**, summarised by utility function $u : A \times \Theta \to \mathbb{R}$, where A is a finite set of actions and Θ a finite set of states. We write $u_a := (u(a,\theta), \theta \in \Theta) \in \mathbb{R}^{|\Theta|}$ for the vector of state-contingent payoffs associated with action $a \in A$. For simplicity, we abuse notation and identify an action $a \in A$ with a vector of state-contingent payoffs; a decision problem A is then a set of state-contingent payoff vectors.

For a belief $\mu \in \Delta(\Theta)$, we write $\mu \cdot u_a \equiv \mathbb{E}_{\theta\sim\mu}[u(a,\theta)] = \sum_\theta \mu(\theta)u(a,\theta)$ for the expected payoff of action a. We define $U^A : \Delta(\Theta) \to \mathbb{R}$ as the **maximised expected utility in problem** A, where $U^A(\mu) := \max_{a\in A} \mathbb{E}_{\theta\sim\mu}[u(a,\theta)]$.

Given information structure π and prior μ, the posterior probability of θ given x, denoted as $(\mu|x)$, is defined as $(\mu|x)(\theta) = \mu(\theta|x) = \frac{\mu(\theta)\pi(x|\theta)}{\pi(x)}$. Then, given a realised signal x, the expected payoff of action a is $\mathbb{E}_{\theta\sim\mu|x}[u(a,\theta)] = \sum_\theta (\mu|x)(\theta)u(a,\theta)$ and the maximised expected utility is $U^A(\mu|x)$.

Since an information structure π (together with a prior μ) entails an ex-ante probability distribution over signals, we can define the **value of an experiment** π in a decision problem A with prior μ as the expectation over the maximised expected utility given each possible signal, which we denote as

$$V^A(\pi,\mu) := \mathbb{E}_{x\sim\pi}\left[U^A(\mu|x)\right] = \mathbb{E}_{x\sim\pi}\left[\max_{a\in A}\mathbb{E}_{\theta\sim\mu|x}[u(a,\theta)]\right] = \sum_x \pi(x)\left[\max_{a\in A}\sum_\theta(\mu|x)(\theta)u(a,\theta)\right]$$

Given a prior μ and a decision problem A, we say that experiment π **is more valuable than** π' if $V^A(\pi,\mu) \geq V^A(\pi',\mu)$.

Definition 1 (Value and Informativeness)**.** We say that information structure π **is more informative than** π' — denoted as $\pi \geq_B \pi'$ — if π is more valuable than π' for any decision problem A and any prior μ.

The order $\geq_B$ is called the **Blackwell order** on information structures.

We want to take a moment to make a parallel between how the Blackwell order is defined and the definitions of stochastic orders, such as the first-order stochastic dominance order (FOSD). Recall that the latter is defined as follows: for any real-valued random variables, X and Y, X FOSD Y if $\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all nondecreasing $u : \mathbb{R} \to \mathbb{R}$. In other words, X delivers a higher expected utility than Y for all expected-utility maximisers satisfying a weak monotonicity condition. Similarly, one can think of π being more informative than π' as π delivering a higher expected utility than π' for all subjective expected-utility maximisers (all u-functions and beliefs μ) in all finite problems (all action sets A).

3.2. Statistical Sufficiency

The above is an instrumental-value-driven definition of informativeness. We need to think of a(ll) Bayesian decision-maker(s), with subjective expected utility, and consider all problems they might encounter to rule whether an information structure is more or less informative than another. This is not very tractable and it is a bit distant from a statistically-oriented notion of being more or less informative.

Let's take a step back and think about what an experiment is. We can think of an experiment as adding noise to the direct observation of the state θ. Instead of observing θ directly, we obtain a signal x according to the distribution $\pi(x|\theta)$.

We can also consider adding noise to π. In this case, we have a new matrix π' that produces a signal $y \in Y$, where $\pi' : \Theta \to \Delta(Y)$ and "$y = x+$ noise". We'll consider any such experiment π as statistically sufficient for π since $\mathbb{E}_\mu[\theta|x,y] = \mathbb{E}_\mu[\theta|x]$. Let's define properly the notion of statistical sufficiency. Denote the set of all $|Y|\times|X|$ column-stochastic matrices as $\mathscr{B}(|Y|,|X|)$. These are also called Markov matrices.

Definition 2 (Sufficiency)**.** An experiment $\pi : \Theta \to \Delta(X)$ is **sufficient** for $\pi' : \Theta \to \Delta(Y)$ if there exists $b \in \mathscr{B}(|Y|,|X|)$ such that $b\pi = \pi'$.

When $\pi' = b\pi$, we say that π' is a **garbling** of π. Note that $\pi'(y|\theta) = \sum_x b(y|x)\pi(x|\theta)$. π' is a garbling of π in the sense that following the realisation of signal x drawn from $\pi(x|\theta)$, b adds noise to this signal in order to produce signal y.

Continuing with our running example,

Example 1 (continued)**.** Note that π is always a garbling of the identity matrix, I, where $I(x|\theta) = 1_{x=\theta}$, since $\pi = \pi I$. Now suppose that an employee need to report whether the equipment is working properly or not, and that they have access to π. Then, they observe signal $x \in \{g, n\}$ and issue a report $y \in \{\tilde{g}, \tilde{n}\}$. However, they make a mistake with some probability $\epsilon \in (0,1)$, in that they report $y = x$ with probability $1-\epsilon$ and $y \neq x$ with probability ϵ. This is a garbling of π, where b is given by $b(\tilde{g}|g) = b(\tilde{n}|n) = 1-\epsilon$ and $b(\tilde{n}|g) = b(\tilde{g}|n) = \epsilon$. Then, the employee's report is also an information structure about the state, $\pi'(y|\theta) = \sum_x b(y|x)\pi(x|\theta)$. Moreover, π' is a garbling of π, in that it is adding noise to the outcome of the quality control procedure.

3.3. Equivalence Result

We can now state the main result of these lecture notes, due to Blackwell (1951); Blackwell (1953).

Theorem 1 (Blackwell (1951, 1953))**.** *(1) If experiment π is sufficient for experiment π', then π is more informative than π'. (2) If π is more informative than π' for some $\mu \in int(\Delta(\Theta))$, then π is sufficient for π'.*

The fact that sufficiency implies informativeness is unsurprising. The reason is that if the decision-maker has access to π, they can always mimic π' by randomising their actions in a way that adds noise, analogously to that of the stochastic matrix b.

The fact that sufficiency is implied by informativeness *is* surprising. This result does not depend on the prior distribution. The proof of the fact that below follows that by Crémer (1982).

Proof. The proof is by contrapositive. That is, we prove that π is not sufficient for π' if and only if we can find a prior and a decision problem such that $V^A(\pi', \mu) > V^A(\pi, \mu)$, thus concluding that π is not more informative than π'. We take the arbitrary signal spaces $\pi : \Theta \to \Delta(X)$ and $\pi' : \Theta \to \Delta(Y)$, where $|X|, |Y| < \infty$.

(i) π is not sufficient for $\pi' \iff \pi' \notin \mathscr{B}\pi := \{b\pi \mid b \in \mathscr{B}(|Y|, |X|)\}$

$\iff$ (ii) There exists $q = [q_{y,\theta}]_{\theta\in\Theta, y\in Y}$ such that for all $b \in \mathscr{B}(|Y|, |X|)$,

$$\sum_{y,\theta} \pi'(y|\theta) q_{y,\theta} > \sum_{y,\theta} (b\pi)(y|x) q_{y,\theta} = \sum_{y,\theta}\sum_x b(y|x)\pi(x|\theta) q_{y,\theta} = \sum_x \sum_y b(y|x) \sum_\theta \pi(x|\theta) q_{y,\theta}.$$

This is because $\mathscr{B}\pi$ is a closed and convex set. By the separating hyperplane theorem, $\pi' \notin \mathscr{B}\pi$

if and only if there exists such q. (Note: since everything is finite dimensional, we can think of π' and q as $|Y|\cdot|\Theta|$ real-valued vectors and $\mathscr{B}\pi$ as a closed convex set of vectors.)

$\iff$ (iii) There exists $q = [q_{y,\theta}]_{\theta\in\Theta, y\in Y}$ such that for all $b \in \mathscr{B}(|Y|,|X|)$,

$$\sum_{y,\theta} \pi'(y|\theta) q_{y,\theta} > \sum_x \max_y \left[\sum_\theta \pi(x|\theta) q_{y,\theta} \right].$$

The reason for this is that $\sum_x \max_y \sum_\theta \pi(x|\theta) q_{y,\theta} \geq \sum_x \sum_y b(y|x) \sum_\theta \pi(x|\theta) q_{y,\theta}$ for any $b \in \mathscr{B}(|Y|,|X|)$ and we can choose b such that $b(y|x) = 1_{y=y^*(x)}$, where $y^*(x)$ is the selection of $\arg\max_y \sum_\theta \pi(x|\theta) q_{y,\theta}$.

$\iff$ (iv) If $\mu : \mu(\theta) > 0$, there exists $u : Y \times \Theta \to \mathbb{R}$, i.e. $u = [u_{y,\theta}]_{\theta\in\Theta, y\in Y}$ such that

$$\sum_y \pi'(y) \sum_\theta \mu(\theta|y) u(y,\theta) > \sum_x \pi(x) \max_y \left[\sum_\theta \mu(\theta|x) u(y,\theta) \right],$$

where $\mu(\theta|y) = \frac{\mu(\theta)\pi'(y|\theta)}{\pi'(y)}$, $\mu(\theta|x) = \frac{\mu(\theta)\pi(x|\theta)}{\pi(x)}$, $\pi'(y) = \sum_\theta \pi'(y|\theta)\mu(\theta)$, and $\pi(x) = \sum_\theta \pi(x|\theta)\mu(\theta)$. The proof for this statement is by construction of the said utility function, where $u(y,\theta) := q_{y,\theta}/\mu(\theta)$. Then, $\pi'(y|\theta) q_{y,\theta} = \frac{\pi'(y)\mu(\theta|y)}{\mu(\theta)} q_{y,\theta} = \pi'(y)\mu(\theta|y)u(y,\theta)$, leading to

$$\sum_{y,\theta} \pi'(y|\theta) q_{y,\theta} = \sum_{y,\theta} \pi'(y)\mu(\theta|y)u(y,\theta) = \sum_y \pi'(y) \sum_\theta \mu(\theta|y)u(y,\theta).$$

Moreover, $\pi(x|\theta) q_{y,\theta} = \frac{\mu(\theta|x)\pi(x)}{\mu(\theta)} q_{y,\theta} = \pi(x)\mu(\theta|x)u(y,\theta)$, and

$$\sum_x \max_y \left[\sum_\theta \pi(x|\theta) q_{y,\theta} \right] = \sum_x \max_y \left[\sum_\theta \pi(x)\mu(\theta|x)u(y,\theta) \right] = \sum_x \pi(x) \max_y \left[\sum_\theta \mu(\theta|x)u(y,\theta) \right].$$

The equivalence between (iii) and (iv) then follows.

$\iff$ (v) There exists a decision problem A and a prior μ such that $V^A(\pi',\mu) > V^A(\pi,\mu)$ (i.e. π is not more informative than π').

That (iv) implies (v) follows immediately from setting $A = Y$ and using the utility function defined in (iv).

To see that (v) implies (iv), suppose that $V^A(\pi',\mu) > V^A(\pi,\mu)$ for some decision problem A with associated utility function $\tilde{u}$ and prior μ. For each $y \in Y$, define $a^*(y)$ as a selection of $\arg\max_{a\in A} \mathbb{E}_{\theta\sim\mu|y}[\tilde{u}(a,\theta)]$, collect all such actions in $A^* := \{a^*(y) | y \in Y\}$, and let $u : Y \times \Theta \to \mathbb{R}$ be such that $u(y,\theta) = \tilde{u}(a^*(y),\theta)$.

Note that

$$\begin{aligned} \max_{a\in A} \mathbb{E}_{\theta\sim\mu|y}[\tilde{u}(a,\theta)] = \mathbb{E}_{\theta\sim\mu|y}[\tilde{u}(a^*(y),\theta)] &= \sum_\theta \mu(\theta|y)u(y,\theta) \\ &= \max_{y'\in Y} \mathbb{E}_{\theta\sim\mu|y}[u(y',\theta)] = U^Y(\mu|y). \end{aligned}$$

Moreover,

$$\begin{aligned}U^A(\mu|x) = \max_{a\in A}\mathbb{E}_{\theta\sim\mu|x}[\tilde{u}(a,\theta)] &\geq \max_{a\in A^*}\mathbb{E}_{\theta\sim\mu|x}[\tilde{u}(a,\theta)]\\ &= \max_{y\in Y}\mathbb{E}_{\theta\sim\mu|x}[\tilde{u}(a^*(y),\theta)]\\ &= \max_{y\in Y}\mathbb{E}_{\theta\sim\mu|x}[u(y,\theta)] = U^Y(\mu|x)\end{aligned}$$

$$\implies \quad V^A(\pi,\mu) = \mathbb{E}_{x\sim\pi(x)}[U^A(\mu|x)] \geq \mathbb{E}_{x\sim\pi(x)}[U^Y(\mu|x)] = V^Y(\pi,\mu) = \sum_x \pi(x)\max_y\left[\sum_\theta \mu(\theta|x)u(y,\theta)\right].$$

Hence,

$$\begin{aligned}\sum_y \pi'(y)\sum_\theta \mu(\theta|y)u(y,\theta) = V^Y(\pi',\mu) &= V^A(\pi',\mu)\\ &> V^A(\pi,\mu) \geq V^Y(\pi,\mu)\\ &= \sum_x \pi(x)\max_y\left[\sum_\theta \mu(\theta|x)u(y,\theta)\right].\end{aligned}$$

Since we have a strict gap, this would still hold for a full-support μ' that is close enough to μ. This concludes the proof.

□

3.4. Payoff Richness

The concept of payoff richness compares two experiments, π and π', based on the set of state-contingent feasible payoffs one could achieve given the information each provides.

Let's make this formal. Given a decision problem A and an experiment $\pi : \Theta \to \Delta(X)$, let a decision rule d be a function mapping from signals X to distributions over actions A. Previously, we considered one particular (class of) decision rule(s), that which maximises expected payoffs according to the posterior belief about θ given each signal x. Since each action pins-down a state-contingent payoff vector $u_a = (u(a,\theta), \theta \in \Theta)$, we can then represent a decision rule d also as a $|\Theta|$ by $|X|$ matrix, where the element $d_{\theta,x}$ corresponds to the expected payoff obtained when taking action $d(x)$ in state θ, i.e. $u(d(x),\theta) = \mathbb{E}_{a\sim d(x)}[u(a,\theta)]$.

Let $v^A_\theta(d,\pi) := \sum_x \pi(x|\theta)u(d(x),\theta)$ denote the expected payoff associated with using decision rule d in problem A given information structure π, and define $v^A(d,\pi) := (v^A_\theta(d,\pi), \theta \in \Theta)$ as the vector of all such payoffs. Note that $v^A_\theta(d,\pi)$ corresponds to the θ-th diagonal element of the $|\Theta| \times |\Theta|$ matrix $d\pi$, which has generic element $(d\pi)_{\theta,\theta'} = \sum_x \pi(x|\theta')u(d(x),\theta)$. Note that, letting μ denote the uniform prior over Θ,

$$\begin{aligned}\sum_\theta \mu(\theta)v^A_\theta(d,\pi) = \sum_\theta v^A_\theta(d,\pi)/|\Theta| = \mathrm{tr}(d\pi)/|\Theta| &= \sum_\theta \mu(\theta)\sum_x \pi(x|\theta)u(d(x),\theta)\\ &= \sum_x \pi(x)\sum_\theta \mu(\theta|x)u(d(x),\theta) = \mathbb{E}_{x\sim\pi}\left[\mathbb{E}_{\theta\sim\mu|x}[u(d(x),\theta)]\right]\end{aligned}$$

That is, the sum of the elements $v^A(d,\pi)$ equals the trace of $d\pi$ which in turn equals $|\Theta|$ times the expected payoff associated with using decision rule d with experiment π when the prior is uniform.

Letting $D^{A,\pi} := \Delta(A)^X$ denote the set of all decision rules in problem A given π, define $P^A(\pi) := \cup_{d \in D^{A,\pi}} v^A(d,\pi)$, corresponding to the range of all state-contingent expected payoffs one could achieve with some decision rule. Note that $P^A(\pi)$ is convex and compact.

We say that π is **payoff richer** than π' if $P^A(\pi) \supseteq P^A(\pi')$ for all decision problems A. Note that payoffs are state-contingent — there is no reference to a prior.

We then have another equivalence result:

Theorem 2 (Blackwell (1953))**.** *Let $\pi \in \Delta(X)^\Theta$ and $\pi' \in \Delta(Y)^\Theta$. π is payoff richer than π' if and only if π is more informative than π'.*

This means than if π is more informative than π' then a Bayesian expected utility maximiser attains higher payoffs in expectation, but if they deviate from maximising posterior expected payoffs, then they may actually do worse. This is because if $P^A(\pi) \supseteq P^A(\pi')$, then the can do at least as well under π as under π', but the worst we can do under π is also at least as bad as how we do under π'!

Before we go into the proof, we'll prove a lemma first.

Lemma 1 (Blackwell (1953))**.** *Let $\pi \in \Delta(X)^\Theta$ and $\pi' \in \Delta(Y)^\Theta$. If π is payoff richer than π', then for any real-valued $|\Theta| \times |Y|$ matrix d', there is $b \in \mathscr{B}(|Y|,|X|)$ such that $(d'b\pi)_{\theta,\theta} = (d'\pi')_{\theta,\theta}$ for all $\theta \in \Theta$.*

Proof. Let decision problem Y be given by payoff function $u(y,\theta) := d'_{\theta,y}$.
As $P^Y(\pi) = \text{co}\left(\{v^Y(d,\pi) | d \in Y^X\}\right) \supseteq P^Y(\pi') \ni v^Y(d',\pi')$, and as $D^{Y,\pi} = \text{co}\left(\{d \in Y^X\}\right)$, there is $d \in D^{Y,\pi}$ such that $v^Y(d,\pi) = v^Y(d',\pi')$, i.e. $\text{diag}(d\pi) = \text{diag}(d'\pi')$. As such a decision rule d can be seen as a $|\Theta|$ by $|X|$ matrix, where the element $d_{\theta,x}$ corresponds to the expected payoff obtained when taking action $d(x)$ in state θ, i.e. $u(d(x),\theta) = \mathbb{E}_{y \sim d(x)}[u(y,\theta)] = \mathbb{E}_{y \sim d(x)}[d'_{\theta,y}]$, there is a $b \in \mathscr{B}(|Y|,|X|)$ such that $d'b = d$, hence $(d'b\pi)_{\theta,\theta} = (d'\pi')_{\theta,\theta}$ for all $\theta \in \Theta$. □

We're now ready to prove the main result of this section.

Proof of Theorem 2. (i) π is payoff richer than $\pi' \Longleftarrow \pi$ is more informative than π':

As π is more informative than π', then π is sufficient for π', which means there is $b \in \mathscr{B}(|Y|,|X|)$ such that $b\pi = \pi'$.

Fix any decision problem A and take $v^A(d,\pi') \in P^A(\pi')$. Given decision rule $d \in D^{A,\pi'}$, note that (db) is a valid decision rule for problem A given information structure π, i.e. $(db) \in D^{A,\pi}$, and so $v^A(db,\pi) \in P^A(\pi)$. We then have that $d\pi' = db\pi = (db)\pi$, and so $v^A(db,\pi) = v^A(d,\pi') \in P^A(\pi)$.

This implies that $P^A(\pi') \subseteq P^A(\pi)$ and concludes the proof of (i).

(ii) π is payoff richer than $\pi' \implies \pi$ is more informative than π':

Let D be the set of $|\Theta| \times |X|$ matrices d such $d_{ij} \in [0,1]$. Consider an auxillary two-player zero-sum game, where player 1 chooses $d \in D$ and player 2 chooses $b \in \mathscr{B}(|Y|,|X|)$. Player 1's utility is given by $\text{tr}(d(b\pi - \pi'))/|\Theta|$ and player 2's is negative of that. This can be interpreted as saying that player 1's utility is the ex-ante expected payoff different between using their decision rule on the garbling of π that player 2 chose relative to using it on π', taking a uniform prior over the states.

The zero-sum game satisfies the usual conditions for existence of a pure-strategy Nash equilibrium (compact, convex strategy set, linear payoffs). Denote a PSNE as $(\hat{d},\hat{b})$.

From Lemma 1 implies that against any pure strategy of player 1, player 2 has a garbling that yields zero utility in this game. This implies that player 1's equilibrium payoff must be no higher than zero under any equilibrium $(\hat{d},\hat{b})$. Hence, for all $d \in D$, $\text{tr}(d(\hat{b}\pi - \pi')) \leq \text{tr}(\hat{d}(\hat{b}\pi - \pi')) \leq 0$.
This also implies that $\hat{b}\pi - \pi' \leq 0$, since if any entry (θ^*, y^*) in $\hat{b}\pi - \pi'$ were strictly positive, player 1 would have a strictly profitable deviation by choosing d such that $d_{\theta,y} = 1$ for $(\theta, y) = (\theta^*, y^*)$ and zero elsewhere, achieving a strictly positive payoff.
As $\hat{b}\pi$ and π' are stochastic matrices, then $\hat{b}\pi, \pi' \geq 0$ and the sum of each column of $\hat{b}\pi - \pi'$ equals zero, as $\sum_y(\sum_x \hat{b}(y|x)\pi(x|\theta) - \pi'(y|\theta)) = 0$. But if elements are non-negative and their sum is zero, then $\hat{b}\pi = \pi'$.

Therefore, we found a stochastic matrix that shows that π is sufficient for π' and hence it is more informative than π'.

□

Before moving on, let's just pause for a second and see how to prove a completely unrelated result, Blackwell's insight was to find an auxillary zero-sum game, which let us appeal to known existence results and exploit the particular details of the game to obtain the result. This is idea of solving problems by designing a non-cooperative game is something that is pervasive (and elegant). Here's an example hitting the buzz of the day: a fundamental approach underlying generative AI, generative adversarial networks (GANs), corresponds to a machine learning framework in which two neural networks compete in a zero-sum game.

3.5. Posterior Beliefs

We will go now to one last equivalence result that is arguably the most used in information economics. It states that an information structure is more informative than another if and only if the posterior beliefs under the former are a mean-preserving spread than the posterior beliefs under the latter — a result first discovered for the case where $X \subset \mathbb{R}$ by Rothschild and Stiglitz (1970).

Note that given a prior μ, since $x \sim \pi(x)$ is a random variable, then the posterior belief $\mu|x \in \Delta(\Theta)$ is also a random variable (from an ex-ante perspective). Since for any decision problem A and belief μ, $U^A(\mu) := \max_{a \in A} \mathbb{E}_{\theta\sim\mu}[u(a,\theta)]$, then $U^A : \Delta(\Theta) \to \mathbb{R}$ is a convex function.

Recall the following:

Theorem 3. *Let X and Y be $\mathbb{R}^n$-valued random varibles, where $X \sim P_X$ and $Y \sim P_Y$. The following are equivalent:*

1. *X is a **mean-preserving spread** of Y (denoted $X \geq_{MPS} Y$ or $P_X \geq_{MPS} P_Y$).*
2. *There is a $\mathbb{R}^n$-valued random varible ϵ such that $\mathbb{E}[\epsilon|Y] = 0$ and $X \overset{d}{=} Y + \epsilon$.*
3. *$\mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)]$ for all convex functions $u : \mathbb{R}^n \to \mathbb{R}$ whenever both expectations exist.*
4. *$\mathbb{E}[u(X)] \leq \mathbb{E}[u(Y)]$ for all concave functions $u : \mathbb{R}^n \to \mathbb{R}$ whenever both expectations exist.*

This implies that if $\mu|x$ is a mean-preserving spread of $\mu|y$, then $\mathbb{E}[U^A(\mu|x)] \geq \mathbb{E}[U^A(\mu|y)]$, suggesting an intrinsic connection between mean-preserving spreads of posterior beliefs and informativeness of information structures. The following result makes this connection formal:

Theorem 4. *Let $\pi \in \Delta(X)^\Theta$ and $\pi' \in \Delta(Y)^\Theta$ be two information structures. π is more informative than π' if and only if $\mu|x$ is a mean-preserving spread (MPS) of $\mu|y$.*

Proof. Let μ have full support (this is without loss of generality; otherwise just redefine $\Theta = \text{supp}(\mu)$).

Observe that since beliefs live in the simplex, $\mu|x$ is a mean-preserving spread of $\mu|y$ if and only if there is a $|Y| \times |X|$ row-stochastic matrix $\beta \in \mathscr{B}(|Y|, |X|)$ such that $\sum_x \beta(x|y)\mu|x = \mu|y$, for each signal realisation y.

The "if" part follows from the arguments we presented above: if $\mu|x \sim \pi$ is a mean-preserving spread of $\mu|y \sim \pi'$, then $V^A(\pi) = \mathbb{E}_{x\sim\pi}[U^A(\mu|x)] \geq \mathbb{E}_{y\sim\pi'}[U^A(\mu|y)]$.

We focus on the "only if" part. Let $\mathrm{D}(v)$ represent a diagonal matrix where its diagonal corresponds to vector v. Notice that $(\pi D(\mu))$ is a $|X| \times |\Theta|$ matrix with element $(\pi D(\mu))_{x,\theta} = \mu(\theta)\pi(x|\theta)$, while $\pi\mu$ is a $|X|$-dimensional vector, where $(\pi\mu)_x = \sum_\theta \mu(\theta)\pi(x|\theta)$. Then, posteriors are given by the rows of the $|X| \times |\Theta|$ matrix $\mu|X := D(\pi\mu)^{-1}\pi D(\mu)$.

We have that $\exists b \in \mathscr{B}(|Y|,|X|)$ such that $\pi' = b\pi$. Hence, we now need but to construct $\beta \in \mathscr{B}(|X|,|Y|)$ such that

$$\begin{aligned}
& \beta\mu|X = \mu|Y \\
& \beta(D(\pi\mu))^{-1}\pi D(\mu) = (D(b\pi\mu))^{-1}b\pi D(\mu) \\
\iff \quad & \beta(D(\pi\mu))^{-1}\pi = (D(b\pi\mu))^{-1}b\pi \\
\iff \quad & 0 = \left[\beta(D(\pi\mu))^{-1} - (D(b\pi\mu))^{-1}b\right]\pi \\
\iff \quad & 0 = \left[\beta - (D(b\pi\mu))^{-1}bD(\pi\mu)\right](D(\pi\mu))^{-1}\pi
\end{aligned}$$

and then, $\beta = (D(b\pi\mu))^{-1}bD(\pi\mu)$. We conclude by verifying β is indeed a $|Y| \times |X|$ row-stochastic matrix:

$$\beta(x|y) = \frac{b(y|x)\sum_\theta \pi(x|\theta)\mu(\theta)}{\sum_{x'} b(y|x')\sum_\theta \pi(x'|\theta)\mu(\theta)} \in [0,1]$$
$$\sum_x \beta(x|y) = 1.$$

□

Why is this such an important tool? Because any information structure induces a distribution over posterior beliefs. Hence, we can for all purposes associated information structures with distributions over posteriors, i.e. $\pi \in \Pi := \{\pi \in \Delta(\Delta(\Theta)) \mid \mathbb{E}_{\mu|x\sim\pi}[\mu|x] = \mu\}$, which starts making our life easier, as we don't need to keep talking about arbitrary signal spaces. Then, we have that $\pi \geq_B \pi' \iff \pi \geq_{MPS} \pi'$, and since $(\Pi, \geq_{MPS})$ is a lattice, then we get a lot of structure on the space of information structures that we can exploit to obtain monotone comparative statics. A good tool for these distributional comparative statics is Jensen (2018) (with macro-based examples).

3.6. Beyond Finiteness

Infinite Signal Space. The results above do generalise to infinite signal spaces. For simplicity, let's keep $|\Theta| < \infty$.

A statistical experiment can be taken to be a tuple $(X, \mathbb{X}, \pi)$, where X is a subset of an Euclidean space, $\mathbb{X}$ is the Borel σ-algebra, and $\pi : \Theta \to \Delta(X)$ is such that $\pi(\cdot|\theta)$ is a Borel probability measure on X. Then we can talk of **Markov kernel** b from $(X, \mathbb{X})$ to $(Y, \mathbb{Y})$, which is a map $b : X \times \mathbb{Y} \to [0,1]$ such that (i) for every $x \in X$, $b(\cdot|x)$ is a probability measure on $(Y, \mathbb{Y})$, and (ii) for every $S \in \mathbb{Y}$, the map $x \mapsto b(S|x)$ is $\mathbb{X}$-measurable.

Definition 3. Fix Θ and let $(X, \mathbb{X}, \pi)$ and $(Y, \mathbb{Y}, \pi')$ be two statistical experiments. The former is sufficient for the latter if and only if there is a Markov kernel b from $(X, \mathbb{X})$ to $(Y, \mathbb{Y})$ such that, for all $\theta \in \Theta$ and $S \in \mathbb{Y}$,

$$\pi'(S|\theta) = \int_{x\in X} b(S|x)\pi(\mathrm{d}x|\theta).$$

If X and Y are finite, then we'd be back to our previous definition and $b \in \mathscr{B}(|Y|,|X|)$. Moreover, all results extend in the natural way using this definition of sufficiency and straightforward generalisations of the other relevant notions. See Le Cam (1964) for details.

Infinite State Space. With an infinite state space Θ, a definition of statistical sufficiency is as follows:

Definition 4. Fix Θ and let $\pi \in \Delta(X)^{\Theta}, \pi' \in \Delta(Y)^{\Theta}$ be two experiments, with $x|\theta \sim \pi(\cdot|\theta)$ and $y|\theta \sim \pi'(\cdot|\theta)$. π is said to be **sufficient** for π' if there there is a random variable z independent of θ and a function $h(y,z)$ such that, for all $\theta \in \Theta$, $h(y|\theta,z) \overset{d}{=} x|\theta$.

The leading example of the above is that of Gaussian variables, where $x \sim \mathcal{N}(\theta,\sigma^2)$ and $y \sim \mathcal{N}(\theta,\lambda\sigma^2)$ (and λ,σ known). Then $x \geq_B y$ if and only if $\lambda \geq 1$. Note that $y \overset{d}{=} x + \sqrt{\lambda - 1}\epsilon$, where $\epsilon \sim \mathcal{N}(0,\sigma^2)$ and ϵ is independent from x.

3.7. Comments

Note that the Blackwell order is very incomplete: it is *demanding* that an information be better than another in *any* decision problem!

For instance, let $x \sim \text{Unif}(\theta-1,\theta+1)$ and $y \sim \text{Unif}(\theta-k,\theta+k)$, $\Theta = [0,1]$. Then x is more informative than y for any $k \in \mathbb{N}$, but they are not ranked if $k \notin \mathbb{N}$.
Another example: $x = \theta + \epsilon$ and $y = \theta + \delta$. If ϵ,δ are both zero-mean Gaussian random variables and independent of x and y, respectively, then x and y are Blackwell-ranked. If ϵ is zero-mean Gaussian random variable and independent of x, but δ is not normally distributed, then they are not Blackwell-ranked.

The coarseness of this ordering[1] has led people to consider some refinements of Blackwell (similar to what happens when going from FOSD to SOSD), but restricting the class of decision problems one allows for. One of which is the Lehmann (1988) order, which has been exploited by Persico (2000) to study information acquisition in auction settings.

Another way to complete the Blackwell order, is to consider instead of the value of information, a notion of its cost.

Despite these rather negative perspectives, there Blackwell is the gold standard of ranking information in decision problems. More: it turns out that an information structure makes a player better off across all in two-player zero-sum games if and only if it improves that player's information or if it worsens the opponent's, in the Blackwell/sufficiency sense (Peşki, 2008).

[1]It's only a proper partial order if we consider experiments as distributions over posterior beliefs.

References

Blackwell, David. 1951. "Comparison of Experiments." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1. 93–102.

Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *Annals of Mathematical Statistics* 24 (2): 265–272. 10.1214/aoms/1177729032.

Crémer, Jacques. 1982. "A simple proof of Blackwell's "comparison of experiments" theorem." *Journal of Economic Theory* 27 (2): 439–443. 10.1016/0022-0531(82)90040-0.

Jensen, Martin Kaae. 2018. "Distributional Comparative Statics." *The Review of Economic Studies* 85 (1): 581–610. 10.1093/restud/rdx021.

Le Cam, Lucien. 1964. "Sufficiency and Approximate Sufficiency." *Annals of Mathematical Statistics* 35 (5): 1419–1455. 10.1214/aoms/1177700372.

Lehmann, Erich. 1988. "Comparing Location Experiments." *The Annals of Statistics* 16 (2): 521–533. 10.1214/aos/1176350818.

Peşki, Marcin. 2008. "Comparison of information structures in zero-sum games." *Games and Economic Behavior* 62 (3): 732–735. 10.1016/j.geb.2007.06.004.

Persico, Nicola. 2000. "Information Acquisition in Auctions." *Econometrica* 68 (1): 135–148. 10.1111/1468-0262.00096.

Rothschild, Michael, and Joseph E. Stiglitz. 1970. "Increasing risk: I. A definition." *Journal of Economic Theory* 2 (3): 225–243. 10.1016/0022-0531(70)90038-4.