Chapter 8

Information Design

8.1 Bayesian Persuasion

8.1.1 Example

A judge and a prosecutor are involved in a court case. The unknown payoff-relevant state is whether the defendant in this case (who will not take an action) is *innocent* (I) or *guilty* (G). The judge and the prosecutor share a common prior that the defendant is guilty with probability 0.3.

The prosecutor cannot falsify or distort evidence, but can selectively choose what kind of information to present to the court (e.g., deciding who to subpoena or which forensic tests to conduct). Formally, the prosecutor chooses an information structure $\sigma : \{G, I\} \to \Delta(S)$ for some set of signal realizations S. The judge observes the outcome of the signal σ, updates his beliefs, and chooses whether to *acquit* or *convict* the defendant.

The judge and prosecutor's payoffs are determined by the judge's action and by the unknown state. The judge receives a payoff of 1 from convicting a guilty defendant or from acquitting an innocent defendant, and otherwise receives a payoff of zero. The prosecutor receives a payoff of 1 if the judge convicts the defendant and a payoff of 0 if the judge acquits the defendant, independent of the defendant's guilt. What information structure should the prosecutor choose, and what is the best expected payoff he can achieve?

Let's start with some benchmarks. One possibility is to send a completely uninformative signal. Since innocence is more likely than guilt under the judge's prior, the judge chooses to acquit given no information, yielding a payoff of zero for the prosecutor. Alternatively, the prosecutor can choose a perfectly informative signal that reveals the defendant's guilt. The judge convicts precisely when the defendant is guilty, yielding an expected payoff (under the prior) of 0.3 for the prosecutor.

Can the prosecutor do better? The perfectly revealing signal splits defendants into two bins—one labeled "convict" and one labelled "acquit" (Figure 8.1).

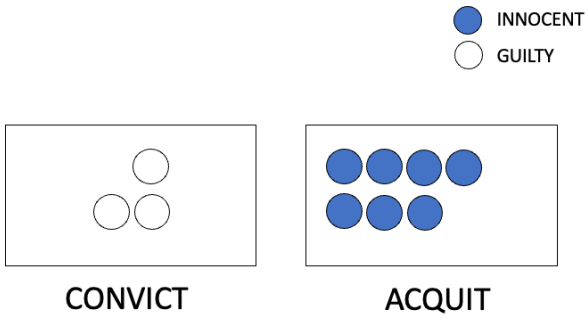

Figure 8.1: Depiction of the perfectly revealing signal, where each circle represents 1/10 of the population.

The judge's posterior for individuals labeled "convict" is that they are guilty with probability 1, so he optimally convicts any individual with this label. Likewise his posterior for individuals labeled "acquit" is that they are innocent with probability 1, so he acquits any individual with this label.

Now consider moving one unit of innocent individuals from the acquit bin to the convict bin (Figure 8.2).

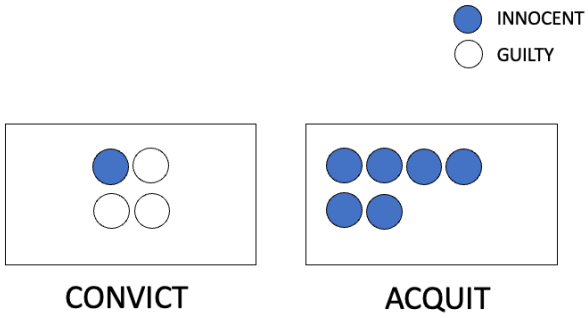

Figure 8.2: Deviation from the perfectly revealing signal.

Remark 8.1. Every "bin representation" as shown in Figures 8.1 and 8.2 corresponds to a unique signal. For each $\theta \in \Theta$ and $s \in \{\text{convict}, \text{acquit}\}$, let $P(\theta, s)$ be the mass of θ-type units in bin s (interpreting each circle as 1/10 of the population). Then P is a probability measure on $\Theta \times S$, and the corresponding signal $\sigma : \Theta \to \Delta(S)$ can be derived by Bayes' rule. As we see in the proof of Proposition 25, every signal also admits a bin representation.[1]

Following this modification on the perfectly revealing signal, the posterior probability of guilt in the acquit bin is unchanged. The posterior probability of

[1] In particular, every signal admits a "bin representation" that consists of two bins—a convict bin, and an acquit bin—where the judge optimally convicts all individuals in the convict bin and acquits all individuals in the acquit bin.

guilt for individuals labeled "convict" drops to 3/4—but crucially, the judge's optimal action remains the same. Intuitively, by pooling innocent defendants with guilty defendants (but maintaining sufficiently guilty defendants that the judge still wants to convict), the prosecutor is able to induce the judge to wrongly convict a larger number of defendants.

Iterating this logic, we can continue moving units of innocent individuals into the convict bin, up until the judge is indifferent between convicting and acquitting (Figure 8.3).

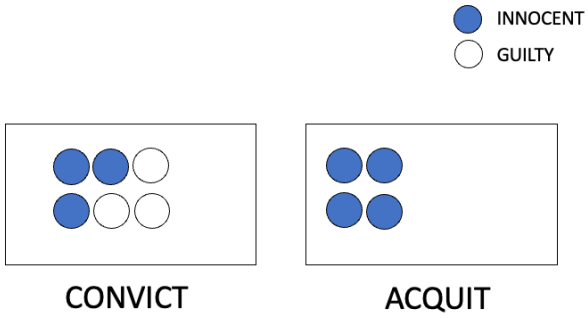

Figure 8.3: Depiction of the prosecutor-optimal signal structure.

These bins correspond to the following signal structure:

$$\begin{array}{ccc} & \textit{convict} & \textit{acquit} \\ G & 1 & 0 \\ I & 3/7 & 4/7 \end{array} \tag{8.1}$$

That this signal structure is optimal will follow from the results in the subsequent section. Strikingly, although the judge knows that only 30% of defendants are guilty, he ends up convicting 60% of them.

8.1.2 Model

There are two agents, a Sender and a Receiver. The unknown parameter θ takes values in the finite set Θ, and agents share a common prior $\mu_0 \in \Delta(\Theta)$. A signal is any mapping $\sigma : \Theta \to \Delta(S)$ from the set of states into distributions over a finite set of signal realizations S.

The Receiver chooses from a compact set of actions A. Both agents' payoffs depend on the Receiver's action and the unknown state. We'll denote the Receiver's utility function by $u_R : A \times \Theta \to \mathbb{R}$ and the Sender's utility function by $u_S : A \times \Theta \to \mathbb{R}$, where both are assumed to be continuous.

The timeline is as follows: First, the Sender chooses a signal σ. The realization of this signal is then observed by the Receiver, who updates his beliefs and chooses an action $a \in A$. Finally payoffs are realized. The solution concept is

Sender-Preferred subgame perfect equilibrium; that is, the Receiver chooses an action to maximize his expected payoffs, breaking ties between optimal actions by maximizing Sender's payoffs.[2]

8.1.3 Solution and Geometric Representation

Consider any Sender-Preferred subgame perfect equilibrium, and let $\hat{a}(\mu)$ denote the Receiver's action given belief $\mu \in \Delta(\Theta)$ in this equilibrium. That is,

$$\hat{a}(\mu) \in \arg\max_{a \in A(\mu)} \mathbb{E}_\mu \left[u_S(a, \theta) \right] \tag{8.2}$$

where

$$A(\mu) = \arg\max_{a \in A} \mathbb{E}_\mu \left[u_R(a, \theta) \right]$$

is the set of actions that maximize the Receiver's expected payoff given belief μ. (If the RHS of (8.2) is non-empty, set $\hat{a}(\mu)$ to be any action in this set.) Let

$$\hat{v}(\mu) := \mathbb{E}_\mu \left[u_S(\hat{a}(\mu), \theta) \right] \tag{8.3}$$

be the Sender's expected payoff given belief μ and Receiver-action $\hat{a}(\mu)$. A signal's *value* is the Sender's (ex-ante) expected payoff given choice of that signal.

Proposition 25 (Kamenica and Gentzkow (2011))**.** *The following are equivalent:*

(i) There exists a (finite-valued) signal with value v^*.

(ii) There exists a (finite-valued) signal taking realizations in $S \subseteq A$ with value v^*.

(iii) There exists a Bayes-plausible distribution over posterior beliefs, $\tau \in \Delta(\Delta(\Theta))$, such that $\mathbb{E}_\tau \left[\hat{v}(\mu) \right] = v^*$.

Proof. The implication (ii) $\Rightarrow$ (i) is immediate. The implication (ii) $\Rightarrow$ (iii) follows from Fact 2.1 (every signal induces a Bayes-plausible distribution over posterior beliefs).

To show (i) $\Rightarrow$ (ii), observe that for any signal $\sigma : \Theta \to \Delta(S)$ with value v^*, we can define a new signal $\widetilde{\sigma} : \Theta \to \Delta(A)$ that maps types into the recommended action under σ. That is,

$$\widetilde{\sigma}(a \mid \theta) = \sum_{s : \hat{a}(\mu_s) = a} \sigma(s \mid \theta)$$

for every $a \in A$ and $\theta \in \Theta$, where μ_s denotes the Receiver's posterior given signal realization s under σ. (The number of distinct action recommendations cannot exceed the size of S and so is finite.) Clearly the optimal action given recommendation of a remains the action a, so the distribution of optimal actions induced by $\widetilde{\sigma}$ and σ are the same.

[2] If there are multiple such actions, the Receiver chooses any action between them.

The direction (iii) $\Rightarrow$ (i) is nearly immediate from Proposition 3 (every Bayes-plausible distribution over posterior beliefs can be induced by a signal), but we need to show that it is possible to construct a *finite-valued* signal for arbitrary τ (even ones with infinite support).[3]

We'll use the following result from convex analysis.

Proposition 26 (Caratheodory's Theorem)**.** *Let $X \subseteq \mathbb{R}^n$ be a nonempty subset of finite-dimensional Euclidean space. Let $conv(X)$ denote the convex hull of X. Then every vector in $conv(X)$ can be represented as a convex combination of at most $n+1$ vectors from X.*

Fix any v^* and Bayes-plausible τ such that $\mathbb{E}_\tau[\hat{v}(\mu)] = v^*$. Define

$$C = \{(\mu, \hat{v}(\mu)) \mid \mu \in \Delta(\Theta)\}$$

to be the set of all beliefs and valuations of those beliefs, noting that $C \subseteq \mathbb{R}^n$ where $n \equiv |\Theta|$.[4] Moreover, by assumption that $v^* = \mathbb{E}_\tau[\hat{v}(\mu)]$ for some Bayes-plausible distribution τ over posterior beliefs, the vector (μ_0, v^*) belongs to the convex hull of C.

Then by Caratheodory's Theorem, there exists a sequence of beliefs $(\mu_i)_{i=1}^{n+1}$ and a sequence of nonnegative weights $(\alpha_i)_{i=1}^{n+1}$ summing to 1, such that

$$(\mu_0, v^*) = \sum_{i=1}^{n+1} \alpha_i \cdot (\mu_i, \hat{v}(\mu_i))$$

Let τ^* be the distribution over posterior beliefs that assigns probability α_i to each belief μ_i, $1 \leq i \leq n+1$. Then

$$\mathbb{E}_{\tau^*}[\hat{v}(\mu)] = \sum_{i=1}^{n+1} \alpha_i \cdot \hat{v}(\mu_i) = v^*$$

as desired. Follow the construction in Section 2.2.2 (setting the set of signal realizations S to be the posterior beliefs in the support of τ^*) to complete the proof. ■

Proposition 25 tells us that we can determine when the Sender benefits from persuasion by studying how $\mathbb{E}_\tau[\hat{v}(\mu)]$ varies over the set of Bayes-plausible distributions.

Corollary 8.1. *The Sender benefits from persuasion if and only if there exists a Bayes-plausible distribution τ such that $\mathbb{E}_\tau[\hat{v}(\mu)] > \hat{v}(\mu_0)$.*

Corollary 8.2. *The value of an optimal signal is*

$$\max_{\tau \in \Delta(\Theta)} \mathbb{E}_\tau[\hat{v}(\mu)] \quad s.t. \int \mu d\tau(\mu) = \mu_0$$

[3] The construction in Section 2.2.2 chooses S to be the set of all beliefs in the support of τ, which need not be finite.

[4] The simplex $\Delta(\Theta)$ is a subset of $\mathbb{R}^{n-1}$ and the valuation belongs to $\mathbb{R}$, hence $C \subseteq \mathbb{R}^n$.

The value of information for the Sender at any prior μ can be represented geometrically using the upper concave envelope of $\hat{v}$.

Definition 8.1. *Define*

$$V(\mu) \equiv \sup\{z \mid (\mu, z) \in Conv(\hat{v})\} \quad \forall \mu \in \Delta(\Theta)$$

where $Conv(\hat{v})$ denotes the convex hull of the graph $\hat{v}$. That is, V is the smallest concave function that is everywhere weakly greater than $\hat{v}$.

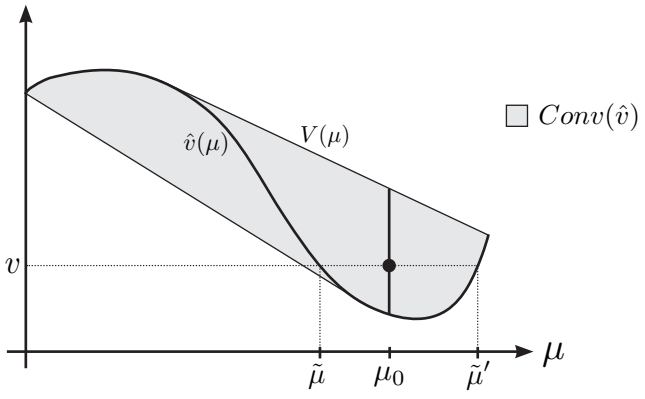

Figure 8.4: Illustration of Definition 8.1.

By Proposition 25, the set $\{z \mid (\mu_0, z) \in Conv(\hat{v})\}$ is precisely those expected payoffs that the Sender can achieve when the prior μ_0. For example, in Figure 8.1, the value v is achievable from the prior μ_0 via a signal that splits the prior into two posterior $\tilde{\mu}$ and $\tilde{\mu}'$ (setting the weights so that the expected posterior equals the prior). So $V(\mu_0) = \sup\{z \mid (\mu_0, z) \in Conv(\hat{v})\}$ is the largest payoff Sender can achieve when the prior is μ_0, and the Sender strictly benefits from persuasion if and only if $V(\mu_0) > \hat{v}(\mu_0)$.

The following corollary is immediate from the previous analysis.

Corollary 8.3. *If $\hat{v}$ is concave, then the Sender does not benefit from persuasion for any prior. If $\hat{v}$ is strictly convex, the Sender benefits from persuasion for every prior.*

8.1.4 Back to the Example

Returning to the setting of Section 8.1.1, observe that in any Sender-preferred subgame equilibrium, the judge's action given probability of guilt μ is

$$\hat{a}(\mu) = \begin{cases} \textit{convict} & \text{if } \mu \geq 0.5 \\ \textit{acquit} & \text{if } \mu < 0.5 \end{cases}$$

where the tie at $\mu = 0.5$ is broken in favor of the prosecutor. So the prosecutor's expected payoff is

$$\hat{v}(\mu) = \begin{cases} 1 & \text{if } \mu \geq 0.5 \\ 0 & \text{if } \mu < 0.5 \end{cases}$$

as depicted in Panel (a) of Figure 8.5.

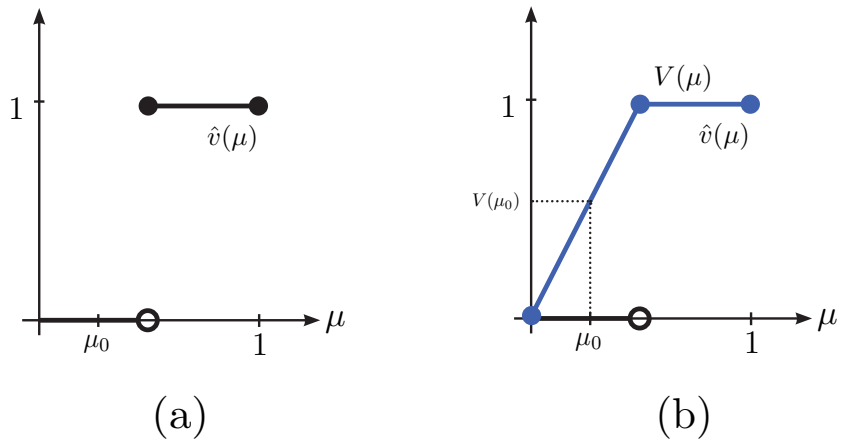

Figure 8.5: Depiction of $\hat{v}(\mu)$ in the prosecutor-judge example.

The upper concave envelope of $\hat{v}$ is

$$V(\mu) = \begin{cases} 1 & \text{if } \mu \geq 0.5 \\ 2\mu & \text{if } \mu < 0.5 \end{cases}$$

as depicted in Panel (b) of Figure 8.5. At the prior belief of $\mu_0 = 0.3$, we have $V(0.3) = 0.6$, confirming that the signal structure in (8.1) delivers the best possible expected payoff for the prosecutor. We see moreover that the prosecutor benefits from persuasion whenever $\mu_0 < 0.5$ (i.e., whenever the judge would optimally acquit under the prior), but cannot improve his expected payoff through choice of any signal structure when $\mu_0 \geq 0.5$.

8.2 Additional Exercises

EXERCISE 8.1 (U). *A student (Sender)'s quality is $\theta \in \{L, H\}$. The employer chooses an action from $A = \{l, m, h\}$ where l is a low-responsibility position, m is a medium-responsibility position, and h is a high-responsibility position. The employer's payoffs are:*

$$u_E(a, \theta) = \begin{cases} 1 & \text{if } (a, \theta) = (H, h) \\ 0 & \text{if } (a, \theta) \in \{(H, m), (H, l), (L, l)\} \\ -1 & \text{if } (a, \theta) \in \{(L, m), (L, h)\} \end{cases}$$

The student's (state-independent) payoff function u_S takes value 1 if $a = h$, 0 if $a = m$, and -1 if $a = l$.

(a) Suppose the employer's beliefs are described as $(p, 1-p)$, where p is the probability of $\theta = L$. Let

$$\hat{a}(p) = \arg\max_{a \in \{l,m,h\}} \mathbb{E}_{(p,1-p)}[u_E(a,\theta)].$$

(This is the same as in (8.2), except we simplify notation by writing $\hat{a}(p)$ instead of $\hat{a}(p, 1-p)$.) Solve for $\hat{a}(p)$ on the domain $p \in [0,1]$, assuming that the employer breaks ties in favor of the action that maximizes the student's payoffs.

(b) Suppose the student's beliefs are described as $(p, 1-p)$, where p is the probability of $\theta = L$, and the student knows that the employer chooses action $\hat{a}(p)$. Let

$$\hat{v}(p) = \mathbb{E}_{(p,1-p)}[u_S(\hat{a}(p),\theta)]$$

denote the student's expected payoff at this belief. Solve for $\hat{v}(p)$ on the domain $p \in [0,1]$ and plot it.

(e) Let $V(p)$ be the smallest concave function that is everywhere above $\hat{v}(p)$. Reproduce your plot from part (d) with $V(p)$ and $\hat{v}(p)$ depicted in the same figure. Clearly label $V(p)$ and $\hat{v}(p)$.

(f) Identify all $p \in [0,1]$ such that $V(p) > \hat{v}(p)$. These are the prior beliefs at which the student can strictly benefit from design of the signal structure.

Exercise 8.2 (G). *Fix an arbitrary finite set of states Θ and finite set of actions A. Suppose that the Sender and Receiver's payoff functions satisfy*

$$u_S(a,\theta) = -u_R(a,\theta)$$

for every $a \in A$ and $\theta \in \Theta$. Prove that $V(\mu) = \hat{v}(\mu)$ for every belief μ, where $\hat{v}$ is as defined in (8.3) and V is as given in Definition 8.1. Interpret this result.