Chapter 4

Comparing Information I: The Blackwell Order

When an agent has access to a choice between multiple signals, we may desire to order these signals based on how informative they are. Intuition can guide us on how to define such an ordering in specific cases, for example:

- Adding noise to a signal decreases its informativeness.

- Observing the realization of (X, Y) is more informative than observing the realization of X alone.

Any informativeness ordering should satisfy these properties, but there are different ways to generalize from here. One approach is to fix a decision problem and characterize the instrumental value of the signal for that decision problem. Alternatively, we could look for a universal informativeness ordering over signals that holds for all decision problems (as will be the focus of this chapter). Yet another approach is to quantify the "signal content" contained within the signal based on the physical difficulty of producing or processing that information (see Chapter 5).

In this chapter we introduce the Blackwell partial order on signals, which considers one signal to be more informative than another if it is more useful for all decision problems. If σ dominates σ' in this Blackwell order, we will say that σ *is more informative than* σ' or that σ *Blackwell-dominates* σ'. The following sections demonstrate five perspectives on this order, culminating in Blackwell's theorem (establishing their equivalence) and the proof of this theorem.

4.1   Garblings

We may consider a signal to be more informative than another if the latter is a noised-up version of the former.

DEFINITION 4.1 (Markov matrix). *A matrix M is a* Markov matrix *if its entries are nonnegative and its rows sum to 1.*

Recall that when the set of states and the set of signal realizations are finite, we can represent any signal as a Markov matrix.

DEFINITION 4.2 (Garblings, Finite Version). *Markov matrix P is a* garbling *of Markov matrix Q if there exists a Markov matrix M s.t. $QM = P$*

EXAMPLE 4.1. Let $\Theta = \{\theta_1, \theta_2\}$ and consider the signals

$$P = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix} \qquad Q = \begin{pmatrix} 9/16 & 3/16 & 3/16 & 1/16 \\ 1/16 & 3/16 & 3/16 & 9/16 \end{pmatrix}$$

where as usual the rows are indexed to states and the columns are indexed to signal realizations. Then since

$$\underbrace{\begin{pmatrix} 9/16 & 3/16 & 3/16 & 1/16 \\ 1/16 & 3/16 & 3/16 & 9/16 \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}}_{M} = \underbrace{\begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}}_{P}$$

where M is a Markov matrix, we can conclude that P is a garbling of Q.

This example has a particularly nice intuition. Label the possible realizations of the first information structure P as s_1 and s_2, and consider the signal which is two independent realizations of P. The set of possible realizations of this new signal is then $\{s_1 s_1, s_1 s_2, s_2 s_1, s_2 s_2\}$ with the conditional distributions over these realizations given precisely by Q. So observing P is statistically equivalent to observing Q and forgetting the second realization. Clearly then P is less informative than Q.

EXERCISE 4.1 (U). *The state space is $\Theta = \{\theta_1, \theta_2\}$ and two signals are described by the following signal structures*

$$P = \begin{pmatrix} & s_1 & s_2 & s_3 \\ \theta_1 & 2/3 & 1/3 & 0 \\ \theta_2 & 0 & 1/3 & 2/3 \end{pmatrix}$$

$$Q = \begin{pmatrix} & \tilde{s}_1 & \tilde{s}_2 \\ \theta_1 & 1/6 & 5/6 \\ \theta_2 & 5/6 & 1/6 \end{pmatrix}$$

Show that Q is a garbling of P. (How does this exercise relate to Example 4.1?)

EXERCISE 4.2 (G). *For any $q \in [1/2, 1]$, define the matrix*

$$S_q = \begin{pmatrix} q & 1-q \\ 1-q & q \end{pmatrix}$$

Suppose $q, q' \in [1/2, 1]$ with $q > q'$. Prove that $S_{q'}$ is a garbling of $S_{q'}$.

More generally, we can replace the Markov matrix M in Definition 4.2 with a Markov kernel.

DEFINITION 4.3 (Garblings, General Version). *The signal $\sigma' : \Theta \to \Delta(S')$ is a garbling of the signal $\sigma : \Theta \to \Delta(S)$ if there exists a Markov kernel $\gamma : S \to \Delta(S')$ such that*

$$\sigma'(s' \mid \theta) = \int_{s \in S} \gamma(s' \mid s)\sigma(s \mid \theta)ds$$

EXAMPLE 4.2. Let θ, ε, and δ be independent real-valued random variables with densities f_θ, f_ε, and f_δ. Then the signal $X = \theta + \varepsilon + \delta$ is a garbling of $Y = \theta + \varepsilon$, since

$$f_{X|\theta}(x \mid t) = \int_{y \in \mathbb{R}} f_\delta(x - y)f_{Y|\theta}(y \mid t)dy$$

where f_δ is a Markov kernel.

EXAMPLE 4.3. Consider an arbitrary finite set Θ and let I be the $|\Theta| \times |\Theta|$ identity matrix. Then for any set of signal realizations S and any $|\Theta| \times |S|$ Markov matrix Q, we have $IQ = Q$, so Q is a garbling of I.

EXERCISE 4.3 (U). *Is it possible for P and Q to both be garblings of one another if $P \neq Q$? Provide an example if so, and otherwise prove that it is not possible.*

REMARK 4.1. Let X and X' respectively denote the random realizations of the signals σ and σ'. Then θ, X, and X' are random variables which can be defined on a common probability space. The property that σ' is a garbling of σ does not however pin down the joint distribution of (θ, X, X'). What it guarantees is that there is a way of generating these variables such that θ is independent of X' conditional on X, in which case $\theta \mid X$ is identical in distribution to $\theta \mid X, X'$.[1] Other ways of generating these variables—still consistent with σ' being a garbling of σ—can yield different relationships.

For example, suppose $\theta \sim \mathcal{N}(0, 1)$ while

$$X = \theta + \varepsilon_1$$
$$X' = \theta + \varepsilon_2$$

where $\varepsilon_1 \sim \mathcal{N}(0, 1)$ and $\varepsilon_2 \sim \mathcal{N}(0, 2)$ are both independent of θ. Then clearly the latter signal is a garbling of the former. If we further assume that $\varepsilon_2 = \varepsilon_1 + \delta$ where $\delta \sim \mathcal{N}(0, 1)$ is an independent noise term, then the following statements are true:

---

[1] First draw the state θ, then draw X according to its conditional distribution, and finally draw X' according to the garbling kernel γ, independent of θ.

- X' is independent of θ conditional on X.

- X' is not independent of X conditional on θ (since they are further related through the common component ε_1).

On the other hand, if we assume that ε_1 and ε_2 are independent, then the statements above are reversed:

- X' is not independent of θ conditional on X (since X' provides additional information about θ beyond what is revealed by X).

- X' is independent of X conditional on θ.

Thus in general, the assumption that two signals are related by a garbling does not imply either conditional independence statement given above.

4.2   Decision Problems

Our next two definitions are based on the instrumental value of the signal for decision problems.

DEFINITION 4.4. *A decision problem is any pair* $\mathbf{D} = (A, u)$ *where A is an action set and $u : A \times \Theta \to \mathbb{R}$ is a payoff function.*

The full decision problem is described as follows. Fix a prior $p \in \Delta(\Theta)$ and a signal $\sigma : \Theta \to \Delta(S)$.

1. The agent chooses a strategy $\alpha : S \to A$.

2. The state $\theta \sim p$ and signal realization $s \sim \sigma(\cdot \mid \theta)$ are realized, and the agent takes action $\alpha(s)$. The agent's payoff is $u(\alpha(s), \theta)$.

Without the benefit of further information, the best expected payoff the agent can achieve is

$$\sup_{a \in A} \mathbb{E}\left[u(a, \theta)\right] \tag{4.1}$$

With the benefit of the signal, the agent can achieve an expected payoff of

$$\sup_{\alpha : S \to A} \mathbb{E}\left[u(\alpha(s), \theta)\right] = \mathbb{E}\left[\sup_{a \in A} \mathbb{E}\left[u(a, \theta) \mid s\right]\right] \tag{4.2}$$

where we abuse notation on the LHS by using s to denote the random variable which is the realization of the signal. On the RHS, the inner expectation is with respect to uncertainty about θ (conditional on the realization of s) and the outer expectation is with respect to uncertainty about s. One measure of the value of the signal is the difference in these expected payoffs, i.e.,

$$V_{\mathbf{D}, p}(\sigma) \equiv \mathbb{E}\left[\sup_{a \in A} \mathbb{E}\left[u(a, \theta) \mid s\right]\right] - \sup_{a \in A} \mathbb{E}\left[u(a, \theta)\right]$$

where $\mathbf{D} = (A, u)$ is the decision problem and $p \in \Delta(\Theta)$ is the agent's prior.

REMARK 4.2. It is without loss to assume the use of pure strategies above, but in the subsequent development of the Blackwell order it will be useful to replace a with a mixed strategy $\alpha \in \Delta(A)$ in (4.1) and α with a stochastic map $\alpha : S \to \Delta(A)$ in (4.2).

EXAMPLE 4.4. Suppose $\Theta = \{\theta_1, \theta_2\}$ with a uniform prior p. The decision problem is (A, u) where $A = \{a_1, a_2\}$ and the utility function $u : A \times \Theta \to \mathbb{R}$ assigns a payoff of 1 when the action matches the state, and zero otherwise. The signal σ is

$$
\begin{array}{ccc}
 & s_1 & s_2 \\
\theta_1 & q & 1-q \\
\theta_2 & 1-q & q
\end{array}
$$

where $q > 1/2$. Then the agent's ex-ante payoff is maximized by choosing the strategy α that maps s_1 to action a_1 and s_2 to action a_2, with an expected payoff of q. In the absence of information the agent's best payoff is $1/2$, so $V_{\mathbf{D},p}(\sigma) = q - 1/2$.

EXAMPLE 4.5. Suppose $\Theta = \mathbb{R}$ with a prior $\theta \sim \mathcal{N}(0, \sigma_\theta^2)$. The decision problem is (A, u) where $A = \mathbb{R}$ and $u(a, \theta) = -(a - \theta)^2$. The signal is $X = \theta + \varepsilon$ where $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ is independent of θ. Then the agent's ex-ante payoff is maximized by choosing the strategy $\alpha(x) = \mathbb{E}(\theta \mid X = x)$, with an expected payoff of

$$
\mathbb{E}_X \left[-(\mathbb{E}(\theta \mid X) - \theta)^2 \right] = \mathbb{E}_X \left[-\operatorname{Var}(\theta \mid X) \right] = -\frac{\sigma_\theta^2 \sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2}
$$

using Fact 2.2 in the final equality (and in particular, the property that posterior variance is independent of the signal realization). In the absence of information the agent's best payoff is $-\operatorname{Var}(\theta) = -\sigma_\theta^2$, so $V_{\mathbf{D},p}(X) = \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{\sigma_\theta^2 + \sigma_\varepsilon^2} - \sigma_\theta^2 = \frac{\sigma_\theta^4}{\sigma_\theta^2 + \sigma_\varepsilon^2}$.

In any specific decision problem, a signal that is informative (in the sense of moving the agent's beliefs about θ) may nevertheless have no instrumental value, as demonstrated in the following exercise.

EXERCISE 4.4 (U). *Suppose $\Theta = \{1, 2\}$ and let the prior p assign equal probability to either state. Consider the decision problem (A, u) with $A = \{1, 2\}$ and $u(a, \theta) = \mathbb{1}(a = \theta)$. Let σ_P and σ_Q respectively be the two signals described by P and Q in Example 4.1. Show that $V_{\mathbf{D},p}(\sigma_P) = V_{\mathbf{D},p}(\sigma_Q)$. That is, the second independent observation of signal P has no value to the agent over the first.*

4.2.1 Uniformly Better

We'll say that a signal is more informative than another if it is more useful in every decision problem and for every prior belief.

DEFINITION 4.5. *The signal σ is more informative than σ' if $V_{\mathbf{D},p}(\sigma) \geq V_{\mathbf{D},p}(\sigma')$ for every decision problem $\mathbf{D}$ and every prior p.*

This is a strong condition, and we generally won't be able to order signals in this way.

EXERCISE 4.5 (U). *Let $\Theta = \{\theta_1, \theta_2, \theta_3\}$ with a uniform prior p. Let $A = \{a_1, a_2\}$ and consider two utility functions: Let $u : A \times \Theta \to \mathbb{R}$ take value 1 if $(a, \theta) \in \{(a_1, \theta_1), (a_2, \theta_2), (a_2, \theta_3)\}$, and value 0 otherwise. Let $u' : A \times \theta \to \mathbb{R}$ take value 1 if $(a, \theta) \in \{(a_1, \theta_1), (a_2, \theta_2), (a_1, \theta_3)\}$, and value 0 otherwise. Consider the two information structures*

	s_1	s_2			s_1	s_2
θ_1	1	0		θ_1	1	0
σ: θ_2	0	1		σ': θ_2	0	1
θ_3	0	1		θ_3	1	0

Show that $V_{\mathbf{D},p}(\sigma) > V_{\mathbf{D},p}(\sigma')$ where $\mathbf{D} = (A, u)$, but $V_{\mathbf{D}',p}(\sigma) > V_{\mathbf{D}',p}(\sigma')$ where $\mathbf{D}' = (A, u')$, i.e. the agent prefers the first information given payoffs u and the second given payoffs u'.

The definition of uniformly better varies both the decision problem and also the prior, but the additional flexibility due to arbitrary priors is not substantial:

EXERCISE 4.6 (G). *Prove that if there is a full-support prior $p_0 \in \Delta(\Theta)$ such that*

$$V_{\mathbf{D},p_0}(\sigma) \geq V_{\mathbf{D},p_0}(\sigma') \quad \text{for every decision problem } \mathbf{D}$$

then σ is more informative than σ'.

4.2.2 Feasible Actions

Our third definition says that a signal is more informative if observing the realization of the signal allows the agent to more effectively tailor his action to the state.

DEFINITION 4.6. *Fix any action set A. A conditional distribution over actions $d : \Theta \to \Delta(A)$ is feasible under $\sigma : \Theta \to \Delta(S)$ if there exists a mapping $\alpha : S \to \Delta(A)$ such that*

$$d(a \mid \theta) = \int_{s \in S} \alpha(a \mid s)\sigma(s \mid \theta)ds$$

We'll use $\Lambda_\sigma(A)$ to denote the set of all feasible distributions under σ given action set A.

When σ is a fully revealing signal (e.g., $\sigma : \Theta \to \Delta(\Theta)$ satisfying $\sigma(\theta \mid \theta) = 1$ for every θ), then every mapping $d : \Theta \to \Delta(A)$ is feasible under σ. (Simply set $\alpha = d$.) When σ is uninformative—for example, a constant—then $\Lambda_\theta(A)$ consists of all mappings $d : \Theta \to \Delta(A)$ that take each state into the same distribution over actions. Larger sets $\Lambda_\sigma(A)$ allow the agent more flexibility in tailoring his action to the state, and in this sense are more valuable.

REMARK 4.3. Observe that α is itself a Markov kernel, so d can be interpreted as a garbling of σ where A is the set of signal realizations.

4.3 Dispersion of Posterior Beliefs

Our final perspective adopts the view on a signal introduced in Section 2.2, where a signal is identified with the distribution over posterior beliefs that it induces. We consider the dispersion of these posterior beliefs. Given an uninformative signal, the agent's posterior is deterministically equal to the agent's prior, so there is no dispersion. And if the signal reveals the state directly, then the posterior belief is a point mass on the true state, which "maximally varies" depending on the realization of the signal.

We may expect more informative signals to be associated with more dispersed beliefs, but the measure of dispersion is important. For example, using variance to measure dispersion yields a complete order on signals, which cannot possibly be equivalent to the (strict) partial order described in the previous definitions. Below we define two alternative measures for dispersion—mean-preserving spreads and dominance in the convex order—which will turn out to again characterize the previous partial order on signals.

4.3.1 Mean-Preserving Spreads

DEFINITION 4.7. *A distribution of posterior beliefs $F \in \Delta(\Delta(\Theta))$ is a* mean-preserving spread *of another distribution $\widetilde{F}$ if there exist $\Delta(\Theta)$-valued random variables $Z, \widetilde{Z}$ satisfying the following conditions:*

1. $Z \sim F, \widetilde{Z} \sim \widetilde{F}$

2. $\mathbb{E}(Z \mid \widetilde{Z}) = \widetilde{Z}$ *(thus in particular $\mathbb{E}(Z) = \mathbb{E}(\widetilde{Z})$)*

The name "mean-preserving spread" reflects that each realization of $\widetilde{Z}$ is spread out into a random Z with the same mean. When Z and $\widetilde{Z}$ are both real-valued, then the second condition can also be stated as $Z = \widetilde{Z} + \varepsilon$ for some random variable ε satisfying $\mathbb{E}(\varepsilon \mid \widetilde{Z}) = 0$.

EXAMPLE 4.6. Consider the two signals

$$P = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix} \qquad Q = \begin{pmatrix} 9/16 & 3/16 & 3/16 & 1/16 \\ 1/16 & 3/16 & 3/16 & 9/16 \end{pmatrix}$$

from Example 4.1, where the set of states is $\Theta = \{\theta_1, \theta_2\}$. Let the agent's prior be uniform over these states. Then the agent has two possible posterior beliefs after observing P, $(3/4, 1/4)$ and $(1/4, 3/4)$, which are equally likely. We will write the distribution of posterior beliefs as

$$F_P = 1/2 \cdot (3/4, 1/4) + 1/2 \cdot (1/4, 3/4).$$

Under Q, the distribution of posterior beliefs is instead

$$F_Q = 5/16 \cdot (9/10, 1/10) + 3/8 \cdot (1/2, 1/2) + 5/16 \cdot (1/10, 9/10).$$

We will now show that F_Q is a mean-preserving spread of F_P. Let $\widetilde{Z}$ be a random variable satisfying $\widetilde{Z} \sim F_P$ and construct the random variable Z given $\widetilde{Z}$ as follows:

- If $\widetilde{Z} = (1/4, 3/4)$ then $Z = (1/10, 9/10)$ with probability $5/8$ and $Z = (1/2, 1/2)$ with probability $3/8$.

- If $\widetilde{Z} = (3/4, 1/4)$ then $Z = (9/10, 1/10)$ with probability $5/8$ and $Z = (1/2, 1/2)$ with probability $3/8$.

Then $\mathbb{E}(Z \mid \widetilde{Z}) = \widetilde{Z}$ and also $Z \sim F_Q$, so F_Q is a mean-preserving spread of F_P as desired. This construction is depicted in Figure 4.1.
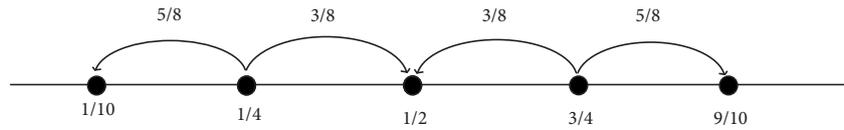

Figure 4.1: Depiction of the mean-preserving spread, where the numbers represent the probability of state θ_1.

Exercise 4.7 (G). *Let $\Theta = \{\theta_1, \theta_2\}$ and consider the two signals*

$$P = \begin{pmatrix} 2/3 & 1/3 \\ 1/4 & 3/4 \end{pmatrix} \qquad Q = \begin{pmatrix} 1/3 & 1/2 & 1/6 \\ 1/8 & 1/2 & 3/8 \end{pmatrix}$$

Define F_P to be the distribution of posterior beliefs induced by P and F_Q to be the distribution of posterior beliefs induced by Q. Prove that F_P is a mean-preserving spread of F_Q.

Exercise 4.8 (G). *Suppose $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random variables, and define $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ to be their sample average. Let $n' < n$ and define $\overline{Y}_{n'} = \frac{1}{n'} \sum_{i=1}^{n'} Y_i$. Prove that the distribution of $\overline{Y}_{n'}$ is a mean preserving spread of the distribution of $\overline{Y}_n$.*

4.3.2   Convex Order

Another partial order of dispersion is the following:

Definition 4.8. *A distribution of posterior beliefs $F \in \Delta(\Delta(\Theta))$ dominates another distribution G in the convex order *if for every continuous convex function* $h : \Delta(\Theta) \to \mathbb{R}$,*

$$\int_{\Delta(\Theta)} h(p) dF(p) \geq \int_{\Delta(\Theta)} h(p) dG(p)$$

This implies that F and G have the same mean (choosing $h(p) = p$) and that F has the larger variance (choosing $h(p) = \|p\|^2$).

You may recall the concept of *second order stochastic dominance*:

DEFINITION 4.9. *For any lotteries F and G, F second-order stochastically dominates G if and only if*

$$\int_{\Delta(\Theta)} u(p)dF(p) \geq \int_{\Delta(\Theta)} u(p)dG(p)$$

for every nondecreasing and concave utility function u.

Dominance in the convex order is stronger than SOSD.

EXERCISE 4.9 (G). *Prove that if F dominates G in the convex order, then G second order stochastically dominates F.*

The converse is not in general true.

EXAMPLE 4.7. *Let G be a distribution uniform on* $[1, 2]$ *and let F be a point mass at zero. Then G second order stochastically dominates F but F does not dominate G in the convex order.*

Intuitively, second-order stochastic dominance confounds changes in the dispersion of the distribution with shifts in the distribution, while dominance in the convex order isolates the former comparison.

4.4 Blackwell's Theorem and Proof

We now state and prove Blackwell (1951)'s theorem, which demonstrates equivalence of these five definitions. For the proof we will work with finite sets (in particular assuming finite Θ) but several parts of the proof extend more generally.

Theorem 4.1. *The following are equivalent:*

1. σ' *is a garbling of* σ.

2. σ *is more informative than* σ'.

3. $\Lambda_\sigma(A) \supseteq \Lambda_{\sigma'}(A)$ *for every finite action set A.*

4. *For any prior on* Θ, *if we define F and F' to be the distributions of posterior beliefs induced by* σ *and* σ' *(under this prior), then F is a mean-preserving spread of F'.*

5. *For any prior on* Θ, *if we define F and F' to be the distributions of posterior beliefs induced by* σ *and* σ' *(under this prior), then F dominates F' in the convex order.*

Several proofs exist for different parts of this result (see e.g., Blackwell (1951) and Leshno and Spector (1992)). Our proof of the equivalence of (1)-(3) below is based on de Oliveira (2019), which presents a particularly simple and elegant argument.

Proof. Throughout, given stochastic mappings $\alpha : X \to \Delta(Y)$ and $\beta : Y \to \Delta(Z)$, let

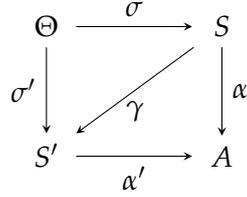$$\beta \circ \alpha(z \mid x) \equiv \sum_{y \in Y} \beta(z \mid y)\alpha(y \mid x) \qquad \forall x \in X, z \in Z.$$

$(1 \Rightarrow 3)$ 1 implies existence of a mapping $\gamma : S \to \Delta(S')$ such that $\gamma \circ \sigma = \sigma'$, as illustrated below:
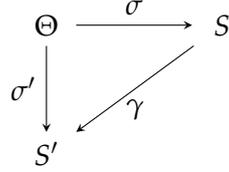
Consider any action set A and mapping $\alpha' : S' \to \Delta(A)$, where $d = \alpha' \circ \sigma'$ is a feasible distribution under σ'. Define $\alpha = \alpha' \circ \gamma$. Then

$$\alpha \circ \sigma = (\alpha' \circ \gamma) \circ \sigma = \alpha' \circ (\gamma \circ \sigma) = \alpha' \circ \sigma' = d$$

using associativity of the operation $\circ$. So d is feasible also under σ, as depicted in the figure below:

$(3 \Rightarrow 1)$ Let the action set be S' and define α' to be the identity mapping $id_{S'} : S' \to \Delta(S')$ which satisfies $id_{S'}(s') = \delta_{s'}$ for all $s' \in S'$ (where $\delta_{s'}$ denotes a point mass at s'). By 3, there must exist some $\alpha : S \to \Delta(S')$ such that

$$\alpha \circ \sigma = id_{S'} \circ \sigma'$$

The RHS reduces to σ' since for any $s' \in S'$,

$$id_{S'} \circ \sigma'(s' \mid \theta) = \sum_{s \in S'} id_{S'}(s' \mid s)\sigma'(s \mid \theta) = \sigma'(s' \mid \theta).$$

Thus $\alpha \circ \sigma = \sigma'$. But this implies that σ' is a garbling of σ, as depicted below.
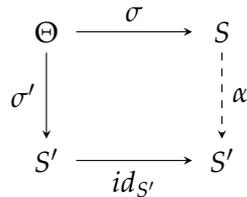
(3 $\Rightarrow$ 2) Clear.

(2 $\Rightarrow$ 3) Suppose 3 fails. Then there is a finite action set A and a vector $\lambda' \in \Lambda_{\sigma'}(A)$ such that $\lambda' \notin \Lambda_{\sigma}(A)$. The set Λ_{σ} is a compact and convex subset of $\mathbb{R}^{|\Theta| \times |A|}$ (you will be asked to prove this in Exercise 4.12). Thus by the Separating Hyperplane Theorem, there exists a vector $v \in \mathbb{R}^{|\Theta| \times |A|}$ such that for all $\lambda \in \Lambda_{\sigma}(A)$,

$$\sum v(a,\theta)\lambda(a,\theta) < \sum v(a,\theta)\lambda'(a,\theta) \tag{4.3}$$
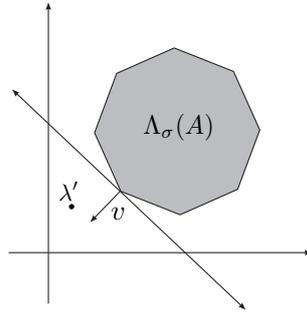
as depicted in Figure 4.2.

Figure 4.2: Separation of λ' from $\Lambda_{\sigma}(A)$.

Consider an agent with a uniform prior p on Θ and utility function v, and define $d(a \mid \theta) \equiv \lambda(a,\theta)$ and $d'(a \mid \theta) \equiv \lambda'(a,\theta)$. Then

$$\sup_{\alpha:S\to\Delta(A)} \sum_{a,s,\theta} v(a,\theta)\alpha(a \mid s)p(\theta,s) = \sup_{\alpha:S\to\Delta(A)} \frac{1}{|\Theta|} \sum_{\theta,a,s} \sigma(s \mid \theta)\alpha(a \mid s)v(a,\theta)$$

$$= \sup_{d\in\Lambda_{\sigma}(A)} \frac{1}{|\Theta|} \sum_{\theta,a} d(a \mid \theta)v(a,\theta)$$

$$< \frac{1}{|\Theta|} \sum_{\theta,a} d'(a \mid \Theta)v(a,\theta)$$

using (4.3) in the final inequality. Thus there is a decision problem and a prior for which an agent can achieve a strictly higher payoff by conditioning on σ' rather than on σ, and so 2 fails.

(1 $\Rightarrow$ 4) Let X and X' respectively denote the random realizations of the signals σ and σ'. Since by assumption σ' is a garbling of σ, we can generate θ, X, X' in a way such that X' is independent of θ conditional on X (see Remark 4.1). On this probability space, define Z to be the random posterior belief of θ given X, i.e., the distribution of $\theta \mid X$, and define Z' to be the random posterior belief of θ given X', i.e., the distribution of $\theta \mid X'$. We need to show that $\mathbb{E}[Z \mid Z'] = Z'$.

For any realization θ_i of θ, define $Z_i \equiv \mathbb{E}[\mathbb{1}_{\theta_i} \mid X] = \mathbb{E}[\mathbb{1}_{\theta_i} \mid X, X']$ (where the second equality is due to independence of θ and X' conditional on X) and

define $Z_i' \equiv \mathbb{E}[\mathbb{1}_{\theta_i} \mid X']$. Then

$$
\begin{aligned}
\mathbb{E}[Z_i \mid X'] &= \mathbb{E}[[\mathbb{1}_{\theta_i} \mid X, X'] \mid X'] \\
&= \mathbb{E}[\mathbb{1}_{\theta_i} \mid X'] \\
&= Z_i'
\end{aligned}
\tag{4.4}
$$

where the second equality follows from the law of iterated expectations (henceforth abbreviated to L.I.E.). Moreover,

$$
\begin{aligned}
\mathbb{E}[Z_i \mid Z'] &= \mathbb{E}[\mathbb{E}[Z_i \mid X', Z'] \mid Z'] &&\text{by L.I.E.} \\
&= \mathbb{E}[\mathbb{E}[Z_i \mid X'] \mid Z'] &&\text{since } Z' \text{ is a function of } X' \\
&= \mathbb{E}[Z_i' \mid Z'] &&\text{using (4.4)} \\
&= Z_i'
\end{aligned}
$$

Repeating this argument for every θ_i, we have the desired result.

$(4 \Rightarrow 5)$ Suppose F is a MPS of F' with associated random variables Z and Z' satisfying $\mathbb{E}(Z \mid Z') = Z'$. Then for any continuous and convex function $h : \Delta(\Theta) \to \mathbb{R}$,

$$
\begin{aligned}
\int_{\Delta(\Theta)} h(p)dF(p) &= \mathbb{E}[h(Z)] \\
&= \mathbb{E}[\mathbb{E}[h(Z) \mid Z']] &&\text{by L.I.E.} \\
&\geq \mathbb{E}[h(\mathbb{E}[Z \mid Z'])] &&\text{by Jensen's inequality} \\
&= \mathbb{E}[h(Z')] &&\text{by assumption of MPS} \\
&= \int_{\Delta(\Theta)} h(p)dF'(p)
\end{aligned}
$$

So F dominates F' in the convex order.

$(5 \Rightarrow 2)$ Fix any action set A and utility function u, and define $h : \Delta(\Theta) \to \mathbb{R}$ by

$$
h(p) = \max_{a \in A} \sum_{\theta \in \Theta} p(\theta)u(a, \theta)
$$

to be the maximum achievable payoff under belief p. The function h is the pointwise maximum of linear functions, and hence it is continuous and convex. Letting $p \sim F$ denote the agent's posterior belief, the maximum *ex-ante* payoff is

$$
\int_{\Delta(\Theta)} h(p)dF(p)
$$

So dominance in the convex order implies "more valuable." ∎

4.5 Additional Exercises

EXERCISE 4.10 (U). *The state space is $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and an agent's prior belief is $p = (1/2, 1/4, 1/4)$. The agent chooses from actions in the set $A = \{a_1, a_2, a_3\}$ and*

has the payoff function

$$u(a, \theta) = \begin{cases} 1 & \text{if } a = \theta \\ 0 & \text{otherwise} \end{cases}$$

(a) What is the highest expected payoff that the agent can achieve?

(b) Now suppose the agent gets to see the outcome of the following signal structure before choosing an action:

	s_1	s_2
θ_1	0	1
θ_2	1/2	1/2
θ_3	1	0

(i) Suppose the realized signal outcome is s_1. Solve for the agent's posterior belief and optimal action.

(ii) Suppose the realized signal outcome is s_2. Solve for the agent's posterior belief and optimal action.

(iii) What is the agent's best expected payoff (where the expectation is taken prior to the realization of the signal outcome)?

EXERCISE 4.11 (U). *(This problem is based on Meyer (1991).) Consider the setting of Example 4.4. It turns out that we can make the second realization of P strictly valuable again by biasing it in favor of the more likely signal realization. That is, let the realizations of P be denoted s_1 and s_2, where*

$$P = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix}$$

and modify the second signal in the following way: If the first realization is s_1, then the second signal realization is determined by

$$Q_1 = \begin{pmatrix} 3/4 + c & 1/4 - c \\ 1/4 + c & 3/4 - c \end{pmatrix}$$

and if the first realization is s_2, the second signal realization is determined by

$$Q_2 = \begin{pmatrix} 3/4 - c & 1/4 + c \\ 1/4 - c & 3/4 + c \end{pmatrix}$$

where in both cases the realization of the second signal is independent of the first conditional on the state.

(a) *Show that for any $c \in (0, 1/4]$, the value of observing this second (biased) signal is strictly positive.*

(b) *Solve for the size of the bias $c \in (0, 1/4]$ that leads to the highest expected payoff for the agent.*

EXERCISE 4.12 (G). *Let the sets A, Θ, and S be finite, and prove that the set $\Lambda_\sigma(A)$ (from Definition 4.6) is compact and convex for every $\sigma : \Theta \to \Delta(S)$.*

EXERCISE 4.13 (G). *Consider two random variables $X = \theta + \varepsilon$ and $Y = \theta + \varepsilon'$, where θ, ε, and ε' are mutually independent.*

(a) *Suppose that $\theta \sim \mathcal{N}(0, 1)$ and $\varepsilon, \varepsilon' \in \mathbb{R}$ are distributed $(\varepsilon, \varepsilon') \sim \mathcal{N}(\mu, \Sigma)$. Prove that X and Y are Blackwell comparable for all mean vectors μ and covariance matrices Σ.*

(b) *Suppose that*

$$\theta \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

and $\varepsilon, \varepsilon' \in \mathbb{R}^2$ are distributed $(\varepsilon, \varepsilon') \sim \mathcal{N}(\mu, \Sigma)$. Prove that X and Y are not always Blackwell ranked by demonstrating a pair (μ, Σ) such that X allows for a strictly higher expected payoff for one decision problem, and Y allows for a strictly higher expected payoff given another.

EXERCISE 4.14 (G). *In each of the following parts, determine whether the statement is true or false and prove your claim in either case.*

(a) *The state θ belongs to $\{\theta_1, \theta_2\}$ and the two signals are defined as*

$$X = \theta + \varepsilon_1, \quad \varepsilon_1 \sim U([-1/2, 1/2])$$
$$\widetilde{X} = \theta + \varepsilon_2, \quad \varepsilon_2 \sim U([-1/3, 1/3])$$

where U denotes the uniform distribution. The signals X and $\widetilde{X}$ can be Blackwell ranked.

(b) *The state θ belongs to $\{0, 1/3, 2/3, 1\}$ and the two signals are defined as*

$$X = \theta + \varepsilon_1, \quad \varepsilon_1 \sim U([-1/2, 1/2])$$
$$\widetilde{X} = \theta + \varepsilon_2, \quad \varepsilon_2 \sim U([-1/3, 1/3])$$

The signals X and $\widetilde{X}$ can be Blackwell ranked.

EXERCISE 4.15 (G). *(This problem is based on Brooks, Frankel and Kamenica (2022a).) Consider the following strengthening of the Blackwell order. Let θ, X, and X' be random variables defined on the same probability space (Ω, Σ, P).*

DEFINITION 4.10. Say that X strongly Blackwell dominates X' if $(X, \widetilde{X})$ Blackwell dominates $(X', \widetilde{X})$ for every random variable $\widetilde{X}$ also defined on (Ω, Σ, P).

Clearly a necessary condition is for X to Blackwell dominate X' (choose $\widetilde{X}$ to be null information). A sufficient condition is for the realization of X' to be known from the realization of X', i.e., for the distribution of $X' \mid X$ to be degenerate for every realization of X (what Brooks, Frankel and Kamenica (2022b) call the refinement order*). Provide an example in between, namely a signal X that strongly Blackwell dominates X', where the realization of X' is not known from X.*