Unraveling

Monday, Jan 12, 2015

We've seen two criteria for evaluating gambles (lotteries or risky assets). FOSD and SOSD.

Next we study criteria for evaluating information structures.

1 Information Structures

1.1 Information Structure without Noise

- Let $(\Omega, \mathcal{O})$ be a measurable space of states of nature

- The prior information of an agent is represented by a probability measure π on $(\Omega, \mathcal{O})$

- For example, if the agent is buying a commondity, $\Omega = \{\omega_1, \omega_2, \omega_3\}$ can be good, average and bad quality

- $\pi(\omega_1)$ is the probability the agent assigns to the product being of good quality

Definition 1 *An information structure* without noise *consists of a space of signals Y and a measurable function φ from the space of states to Y*

INSERT FIGURE

- The function φ defines a partition of Ω, the elements of which are given by

$$O_i = \varphi^{-1}(y_i) \text{ for } y_i \in Y.$$

- A decision maker wishes to maximize his objective function $u(\omega, a)$ with respect to his action $a \in A$ without knowing ω

- If he is rational, he maximizes expected utility

$$\max_{a \in A} \int_{\Omega} u(a, \omega) \pi(\omega) d\pi \quad (1)$$

- Let a^{*0} be the solution to problem (1).

- Let $\mathcal{P}_1 = \{O_1(y), y \in Y_1\}$ be the partition generated by information structure 1.

- EXAMPLE.

- Denote $v(\omega|y)$ the posterior probability distribution. If $\omega \notin O_1(y)$ then $v(\omega|y) = 0$, otherwise

$$v(\omega|y) = \frac{\pi(\omega)}{\int_{O_1(y)} \pi(\tilde{\omega})\, d\tilde{\omega}}.$$

- Namely, the agent revises his beliefs using Bayes' Theorem.

- For each value of y the agent knows that he will solve the following problem

$$\begin{aligned} & \max_{a \in A} \int_{\Omega} u(a, \omega)\, v(\omega|y)\, d\omega \\ = & \int_{\Omega} u\left(a_1^*(y), \omega\right) v(\omega|y)\, d\omega = V(y). \end{aligned}$$

- He can evaluate ex ante the value of having information structure $\mathcal{P}_1$

$$U(\mathcal{P}_1) = \int_{Y^1} V(y)\, \pi(y)\, dy$$

- where $\pi(y)$ is the prior probability of having the signal y, that is

$$\pi(y) = \int_{O_1(y)} \pi(\omega)\, d\omega.$$

- We say that information structure 1 is better than information structure 2 for the agent if

$$U(\mathcal{P}_1, \pi, u) > U(\mathcal{P}_2, \pi, u).$$

- Clearly, this comparision depends on the agent's preferences u and prior beliefs π.

- But can we compare information structures independently from these characteristics?

Definition 2 *We say that information structure* 1 *is* finer *than information structure* 2 *if the partition generated by structure 1 is finer than the one generated by structure 2, that is* $\forall O_2 \in \mathcal{P}_2$ *there is* $\left\{O_i^1\right\}_{i=1}^k : \cup_{i=1}^k O_i^1 = O_2$.

Theorem 1 *Information structure 1 is finer than information structure 2 if and only if for any prior probability distribution* π *and for any utility function* u*:* $U(\mathcal{P}_1, \pi, u) \geq U(\mathcal{P}_2, \pi, u)$.

Proof. if (1) is finer than (2), then $\forall y^2$ and O^2, there exists $O_1^1, .., O_k^1$ such that $\cup_{j=1}^k O_j^1$. Fix a signal y_2. Let $a^{*2}(y_2)$ be the optimal action given signal y^2. Since information structure 1 is finer than information structure 2, then there is a set of k signals $y_1^1, y_2^1, y_j^1, ..., y_k^1$ that I would have observed if info struct 1 was in place. By definition we have given signal y_j^1

$$\max_a \int_{O_j^1} u(a, \omega) v\left(\omega | y_j^1\right) d\omega \geq \int_{O_j^1} u\left(a^{*2}\left(y^2\right), \omega\right) v\left(\omega | y_j^1\right) d\omega.$$

Let $a^{*1}\left(y_j^1\right)$ be the solution to the above problem for $j = 1..., k$. We have

$$\begin{aligned} & \sum_{j=1}^{k} \pi\left(y_j^1\right) \int_{O_j^1} u\left(a^{*1}\left(y_j^1\right), \omega\right) v\left(\omega | y_j^1\right) d\omega \\ \geq & \sum_{j=1}^{k} \pi\left(y_j^1\right) \int_{O_j^1} u\left(a^{*2}\left(y_2\right), \omega\right) v\left(\omega | y_j^1\right) d\omega, \end{aligned}$$

since this holds for any signal $y^2 \notin Y^2$ the result has been shown. To show the converse result we must show that for any pair of partitions $(\mathcal{P}_1, \mathcal{P}_2)$ such that neither partition is finer than the other we can find a decision problem in which $\mathcal{P}_1$ is preferred to $\mathcal{P}_2$ and vice versa. FIGURE. ■

- Prior information can be identified with an uninformative information structure $\mathcal{P}_0$

$$U\left(\mathcal{P}_0, \pi, u\right) = \max_{a \in A} \int_{\Omega} u\left(a, \omega\right) \pi\left(\omega\right) d\pi.$$

- Every information structure is finer than $\mathcal{P}_0$. From Theorem 1 we conclude that an information structure without noise is always valuable to an agent and we can define the value of information structure by

$$V(\mathcal{P}_1, \pi, u) = U(\mathcal{P}_1, \pi, u) - U(\mathcal{P}_0, \pi, u) \geq 0.$$

1.2 Information Structure with Noise

- Consists of a space of signals and a function from Ω to the state of probability measures over Y

- In other words it's given by a conditional probability function $f(y|\omega)$ over Y.

- For example, if y is normally distributed then $f(y|\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\omega)^2}{2\sigma^2}\right)$

- Given $f(y|\omega)$ beliefs are revised according to Bayes, as follows:

$$v(\omega|y) = \frac{f(y|\omega)\,\pi(\omega)}{\int_\Omega f(y|\tilde{\omega})\,\pi(\tilde{\omega})\,d\tilde{\omega}}.$$

- The important point is that the decision maker knows $f(.|.)$, namely he knows the probability the information system will make a mistake.

- Without information the decision maker solves

$$\max_{a\in A} \int_\Omega u(a,\omega)\,\pi(\omega)\,d\pi$$

- which yields the optimal decision a^{*0}

- With information structure characterized $f(y|\omega)$, for any value of y he solves

$$\max_{a\in A}\int_{\Omega} u(a,\omega)\, v(\omega|y)\, d\omega$$

- which yields decision $a^*(y)$. By definition of $a^*(y)$ we have

$$\int_{\Omega} u(a^*(y),\omega)\, v(\omega|y)\, d\omega \geq \int_{\Omega} u\left(a^{*0},\omega\right) v(\omega|y)\, d\omega$$

- therefore

$$\begin{aligned} & \int_Y \int_\Omega u\left(a^*\left(y\right), \omega\right) v\left(\omega|y\right) d\omega \pi\left(y\right) dy \\ \geq & \int_Y \int_\Omega u\left(a^{*0}, \omega\right) v\left(\omega|y\right) d\omega \pi\left(y\right) dy \\ = & \int_Y \int_\Omega u\left(a^{*0}, \omega\right) f\left(y|\omega\right) dy \pi\left(\omega\right) d\omega \text{ (By Bayes)} \\ = & \int_\Omega u\left(a^{*0}, \omega\right) \pi\left(\omega\right) d\omega. \end{aligned}$$

- Comparing information structures with Noise is more delicate and is the subject of BlacKwell's Theorem.

- Let

$$U\left[Y, f, \pi, u\right] = \sum_Y \pi\left(y\right) \int u\left(a^*\left(y\right), \omega\right) v\left(\omega|y\right) d\omega.$$

Definition 3 *Information structure* $\left[Y^1, f^1\right]$ *is more valuable than* $\left[Y^2, f^2\right]$ *iff*

$$U\left[Y^1, f^1, \pi, u\right] \geq U\left[Y^2, f^2, \pi, u\right], \forall u, \pi.$$

Theorem 2 *Blackwell (1951,1953). Information structure* $\left[Y^1, f^1\right]$ *is more valuable than* $\left[Y^2, f^2\right]$ *iff* $\left[Y^1, f^1\right]$ *is* sufficient *for* $\left[Y^2, f^2\right]$*,namely iff there exists non negative numbers* $\beta_{y_k^1, y_{k'}^2}$ *such that*

1. $f^2\left(y_{k'}^2|\omega\right) = \sum_{y_k^1 \in Y_1} \beta_{y_k^1, y_{k'}^2} f\left(y_k^1|\omega\right)$ *for all* ω *and* $y_{k'}^2$

2. $\sum_{y_{k'}^2 \in Y^2} \beta_{y_k^1, y_{k'}^2} = 1$ *for all* $y_k^1 \in Y^1$.

Proof. See Crémer (1982). ■

This (1) is a generalization of the following idea. Each time y_1 is observed, it is garbled by a stochastic mechanism independent of ω and transformed into a vector of signals in Y^2 via the conditional distribution $p(y_2|y_1)$.

Example 1 *When* Y^1, Y^2, Ω *have a finite number of elements say 2 elements then*

$$F^1 = \begin{pmatrix} f^1\left(y_1^1|\omega_1\right) & f^1\left(y_1^1|\omega_2\right) \\ f^1\left(y_2^1|\omega_1\right) & f^1\left(y_2^1|\omega_2\right) \end{pmatrix}$$

$$F^2 = \begin{pmatrix} f^2\left(y_1^2|\omega_1\right) & f^2\left(y_1^2|\omega_2\right) \\ f^2\left(y_2^2|\omega_1\right) & f^2\left(y_2^2|\omega_2\right) \end{pmatrix}$$

$$B = \begin{pmatrix} \beta_{y_1^2 y_1^1} & \beta_{y_1^2 y_2^1} \\ \beta_{y_2^2 y_1^1} & \beta_{y_2^2 y_2^1} \end{pmatrix}$$

- B is a Markov probability matrix (if we fix a column and sum across rows we get 1). Condition 1 can be written as

$$F^2 = BF^1.$$

Example 2 *Consider a perfect expert in the example when a good can be either good or bad.*

$$F^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Now consider a Markov Matrix of the form

$$B = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}.$$

Then

$$F^2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}$$

So expert 2 is no longer perfect, he errs three out of four times when quality is high and one out of two times when is bad.

1.3 Demski 1973 and the Impossibility of General Accounting Standards

- Demski (1973) asks whether there is an optimal accounting standard, defined as a complete ranking of information systems based on (Blackwell) informativeness. As we know from Blackwell (1951), the answer to Demski's question is no.

- For example, we cannot say in general whether a system that provides precise good news is better than another system that provides precise bad news: it depends on the decision problem, or the decision maker preferences.

- That does not mean there we cannot compare information systems. More informative systems, in the sense of Blackwell, are always better. In particular, in single agent decision settings more information is always better.

1.4 Continuously Distributed Signals

- For continuously distributed signals, we say that a signal s is sufficient for signal s' if there is a stochastic transformation $g : S \times S' \to \mathbb{R}_+$, where $\int_{Y'} g(y', y)\, dy' = 1$ for any $y \in Y$. Assuming that g is integrable

$$f'\left(y'|\omega\right) = \int_Y g\left(y', y\right) f(y|\omega)\, dy.$$

Example 3 *Suppose that* ω˜$U[-M, M]$ *and consider the family of signals indexed by* ε, *where* $s = \omega+$*Uniform*$[-\varepsilon, \varepsilon]$. *Show that these signals cannot be ranked based on Blackwell criterion.*

- When are normally distributed signals Blackwell more informative?

Assume $\omega \sim N\left(\mu, \sigma_\omega^2\right)$ and $y|\omega \sim (\omega, \sigma_\varepsilon^2)$, that is

$$y = \omega + \varepsilon$$

where $\varepsilon \sim N\left(0, \sigma_\varepsilon^2\right)$ and cov$(\varepsilon, \omega) = 0$.

- In general, y is more informative than $y' = \omega + \varepsilon'$ if and only if ε' has the same distribution as $\varepsilon + \xi$ where ε and ξ are independent. (So if ε' is normal, then both ε and ξ must be normal for y and y' to be Blackwell comparable (see Lehmann (1988)). This implies than only a normally distributed signal can be Blackwell more informative than other normally distributed signal.)

- It's clear that y is more informative than y' iff $\tau_\varepsilon \geq \tau_{\varepsilon'}$ because y' can be obtained from y by adding noise.

1.4.1 Sufficient Statistic (Digression)

The notion of Blacwkell sufficiency should not be mistaken with that of a sufficient statistic. Roughly, given a set X of independent identically distributed data conditioned on an unknown parameter $\tilde{\theta}$, a sufficient statistic is a function $T(X)$ whose value contains all the information needed to compute any estimate of the parameter (e.g. a maximum likelihood estimate). Due to the factorization theorem, for a sufficient statistic $T(X)$, the joint distribution can be written as

$$p(X) = h(X)g(\theta, T(X)),$$

From this factorization, it can easily be seen that the maximum likelihood estimate of θ will interact with X only through $T(X)$. Typically, the sufficient statistic is a simple function of the data, e.g. the sum of all the data points.

2 Integral Precision

Blackwell does not tell us how to identify empirically the informativeness of an information system. In fact, checking Blackwell conditions is notoriously difficult. Ganuza and Penalva (2010) provide a measure of informativeness based on the dispersion of conditional expectations called integral precision.

- Let V be a random variable representing the state of nature, i.e., the firm value. Let Y_k be a signal (a random variable).

- For a given prior $H(v)$ we compare a signal Y_1 with another Y_2 in terms of information content. We say that Y_1 is more precise than Y_2 if $E[V|Y_1]$ is more disperse than $E[V|Y_2]$. We use the notion of dispersion that the satistics literature refers to as the convex order.

Definition 4 *X is greater than Z in the convex order if for all convex real valued functions ϕ, $E[\phi(X)] \geq E[\phi(Z)]$ provided the expectation exists.*

- If both X and Z have the same finite mean, then $X \geq_{cx} Z$ if and only if X is a mean preserving increase in risk of Z.

Definition 5 *Y_1 is more integral precise than Y_2 if $E[V|Y_1]$ is greater in the convex order than $E[V|Y_2]$.*

- Notice that signals are ordered for a given prior. The prior plays a crucial role in the definition.

- Integral precision is a partial order, and it's consistent with Blackwell in the sense that if two signals are ordered according to Blackwell, they will be equally ordered according to integral precision. But there are signals that can be ranked according to integral precision but can't be ranked according to Blackwell.

The notion of integral precision implies that if we think of stock prices as expectations conditioned on accounting information, then the informativeness of accounting information could be measured by estimating the dispersion of stock prices during earnings announcements. The problem is that we do not know the priors. HOMEWORK.

3 Entropy

- Entropy is a measure of the average uncertainty in a random variable. It was developed by Shannon (1948).

- Mathematically, it is defined as

$$H(\tilde{x}) = E[I(\tilde{x})] = E[-\ln P(\tilde{x})]$$

- We can think of entropy as the average surprise of information.

Example 4 *For example, if $\tilde{x}$ is a binomial r.v then*

$$H(\tilde{x}) = -p_0 \ln p_0 - (1 - p_0) \ln (1 - p_0).$$

Entropy is maximized when $p_0 = \frac{1}{2}$.

PLOT

- Information, according to entropy, must satisfy the following conditions:

  1. $I(p) \geq 0$

  2. $I(1) = 0$

  3. $I(p_1 p_2) = I(p_1) + I(p_2)$ if the two events are independent.

- The information of signal Y can be represented as

$$E_{\tilde{y}}\left[H(\tilde{x}) - H(\tilde{x}|\tilde{y})\right].$$

4 Unraveling

The following notes are based on Milgrom (1981). The unraveling principle was independently stated by Grossman (1981) and Milgrom (1981).

- Let Θ be the possible values of a random parameter $\tilde{\theta}$
- The set of possible signals about $\tilde{\theta}$ is X
- Let $f(x|\theta)$ denote the conditional density on X when $\tilde{\theta}$ takes the particular value θ.

Definition 6 *A signal x is more favorable than another signal y if for every non degenerate prior G for θ, the posterior distribution $G(.|x)$ dominates the posterior distribution $G(.|y)$ in the sense of strict FOSD.*

Proposition 1 *x is more favorable than y if for every $\theta^* > \theta$,*

$$\frac{f(x|\theta^*)}{f(x|\theta)} > \frac{f(y|\theta^*)}{f(y|\theta)}$$

Proof. For sufficiency, consider $\theta^* > \theta$, then for all $\tilde{\theta} > \theta^*$ we have that

$$\frac{f\left(x|\tilde{\theta}\right)}{f(x|\theta)} > \frac{f\left(y|\tilde{\theta}\right)}{f(y|\theta)}.$$

Integrating for all $\tilde{\theta} > \theta^*$

$$\frac{\int_{\tilde{\theta}>\theta^*} f\left(x|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)}{f\left(x|\theta\right)} > \frac{\int_{\tilde{\theta}>\theta^*} f\left(y|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)}{f\left(y|\theta\right)}$$

or equivalently

$$\frac{f\left(x|\theta\right)}{\int_{\tilde{\theta}>\theta^*} f\left(x|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)} < \frac{f\left(y|\theta\right)}{\int_{\tilde{\theta}>\theta^*} f\left(y|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)}.$$

Integrating over $\theta \leq \theta^*$

$$\frac{\int_{\theta<\theta^*} f\left(x|\theta\right) dG\left(\theta\right)}{\int_{\tilde{\theta}>\theta^*} f\left(x|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)} < \frac{\int_{\theta<\theta^*} f\left(y|\theta\right) dG\left(\theta\right)}{\int_{\tilde{\theta}>\theta^*} f\left(y|\tilde{\theta}\right) dG\left(\tilde{\theta}\right)}.$$

By Bayes' rule we have $g(\theta|x) = \frac{f(x|\theta)g(\theta)}{f(x)}$ hence the above inequality boils down to

$$\frac{G(\theta^*|x)}{1 - G(\theta^*|x)} < \frac{G(\theta^*|y)}{1 - G(\theta^*|y)}.$$

which implies

$$G(\theta^*|x) < G(\theta^*|y).$$

■

- MLRP takes its name from the fact that the likelihood ratio $\frac{f(x|\theta^*)}{f(x|\theta)}$ is monotonic in x.

Proposition 2 *Let $\tilde{x}$ be a random variable whose densities have the strict MLRP. For any two intervals $[a, b]$ and $[c, d]$ with $a \geq c$ and $b \geq d$, where* *at least one inequality is strict, the signal* $\{\tilde{x} \in [a, b]\}$ *is more favorable than* $\{\tilde{x} \in [c, d]\}$.

4.1 A Persuasion Game

- A commodity of unknown value $\tilde{\theta}$ is to be exchanged for money.

- If the buyer purchases q units at price p his payoff is $\tilde{\theta}F(q) - pq$ where $F(\cdot)$ is increasing, concave, and differentiable.

- CAPITAL MARKET INTERPRETATION. $\max_p -|\left(p - E(\tilde{\theta}|r(\tilde{x}) = S)\right)|$

- The seller payoff increases in q

- The seller has information about $\tilde{\theta}$ represented by $\tilde{x}$.

- A report S is a closed subset of $\mathbb{R}$

- The report S is an assertion by the seller that $\tilde{x} \in S$.

- A reporting strategy is a function $r(x)$ from $\mathbb{R}$ to the closed non empty subsets of $\mathbb{R}$ with the property that $x \in r(x)$

- In other words the report must be truthful.

- IN WHICH ACCOUNTING CONTEXTS IS THIS A GOOD ASSUMPTION?

- It can be very precise as when $r(x) = \{x\}$ or extremely vague as when $r(x) = \mathbb{R}$

- A purchasing strategy is denoted $b(S)$

- A pair (b, r) is a Nash equilibrium if holding r fixed, b is optimal for the buyer and holding b fixed, r is optimal for the seller.

- Some Nash equilibrium are unnatural (for example No disclosure)

- We focus on sequential equilibrium (see Kreps and Wilson (1982)).

- Let $c(S)$ be the conjecture of the market when the seller reports S. The buyer concludes that $\tilde{x} \in c(S)$

Definition 7 *A sequential equilibrium is a triple (b, r, c) satisfying*

1. *For every S, $b(S)$ solves* $\max_q E\left[\tilde{\theta} F(q) - pq | \tilde{x} \in c(S)\right]$

2. *For every x, $r(x)$ solves* $\max_S b(S)$ *subject to $x \in S$*

3. *For every S in the range of r, $c(S) = r^{-1}(S)$.*

EXPLAIN.

- A strategy is called of *full disclosure* if r together with any optimal response b satisfies $b(r(x)) = b(\{x\})$. Consequently r is a full disclosure strategy if $E[\theta|r(\tilde{x}) = r(x)] = E[\theta|\tilde{x} = x]$

Proposition 3 *At every sequential equilibrium, the seller uses a strategy of full disclosure.*

Proof. Let (b, r, c) be an equilibrium. From condition (ii) if follows that $b(\{x\}) \leq b(r(x))$. Since $c(\{x\}) = x$, the inequality holds only if $E\left[\tilde{\theta}|\tilde{x} = x\right] \leq E\left[\tilde{\theta}|r(\tilde{x}) = r(x)\right]$. Since x is arbitrary, the inequality can be written as $E\left[\tilde{\theta}|\tilde{x}\right] \leq E\left[\tilde{\theta}|r(\tilde{x})\right]$. If the inequality was ever strict, we would have that

$EE\left[\tilde{\theta}|\tilde{x}\right] < EE\left[\tilde{\theta}|r(\tilde{x})\right]$. But by the Law of Iterated Expectations $EE\left[\tilde{\theta}|\tilde{x}\right] = EE\left[\tilde{\theta}|r(\tilde{x})\right] = E\left(\tilde{\theta}\right)$, hence $E\left[\tilde{\theta}|\tilde{x} = x\right] = E\left[\tilde{\theta}|r(\tilde{x}) = r(x)\right]$ therefore $r(x)$ is a strategy of full disclosure. ■

- Very general. $\tilde{x} \in \mathbb{R}^n$.

- Any vagueness is intepreted by the buyer in the most pessimistic possible way by the buyer.

- Unrealistic. This does not seem to match our view of reality.

- We think of this theorem not as a real world prediction but as a conceptual framework, similar to Modigliani Miller.

- Strong assumptions:

  1. The seller always receives information

  2. The buyer knows the seller has information

  3. The seller cannot lie about his information

  4. Common knowledge about everything

  5. Disclosure is costless

  6. Common priors about $\left(\tilde{\theta}, \tilde{x}\right)$

References

Blackwell, David, "Comparison of experiments," in "Proceedings of the second Berkeley symposium on mathematical statistics and probability," Vol. 1 1951, pp. 93–102.

Crémer, Jacques, "A simple proof of Blackwell's Ṣcomparison of experimentsŤ theorem," *Journal of Economic Theory*, 1982, *27* (2), 439–443.

Demski, Joel S, "The general impossibility of normative accounting standards," *Accounting Review*, 1973, pp. 718–723.

Ganuza, Juan-José and Jose S Penalva, "Signal orderings based on dispersion and the supply of private information in auctions," *Econometrica*, 2010, *78* (3), 1007–1030.

Grossman, Sanford J, "The informational role of warranties and private disclosure about product quality," *Journal of law and economics*, 1981, pp. 461–483.

Kreps, David M and Robert Wilson, "Sequential equilibria," *Econometrica: Journal of the Econometric Society*, 1982, pp. 863–894.

Lehmann, E. L., "Comparing Location Experiments," *The Annals of Statistics*, 1988, *16* (2), pp. 521–533.

Milgrom, Paul R., "Good News and Bad News: Representation Theorems and Applications," *The Bell Journal of Economics*, 1981, *12* (2), pp. 380–391.

Shannon, Claude E, "The mathematics theory of communication," *Bell Syst. Tech. J*, 1948, *27*, 379–423.