Nash Equilibrium and information transmission coding and decoding rules*

Penélope Hernández[†], Amparo Urbano and José E.Vila[‡]

June 29, 2010

Abstract

The design of equilibrium protocols in sender-receiver games where communication is noisy occupies an important place in the Economic literature. This paper shows that the common way of constructing a noisy channel communication protocol in Information Theory does not necessarily lead to a Nash equilibrium. Given the decoding scheme, it may happen that, given some state, it is better for the sender to transmit a message that is different from that prescribed by the codebook. Similarly, when the sender uses the codebook as prescribed, the receiver may sometimes prefer to deviate from the decoding scheme when receiving a message.

Keywords: Noisy channel, Shannon's Theorem, sender-receiver games, Nash equilibrium.

JEL: C72, C02

---

*The authors wish to thank participants of the Third World Congress of the Game Theory Society, Chicago 2008, and the Center of Rationality and Interactive Decision Theory, Jerusalem 2009. Special thanks go to Bernhard von Stengel. The authors thank both the Spanish Ministry of Science and Technology and the European Feder Funds for financial support under project SEJ2007-66581.

†Corresponding Author: Departament of Economic Analysis and ERI-CES. University of Valencia. Campus dels Tarongers. 46022 Valencia (Spain). Tel: +34 96 3828783. Fax: +34 96 3828249. e-mail: penelope.hernandez@uv.es.

‡Departament of Economic Analysis and ERI-CES. University of Valencia. Campus dels Tarongers. 46022 Valencia (Spain). e-mails: amparo.urbano@uv.es; jose.e.vila@uv.es.

1 Introduction

A central result of Information Theory is Shannon's noisy channel coding theorem. The purpose of this note is to point out that this theorem is not robust to a game theoretical analysis and thus cannot be directly applied to strategic situations. To demonstrate our inquiry we study the same framework as Shannon: the possibility of a noisy channel communication between a privately informed sender and a receiver who must take an action. Our contribution is to show that the methodology developed for optimal information transmission do not necessarily define equilibria of sender-receiver games.

The issue of information transmission is not new in Economics and actually there is a vast literature starting with the seminal work of Crawford and Sobel [3]. Several papers have additionally addressed the situation where communication may be distorted in the communication process by assuming that messages may not arrive (Myerson [12], Rubinstein [13], among others). This brand of literature points out that players' strategic behavior under "almost common knowledge" is not enough to guarantee coordination. Less research has been undertaken when the noisy communication is of a particular type: while messages are always received by the receiver, they may differ from those sent by the sender (Blume *et al.* [1], Koessler [10], Hernández *et al.* [9], Mitusch and Strausz [11]). Another brand of the literature deals with entropy based communication protocols (See Gossner et al [4], Gossner and Tomala [5], [6], [7], Hernández and Urbano [8]).

Traditional Information Theory, pioneered by Shannon [14], has approached noisy information transmission by considering that agents communicate through a *discrete noisy channel.* Although Shannon does not describe this situation as a game, we consider it as in a standard sender-receiver game with two players: a sender and a receiver. The sender has to communicate through a noisy channel some private information from a message source to the receiver, who must take some action from an action space, and with both receiving 1 if the information is correctly transmitted and 0 otherwise. More precisely, suppose the sender wishes to transmit an input sequence of signals (a message) through a channel that makes errors. One way to compensate for these errors is to send through the channel not the sequence itself but a modified version of the sequence that contains redundant information. The process of modification chosen is called the *encoding* of the message. The receiver receives an output message and he has to decode it, removing the errors and obtaining the original message. He does this by applying a *decoding* function.

The situation that we consider, in line with the set up of Information Theory, is as follows. We have a set Ω of M states. The sender wants to transmit through the channel the chosen state, so there are M possible messages. The communication protocol is chosen, given by a codebook of M possible messages, each of which is represented by a codeword of length n over the communication alphabet. The sender picks the codeword corresponding to the state. This codeword is transmitted and altered by the noisy channel. The receiver decodes the received message (a string of n symbols from the alphabet), according to some decoding scheme. The protocol is common knowledge to the players. Both sender and receiver are supposed to follow the rules of the protocol.

The natural question from the viewpoint of Game Theory is whether following the rules constitutes a Nash equilibrium. The protocol may not define the best possible code in terms of reliability, but in that case one may hope that it constitutes at least a not-so-good Nash equilibrium.

This paper shows that the common way of constructing a communication protocol does not necessarily lead to a Nash equilibrium: Given the decoding scheme, it may happen that, given some state, it is better for the sender to transmit a message that is different from that prescribed by the codebook. Similarly, when the sender uses the codebook as prescribed, the receiver may sometimes prefer to deviate from the decoding scheme when receiving a message.

This common way of choosing a communication protocol is as follows:

1. The channel with its errors is defined as a discrete Markov process where a symbol from the alphabet is transformed into some other symbol according to some error probability.

2. From these characteristics of the channel one can compute a *capacity* of the channel, which determines the maximal *rate* of transmitting information reliably. For example, a rate of 0.2 means that (if the alphabet is binary) for every bit of information on the input side, one needs to transmit 5 bits across the channel.

3. A main insight of Shannon is that as long as the rate is below channel capacity, the probability of error in information transmission can be made arbitrarily small when the length n of the codewords is allowed to be sufficiently long.

The way Shannon achieves the above is the following: The sender selects M codewords of length n at random. That is, for every input message, the encoding is chosen entirely randomly from the set of all possible encoding functions. Furthermore, for every message, this choice is independent of the encoding of every other message. With high probability this random choice leads to a "nearly optimal" encoding function, from the point of view of rate and reliability. The decoding rule is based on a simple idea: A channel outcome will be decoded as a specific input message if that input sequence is "statistically close" to the output sequence. This statistical proximity is measured in terms of the entropy of the joint distribution of both sequences which establishes when two sequences are probabilistically related. The associated decoding function is known as the *jointly typical decoding.*

Our methodological note is organized as follows. The sender-receiver game and the noisy channel are set up in Section 2. Section 3 offers a rigorous presentation of Shannon's communication protocol, specifying players' strategies from a theoretical viewpoint. The reader familiar with Information Theory can skip it. Section 4 presents three simple examples of a sender-receiver game with specific code realizations. The first two examples offer the following code realizations: 1) the "natural one" where the decoding rule translates to the majority rule and where the equilibrium conditions are satisfied; and 2) a worse code realization, where a deviation by the receiver takes place. The last example exhibits a sender's deviation. Concluding remarks close the paper.

2 The basic sender-receiver set up

Consider the possibilities of communication between two players, called the sender (S) and the receiver (R) in an incomplete information game Γ: there is a finite set of feasible states of nature $\Omega = \{\omega_0, \ldots, \omega_{M-1}\}$. Nature chooses first randomly $\omega_j \in \Omega$ with probability q_j and then the sender is informed of this state ω_j, the receiver must take some action in some finite action space A, and payoffs are realized. The agents' payoffs depend on the sender's information or type ω and the receiver's action a. Let $u : A \times \Omega \to \mathbb{R}$ be the players' (common) payoff function, i.e., $u(a_t, \omega_j)$, $j = 0, 1, \ldots, M-1$. Assume that for each realization of ω, there exists a unique receiver's action with positive payoffs: for each state $\omega_j \in \Omega$, there exists a unique action $\widehat{a}_j \in A$ such that:

$$u(a_t, \omega_j) = \begin{cases} 1 & \text{if } a_t = \widehat{a}_j \\ 0 & \text{otherwise} \end{cases}$$

The timing of the game is as follows: the sender observes the value of ω and then sends a message, which is a string of signals from some message space. It is assumed that signals belong to some finite space and may be distorted in the communication process. This distortion or interference is known as *noise*. The noise can be modeled by assuming that the signals of each message can randomly be mapped to the whole set of possible signals. An unifying approach to this noisy information transmission is to consider that agents communicate through a *discrete noisy channel.*

Definition 1 *A discrete channel $(X; p(y|x); Y)$ is a system consisting of an input alphabet X and output alphabet Y, and a probability transition matrix $p(y|x)$ that expresses the probability of observing the output symbol y, given that the symbol x was sent.*

A channel is memoryless if the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs. In addition, a channel is used without feedback if the input symbols do not depend on the past output symbols.

The nth extension of a discrete memoryless channel is the channel $(\mathbf{X} = X^n; p(\mathbf{y} = \mathbf{y}^n | \mathbf{x} = x^n); \mathbf{Y} = Y^n)$, where $p(\mathbf{y}|\mathbf{x}) = p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$.

Consider the binary channel $\nu(\varepsilon_0, \varepsilon_1) = (X = \{0,1\}; p(y|x); Y = \{0,1\})$ where $p(1|0) = \varepsilon_0$ and $p(0|1) = \varepsilon_1$ (i.e., ε_l is the probability of a mistransmission of input message l) and let $\nu^n(\varepsilon_0, \varepsilon_1)$ be its nth extension. While binary channels may seem rather oversimplified, they capture the essence of most mathematical challenges that arise when trying to make communication reliable. Furthermore, many of the solutions found to make communication reliable in this setting have been generalized to other scenarios.

Let Γ_v^n denote the extended communication game. It is a one-stage game where the sender sends a message $\mathbf{x} \in \mathbf{X}$ of length n, using the noisy channel, the receiver observes a realization $\mathbf{y} \in \mathbf{Y}$ of such a message and takes an action in Γ.

A strategy of S in the extended communication game Γ_v^n is a decision rule suggesting the message to be sent at each ω_j: a M-tuple $\{\sigma_j^S\}_j$ where $\sigma_j^S \in \mathbf{X}$ is the message sent by S given that the true state of nature is ω_j. A strategy of R is a 2^n-tuple $\{\sigma_{\mathbf{y}}^R\}_{\mathbf{y}}$, specifying an action choice in Γ as a response to the realized output sequence $\mathbf{y} \in \mathbf{Y}$.

Expected payoffs are defined in the usual way. Let the tuple of the sender's payoffs be denoted by $\{\pi_j^S\}_j = \{\pi_j^S(\sigma_j^S, \{\sigma_{\mathbf{y}}^R\}_{\mathbf{y}})\}_j$, where for each ω_j,

$$\pi_j^S = \pi_j^S(\sigma_j^S, \{\sigma_{\mathbf{y}}^R\}_{\mathbf{y}}) = \sum_{\mathbf{y} \in \mathbf{Y}} p(\mathbf{y}|\sigma_j^S) u(\sigma_{\mathbf{y}}^R, \omega_j)$$

and where $p(\mathbf{y}|\sigma_j^S)$ is the sender's probability about the realization of the output sequence $\mathbf{y} \in \mathbf{Y}$ conditional on having sent message σ_j^S in state ω_j.

Let the tuple of the receiver's payoffs be denoted by $\{\pi_{\mathbf{y}}^R\}_{\mathbf{y}} = \{\pi_{\mathbf{y}}^R(\{\sigma_j^S\}_j, \sigma_{\mathbf{y}}^R)\}_{\mathbf{y}}$, where for each output sequence $\mathbf{y} \in \mathbf{Y}$,

$$\pi_{\mathbf{y}}^R = \pi_{\mathbf{y}}^R(\{\sigma_j^S\}_j, \sigma_{\mathbf{y}}^R) = \sum_{j=0}^{M-1} p(\sigma_j^S|\mathbf{y}) u(\sigma_{\mathbf{y}}^R, \omega_j)$$

and where $p(\sigma_j^S|\mathbf{y})$ is the receiver's probability about input message σ_j^S in state ω_j conditional on having received the output message $\mathbf{y}$.

A pure strategy Nash equilibrium of the communication game is a pair of tuples $(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_{\mathbf{y}}^R\}_{\mathbf{y}})$ such that for each ω_j, and for any other strategy $\widetilde{\sigma}_j^S$ of the sender,

$$\widehat{\pi}_j^S = \pi_j^S(\widehat{\sigma}_j^S, \{\widehat{\sigma}_{\mathbf{y}}^R\}_{\mathbf{y}}) \geq \pi_j^S(\widetilde{\sigma}_j^S, \{\widehat{\sigma}_{\mathbf{y}}^R\}_{\mathbf{y}})$$

and for each $\mathbf{y} \in \mathbf{Y}$ and for any other receiver's strategy $\widetilde{\sigma}_{\mathbf{y}}^R$,

$$\widehat{\pi}_{\mathbf{y}}^R = \pi_{\mathbf{y}}^R(\{\widehat{\sigma}_j^S\}_j, \widehat{\sigma}_{\mathbf{y}}^R) \geq \pi_{\mathbf{y}}^R(\{\widehat{\sigma}_j^S\}_j, \widetilde{\sigma}_{\mathbf{y}}^R)$$

Notice that the set of probabilities $\{p(\sigma_j^S|\mathbf{y})\}_j$ for the receiver (where by Bayes rule $p(\sigma_j^S|\mathbf{y}) = \frac{p(\mathbf{y}|\sigma_j^S)p(\sigma_j^S)}{p(\mathbf{y})}$) is always well-defined ($p(\mathbf{y}) > 0$ for all $\mathbf{y}$). Therefore, the Nash equilibrium is also a perfect Bayesian equilibrium.

Fix the Sender's strategy $\{\sigma_j^S\}_{0,\ldots,M-1}$ where $\sigma_j^S \in X$ is the message sent by S given that the true state of nature is ω_j. The receiver has to take an action a_l in Γ after receiving an output sequence $\mathbf{y}$ such that:

$$a_l = Arg\max_{a_l} \sum_{j=0}^{M-1} p(\sigma_j^S|\mathbf{y})u(\sigma_{\mathbf{y}}^R, \omega_j) = Arg\max_{a_l} \sum_{j=0}^{M-1} p(\sigma_j^S|\mathbf{y}).$$

Equivalently, given the linearity of the receiver's payoff functions in probabilities $\{p(\sigma_l^S|\mathbf{y})\}_l$, $0 \leq l < M-1$, and since by Bayes' rule,

$$\frac{p(\sigma_l^S|\mathbf{y})}{p(\sigma_k^S|\mathbf{y})} = \frac{\frac{p(\mathbf{y}|\sigma_l^S)p(\sigma_l^S)}{p(\mathbf{y})}}{\frac{p(\mathbf{y}|\sigma_k^S)p(\sigma_k^S)}{p(\mathbf{y})}} = \frac{q_l}{q_k}\frac{p(\mathbf{y}|\sigma_l^S)}{p(\mathbf{y}|\sigma_k^S)}$$

then the receiver will choose, for each $\mathbf{y}$, action a_l whenever $q_l p(\sigma_l^S|\mathbf{y}) \geq q_k p(\sigma_k^S|\mathbf{y})$ (i.e., $\frac{q_l p(\sigma_l^S|\mathbf{y})}{q_k p(\sigma_k^S|\mathbf{y})} \geq 1$), for all $k \neq l$, $k = 0, \ldots, M-1$, and will choose a_k otherwise. This condition translates to the receiver choosing action a_l whenever $q_l p(\mathbf{y}|\sigma_l^S) \geq q_k p(\mathbf{y}|\sigma_k^S)$, and choosing a_k otherwise, with $p(\mathbf{y}|\sigma_j^S)$ given by the channel's error probabilities and by the sender's coding. To simplify assume that the states of nature are uniformly distributed, $q_l = \frac{1}{M}$ for $l \in \{0, \ldots, M-1\}$. Then

$$\sigma_{\mathbf{y}}^R = a_l, \text{ whenever } p(\mathbf{y}|\sigma_l^S) \geq p(\mathbf{y}|\sigma_k^S) \ \forall \sigma_k^S \in \mathbf{X} \tag{1}$$

Consider now the sender's best response to the receiver's strategy $\sigma_{\mathbf{y}}^R$. The sender's problem is to choose an input sequence σ_j^S for each state ω_j, $j = 0, \ldots, M-1$, such that

$$\sigma_l^S = Arg\max \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_l^S)u(\sigma_{\mathbf{y}}^R, \omega_l) = Arg\max \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_l^S).$$

Given the receiver's decoding, the above problem amounts to choosing an input sequence σ_j^S in states ω_j such that

$$\sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_l^S) \geq \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|x^S) \tag{2}$$

for any other input sequences x^S in all codebooks over $\{0,1\}^n$.

3 Shannon's communication protocol

For completeness we present first some basic results from Information Theory, largely following Cover and Thomas [2]

Let X be a random variable with probability distribution p. The entropy $H(X)$ of X is defined by $H(X) = -\Sigma_{\theta \in \Theta} p(\theta) \log(p(\theta) = -E_X [\log p(X)]$, where 0 *log* $0 = 0$ by convention. Consider independent, identically distributed (i.i.d.) random variables $X_1, \ldots, X_n$. Then by the definition of entropy,

$$H(X_1, \ldots, X_n) = -\Sigma_{\theta_1 \in \Theta_1} \ldots \Sigma_{\theta_n \in \Theta_n} p(\theta_1, \ldots, \theta_n) \log p(\theta_1, \ldots, \theta_n)$$

where $p(\theta_1, \ldots, \theta_n) = p(X_1 = \theta_1, \ldots, X_n = \theta_n)$.

Let $\mathbf{x}$ be a sequence of length n over a finite alphabet θ of size $|\theta|$. Denote by $\theta_i(\mathbf{x})$ the frequency θ_i over n. We define the empirical entropy of $\mathbf{x}$, denoted by $H(\theta_1(\mathbf{x}), \ldots, \theta_{|\theta|}(\mathbf{x}))$, as the entropy of the empirical distribution of $\mathbf{x}$.

An (M, n) code for the channel $(X, p(y \mid x), Y)$ consists of 1) an index set $\{0, 1, \ldots, M-1\}$; 2) an encoding function $e : \{0, 1, \ldots, M-1\} \longrightarrow X^n$, yielding *codewords*: $e(1), e(2), \ldots, e(M)$. The set of codewords is called the *codebook*; 3) a decoding function $d : Y^n \longrightarrow \{0, 1, \ldots, M-1\}$.

Consider a noisy channel and a communication length n. Let

$$\lambda_i = \Pr(d(Y^n \neq i | X^n = X^n(i)) = \sum_{y^n} p(y^n | x^n(i)) I(d(y^n) \neq i)$$

be the conditional probability of error given that index i was sent, and where $I(.)$ is the indicator function. The maximal probability of error $\lambda^{(n)}$ for an (M, n) code is defined as $\lambda^{(n)} = \max_{i \in \{0,1,\ldots,M-1\}} \lambda_i$ and the average probability of error $P_e^{(n)}$ for an (M, n) code is $P_e^{(n)} = \frac{1}{M} \sum_{i=0}^{M-1} \lambda_i$. Note that $P_e^{(n)} \leq \lambda^{(n)}$.

The *rate* and the *mutual information* are two useful concepts from Information Theory characterizing when information can be reliably transmitted over a communications channel. The rate r of an (M, n) code is equal to $r = \frac{\log_{|\Theta|} M}{n}$, and a rate r is said to be *achievable* if there exists a sequence of $(2^{nr}, n)$ codes such that the maximal probability of error $\lambda^{(n)}$ tends to 0 as n goes to ∞. The *capacity* of a discrete memoryless channel is the supremum of all achievable rates.

The mutual information $I(X; Y)$ measures the information that random variables X and Y share. Mutual information can be equivalently expressed as $I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, where $H(Y|X)$ is the conditional entropy of Y (taking values $\theta_2 \in \Theta_2$) given X (taking values $\theta_1 \in \Theta_1$) defined by

$$H(Y \mid X) = - \sum_{\theta_1 \in \Theta_1} p(\theta_1) \sum_{\theta_2 \in \Theta_2} p(\theta_2 \mid \theta_1) \log p(\theta_2 \mid \theta_1).$$

Then, the capacity C of a channel can be expressed as the maximum of the mutual information. Formally: $C = \sup_{p_X} I(X;Y)$ between the input and output of the channel, where the maximization is with respect to the input distribution. Therefore the channel capacity is the tightest upper bound on the amount of information that can be reliably transmitted over a communications channel.

Theorem 1 *(Shannon):All rates below capacity C are achievable. Specifically, for every rate $r < C$, there exists a sequence of $(2^{nr}, n)$ codes with maximum probability of error $\lambda^{(n)} \longrightarrow 0$. Conversely, any sequence of $(2^{nr}, n)$ codes with $\lambda^{(n)} \longrightarrow 0$ must have $r \leq C$.*

3.1 Shannon's strategies:

Fix a channel and a communication length n. We can compute from the channel its capacity C, and from n the information transmission rate r. Shannon's theorem states that given a noisy channel with capacity C and information transmission rate r, if $r < C$, then there will exist both an encoding rule and a decoding rule which will allow the receiver to make arbitrarily small the average probability of the information transmission error. These two parameters: rate and capacity are the key to the existence of such coding[1].

The sender's strategy: random coding Let us show how to construct a random choice of codewords to generate a (M, n) code for our sender-receiver game. Consider the binary channel $\nu(\varepsilon_0, \varepsilon_1)$ and its nth extension $\nu^n(\varepsilon_0, \varepsilon_1)$. Following Shannon's construction random codes are generated, for each state of nature, according to the probability distribution θ that maximizes the mutual information $I(X;Y)$. In other words, let us assume a binary random variable X_θ that takes value 0 with probability θ and value 1 with probability $1 - \theta$. Then, let Y_θ be the random variable defined by the probabilistic transformation of input variable X_θ through the channel, with probability distribution:

$$Y_\theta = \{(1 - \varepsilon_0)\theta + \varepsilon_1(1 - \theta), \varepsilon_0\theta + (1 - \varepsilon_1)(1 - \theta)\}.$$

Therefore the mutual information between X_θ and Y_θ is equal to:

$$\begin{aligned} I(X_\theta; Y_\theta) &= H(Y_\theta) - H(Y_\theta|X_\theta) = \\ &\quad H(\{(1 - \varepsilon_0)\theta + \varepsilon_1(1 - \theta), \varepsilon_0\theta + (1 - \varepsilon_1)(1 - \theta)\}) - [\theta H(\varepsilon_0) + (1 - \theta)H(\varepsilon_1)], \end{aligned}$$

where θ is obtained as the solution of the optimization problem:

$$\theta = \arg\max_{\overline{\theta}} I(X_{\overline{\theta}}, Y_{\overline{\theta}})$$

[1]Notice that for a fixed C, it is always possible to find a length n, large enough, to guarantee Shannon's Theorem. Alternatively, given a fixed r, we can always find a noisy structure, a channel, achieving this transmission rate.

Denoting by $p(x)$ the distribution of X_θ according to θ, generate 2^{nR} codewords, i.e., a (M,n) code at random according to $p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i)$.

The M codewords can be displayed as the rows of a matrix:

$$\zeta = \begin{bmatrix} x_1(0) & x_2(0) & \dots & x_n(0) \\ \dots & \dots & \dots & \dots \\ x_1(M-1) & x_2(M-1) & \dots & x_n(M-1) \end{bmatrix}$$

and therefore the probability of such a code is: $p(\zeta) = \prod_{\omega=0}^{2^{nR}-1} \prod_{i=1}^{n} p(x_i(\omega))$.

The receiver's strategy: jointly typical decoding The receiver's strategy is based on a statistical property derived from the weak law of large numbers. This property tell us when two sequences are probabilistically related.

Definition 2 *The set A_η^n of* jointly typical *sequences $\{\mathbf{x},\mathbf{y}\}$ with respect to the distribution $p(\mathbf{x},\mathbf{y})$ is the set of n-sequences with empirical entropy η-close to the true entropy,i.e.*

$$A_\eta^n = \begin{array}{c} \{(\mathbf{x},\mathbf{y}) \in \mathbf{X}\times\mathbf{Y} : \left|-\frac{1}{n}\log p(\mathbf{x}) - H(X)\right| < \eta; \left|-\frac{1}{n}\log p(\mathbf{y}) - H(Y)\right| < \eta \quad \textit{and} \\ \left|-\frac{1}{n}\log p(\mathbf{x},\mathbf{y}) - H(X,Y)\right| < \eta\} \end{array}$$

A channel outcome $\mathbf{y} \in \mathbf{Y}$ will be decoded as the *ith* index if the codeword $\mathbf{x}_i \in \mathbf{X}$ is "*jointly typical*" with the received sequence $\mathbf{y}$: two sequences $\mathbf{x}$ and $\mathbf{y}$ are jointly η-typical if the pair $(\mathbf{x},\mathbf{y})$ is η-typical with respect to the joint distribution $p(\mathbf{x},\mathbf{y})$ and both $\mathbf{x}$ and $\mathbf{y}$ are η-typical with respect to their marginal distributions $p(\mathbf{x})$ and $p(\mathbf{y})$. In words, a typical set with tolerance η, A_η^n, is the set of sequences whose empirical entropy differ by no more than η from their true entropy.

Shannon's communication protocol: Let us apply the above concepts to the extended communication game Γ_v^n. The sender communicates her private information, through the *nth* extension of the noisy channel $\nu(\varepsilon_0,\varepsilon_1)$, by generating M codewords of length n from the probability θ which maximizes the capacity of the channel. The communication protocol has the following sequence of events:

1. The realization of such codes is revealed to both the sender and the receiver.
2. The sender is informed about the true state of nature and sends message $\mathbf{x}_i$ associated to $i \in \Omega$.
3. The receiver observes a sequence $\mathbf{y}$, according to $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i)$

4. The receiver updates the possible state of nature, and decides that index $l \in \Omega$ was sent if the following conditions are satisfied:

   - $(\mathbf{x}_l, \mathbf{y})$ are jointly typical.
   - There is no other index $k \in \Omega$ such that $(\mathbf{x}_k, \mathbf{y})$ are jointly typical.
   - If no such $l \in \Omega$ exists, then an error will be declared.

5. Finally, the receiver chooses an action in Γ according to his decoding rule:

   - if $\mathbf{y}$ is only jointly typical with $\mathbf{x}_l$, he takes action a_l,
   - otherwise, no action is taken.

Shannon was the first one to show that good codes exists. Given the above strategies and Shannon's Theorem, we can construct a *good* code for information transmission purposes in the following way:

1. Choose first the θ that maximizes the mutual information $I(X;Y)$ and generate a realization of the random code. Then, for *all* η there exists an n^* such that for all $n \geq n^*$, the empirical entropy of each realized code is at distance $\frac{\eta}{12}$ to $H(X)$.

2. By the jointly typical decoding rule, any output message $\mathbf{y}$ is decoded as either a unique input coding $\mathbf{x}$, or an error is declared. When no error is declared, the decoding rule translates to the condition that the distance between the empirical entropy of the pair $(\mathbf{x}, \mathbf{y})$ and the true entropy $H(X, Y)$ is smaller than $\frac{\eta}{12}$.

3. By the proof of the above Shannon's Theorem (Cover and Thomas, page 200–202), the average probability of error $P_e^{(n)}$, averaged over all codebooks, is smaller than$\frac{\eta}{2}$. Therefore, for a fixed $n \in [n^*, \infty)$, there shall exist a realization of a codebook satisfying that at least half of its codewords have conditional probability of error less than η. In particular, its maximal probability of error $\lambda^{(n)}$ is less than η.

Notice that in order to apply this protocol to a standard sender-receiver game, one needs to define an assignment rule when an error is declared in Shannon's protocol. This rule assigns an action to the decoding errors and allows us to completely specify the receiver's strategy.

Remark:

Shannon's Theorem is an asymptotic result and establishes that for all $\eta-$ approximations there exists a large enough n guaranteeing a small average error related to such η. By the proof of the Theorem (Cover and Thomas, page 200-202), the average error has two terms. The first one comes from the Jointly Typical Set defined by such a threshold η. Here, again for large enough n, the probability that a realized output sequence is not jointly typical with the right code is very low. The second term comes from the declared errors in Shanon's protocol, which have a probability of $2^{\{-n(I(X:Y)-3\eta))\}}$ of taking place and which is very small when n is large enough.

Therefore, both probabilities are bigger or smaller depending on both n and how many outcomes are rightly declared, and they are important to partition the output sequence space.

When we focus on finite-time communication protocols, i.e., when n and η are both fixed, disregarding asymptotic assumptions, we cannot guarantee that the above probabilities are small enough with respect to n. Actually, the η-approximation and the corresponding different associated errors can generate different partitions of the output space. Therefore, careful attention shall be paid to generate a partition in such situations.

3.2 Nash Equilibrium Codes

We have defined good information transmission codes. They come from asymptotic behavior. Now, we look for finite communication-time codes and such that no player has an incentive to deviate.

Let Y_l be the set of $\mathbf{y}$'s in $\mathbf{Y}$ such that the receiver decodes all of them as index $l \in \{0, 1, \ldots, M-1\}$. From the equilibrium conditions 1 and 2 in section 2:

Proposition 1 *A code* (M, n) *is a Nash Equilibrium code if and only if*
i) $p(\mathbf{y}|\mathbf{x}(i)) \geq p(\mathbf{y}|\mathbf{x}(j))\ \forall i \neq j \in M$*, and* $d(y) = i$
ii) $\sum_{y \in Y_i} p(y|\mathbf{x}(i)) \geq \sum_{y \in Y_i} p(y|\mathbf{x})$*, for all* $\mathbf{x} \in \{0,1\}^n$.

The question that arises is whether Shannon's strategies are Nash equilibrium strategies of the extended communication game Γ_ν^n. Particularly, we rewrite condition *i)* above in terms of the entropy condition of the jointly typical sequences. For any two indexes l and k, let $\mathbf{x_l} = \mathbf{x}(l)$, and $\mathbf{x_k} = \mathbf{x}(k)$, then

$$d(\mathbf{y}) = l, \text{ whenever } p(\mathbf{y}|\mathbf{x_l}) \geq p(\mathbf{y}|\mathbf{x_k})\ \forall \mathbf{x_k} \in M$$

Alternatively, there exist $\eta > 0$ such that

$$-\frac{1}{n}\log p(\mathbf{x_l}, \mathbf{y}) - H(X,Y) < \eta \text{ and } -\frac{1}{n}\log p(\mathbf{x_k}, \mathbf{y}) - H(X,Y) > \eta.$$

By Definition 3, set A_η^n is the set of jointly typical sequences. Consider $\mathbf{y} \in \mathbf{Y}^n$ such that $(\mathbf{x}_0, \mathbf{y}) \in A_\eta^n$ and $(\mathbf{x}_1, \mathbf{y}) \notin A_\eta^n$. Formally:

$$\left|-\frac{1}{n}\log p(\mathbf{x_0}, \mathbf{y}) - H(X,Y)\right| < \eta \text{ and } \left|-\frac{1}{n}\log p(\mathbf{x_1}, \mathbf{y}) - H(X,Y)\right| \geq \eta$$

Therefore if $\mathbf{y}$ were decoded as l, we could assert that $\mathbf{y}$ is jointly typical with $\mathbf{x_l}$, and not jointly typical with any other $\mathbf{x_k}$. It is straightforward to check that the opposite is not true, that is, even if the empirical entropy of $p(\mathbf{x_l}, \mathbf{y})$ were closer than that of $p(\mathbf{x_k}, \mathbf{y})$ to the true entropy, then the conditional probability of $\mathbf{x}_l$ given

$\mathbf{y}$ would not need be bigger than the conditional probability of $\mathbf{x}_k$ given $\mathbf{y}$. In fact there are four possible inequalities:

1. $-\frac{1}{n}\log p(\mathbf{x_0},\mathbf{y}) - H(X,Y) < \eta$ and $-\frac{1}{n}\log p(\mathbf{x_1},\mathbf{y}) - H(X,Y) > \eta$. In this case we obtain that

$$p(\mathbf{x}_0|\mathbf{y}) > \frac{2^{-n(H(X,Y)+\eta)}}{p(\mathbf{y})} > p(\mathbf{x}_1|\mathbf{y})$$

and therefore, if $(\mathbf{x}_0,\mathbf{y})$ is more statistically related than $(\mathbf{x}_1,\mathbf{y})$, then the conditional probability of $\mathbf{x}_0$ given $\mathbf{y}$ will be greater than the conditional probability of $\mathbf{x}_1$ given $\mathbf{y}$.

2. $\frac{1}{n}\log p(\mathbf{x_0},\mathbf{y}) + H(X,Y) < \eta$ and $\frac{1}{n}\log p(\mathbf{x_1},\mathbf{y}) + H(X,Y) > \eta$. In this case we obtain the opposite conclusion. Namely,

$$p(\mathbf{x}_0|\mathbf{y}) < \frac{2^{-n(H(X,Y)-\eta)}}{p(\mathbf{y})} < p(\mathbf{x}_1|\mathbf{y})$$

and now the above condition shows that even if the empirical entropy of $p(\mathbf{x_0},\mathbf{y})$ were closer than that of $p(\mathbf{x_1},\mathbf{y})$ to the true entropy, then the conditional probability of $\mathbf{x}_1$ given $\mathbf{y}$ could be bigger than or equal to the conditional probability of $\mathbf{x}_0$ given $\mathbf{y}$.

3. $-\frac{1}{n}\log p(\mathbf{x_0},\mathbf{y}) - H(X,Y) < \eta$ and $\frac{1}{n}\log p(\mathbf{x_1},\mathbf{y}) + H(X,Y) > \eta$. Here,

$$p(\mathbf{x}_0,\mathbf{y}) > \frac{2^{-n(H(X,Y)+\eta)}}{p(\mathbf{y})} \quad \text{and} \quad \frac{2^{-n(H(X,Y)-\eta)}}{p(\mathbf{y})} < p(\mathbf{x}_1,\mathbf{y}).$$

and no relationship between $p(\mathbf{x}_0|\mathbf{y})$ and $p(\mathbf{x}_1|\mathbf{y})$ can be established. Finally,

4. $\frac{1}{n}\log p(\mathbf{x_0},\mathbf{y}) + H(X,Y) < \eta$ and $-\frac{1}{n}\log p(\mathbf{x_1},\mathbf{y}) - H(X,Y) > \eta$.

As the third case above, we cannot establish any order between $p(\mathbf{x}_0|\mathbf{y})$ and $p(\mathbf{x}_1|\mathbf{y})$. Indeed, we get:

$$p(\mathbf{x}_0|\mathbf{y}) < \frac{2^{-n(H(X,Y)+\eta)}}{p(\mathbf{y})} \quad \text{and} \quad \frac{2^{-n(H(X,Y)-\eta)}}{p(\mathbf{y})} > p(\mathbf{x}_1|\mathbf{y}).$$

Condition *i)* above establishes an *order* on the conditional probabilities of each output sequences $\mathbf{y}$, for all input sequences. We have seen that when the entropy condition of the Jointly Typical Set is satisfied without the absolute value, then it properly orders these conditional probabilities. Otherwise it may fail to do so.

Consider now condition *ii)*. Let Y_l be the set of $\mathbf{y} \in \mathbf{Y}$ such that $p(\mathbf{y}|\mathbf{x_l}) \geq p(\mathbf{y}|\mathbf{x_k})\ \forall \mathbf{x_k} \in M$. Summing over all $\mathbf{y}$ in Y_l we get:

$$\sum_{\mathbf{y}\in Y_l} p(\mathbf{y}|\mathbf{x_l}) \geq \sum_{\mathbf{y}\in Y_l} p(\mathbf{y}|\mathbf{x_k}) \text{ for all } \mathbf{x_k} \in M.$$

The second condition says that the aggregated probability of partition Y_l when σ_l^S was sent is higher than such probability[2] when any other code, even those sequences never taken into account in the realized codebook, are sent.

4 Examples: Shannon versus Game Theory

We wish to investigate whether the random coding and jointly typical decoding are robust to a game theoretical analysis, i.e. whether they are ex-ante equilibrium strategies. Since, the ex-ante equilibrium is equivalent to playing a Nash for every code realization, then if for some code realizations the players' strategies are not a Nash equilibrium, then no ex-ante equilibrium will exist.

In the sequel we analyze three examples. The first two examples correspond to two realizations of the random coding. The former consists of the "natural" coding in the sense that the signal strings do not share a common digit, either 0 or 1, and then the decoding rule translates to the "majority" rule; the latter is a worse codebook realization. For each code realization we show how to generate a partition of the output space, the receiver's strategy and the players' equilibrium conditions. In particular, we prove that receiver's equilibrium condition is not fulfilled for the second code realization. The last example offers a sender's deviation.

Fix a Sender-Receiver "common interest" game Γ where nature chooses ω_i, $i = 0, 1$, according to the law $q = (q_0, q_1) = (0.5, 0.5)$. The Receiver's set of actions is $A = \{a_0, a_1\}$ and the payoff matrices for both states of the nature are defined by:

		R	
		a_0	a_1
S	ω_0	$(1,1)$	$(0,0)$
	ω_1	$(0,0)$	$(1,1)$

Consider the noisy channel $\nu(\varepsilon_0,\varepsilon_1)$ where the probability transition matrix $p(y|x)$ expressing the probability of observing the output symbol y, given that the symbol x was sent, is $p(1|0) = \varepsilon_0 = 0.1$ and $p(0|1) = \varepsilon_1 = 0.2$.

Define the binary random variable X_θ which takes value 0 with probability θ and value 1 with probability $1 - \theta$. Let Y_θ be the random variable defined by the channel probabilistic transformation of the input random variable X_θ with probability distribution:

$$Y_\theta = \{(1 - \varepsilon_0)\theta + \varepsilon_1(1 - \theta), \varepsilon_0\theta + (1 - \varepsilon_1)(1 - \theta)\}.$$

[2]Recalling that the error λ_l of decoding the codeword $\mathbf{x_l}$ is $\lambda_l = Pr(\mathbf{y} \in \cup_{k\neq l} Y_k|\mathbf{x_l}) = \sum_{\mathbf{y}\notin Y_l} p(\mathbf{y}|\mathbf{x_l})$, and that the right side $\sum_{\mathbf{y}\in Y_l} p(\mathbf{y}|\mathbf{x_k})$ is part of the λ_k error, then the Sender's condition could be written as $1 - \lambda_l \geq \sum_{\mathbf{y}\in Y_l} p(\mathbf{y}|\mathbf{x_k})$ for all $\mathbf{x_k} \in M$, which means that the aggregated probability of the partition Y_l when σ_l^S was sent is higher than the corresponding part of the k-error of any code even for sequences never taken into account in the realized codebook.

Therefore the mutual information between X_θ and Y_θ is equal to:

$$\begin{aligned} I(X_\theta;Y_\theta) &= H(Y_\theta) - H(Y_\theta|X_\theta) = \\ & H(\{(1-\varepsilon_0)\theta + \varepsilon_1(1-\theta), \varepsilon_0\theta + (1-\varepsilon_1)(1-\theta)\}) - [\theta H(\varepsilon_0) + (1-\theta)H(\varepsilon_1)]. \end{aligned}$$

Let $\widehat{\theta} = \arg\max_\theta I(X_\theta, Y_\theta)$. Then for channel $\nu(\varepsilon_0,\varepsilon_1) = \nu(0.1,0.2)$, this probability $\widehat{\theta} = 0.52$.

Random codes are generated, for each state of nature, according to the probability distribution $\widehat{\theta} = 0.52$. The code corresponding to index 0, i.e. state ω_0, say $\mathbf{x}_0$, is generated by n independent realizations of $\widehat{\theta}$. Similarly, $\mathbf{x}_1$ is the code corresponding to index 1, i.e. state ω_1. Let us consider that a code is chosen *uniformly* at random and sent through the noisy channel (by sending n bits one after the other).

4.1 A code fulfilling the Nash equilibrium conditions

We present first the realization of the "natural code" in full detail because it is quite familiar and will help the reader to follow later a more complicated example. To make the analysis very simple consider that the communication goes for 3 periods and let Γ^3_ν be the noisy communication extended game.

Suppose that a specific and common knowledge realization of the random code is:

$$\begin{bmatrix} x_1(0) & x_2(0) & x_3(0) \\ x_1(1) & x_2(1) & x_3(1) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_0 = 0,0,0 \\ \mathbf{x}_1 = 1,1,1 \end{bmatrix}$$

Nature informs the sender about the true state of nature, therefore, the sender's strategy σ^S_j, $j = 0,1$ is sending:

$$\begin{aligned} \sigma^S_0 &= \mathbf{x}_0 = 000, \text{ if } \omega = \omega_0 \\ \sigma^S_1 &= \mathbf{x}_1 = 111, \text{ if } \omega = \omega_1 \end{aligned}$$

The receiver observes a transformed sequence $\mathbf{y}$, with transition probability $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^3 p(y_i|x_i)$ and tries to guess which message has been sent. He will consider that index j was sent if $(\mathbf{x}_j, \mathbf{y})$ are jointly typical and there is no other index k, such that $(\mathbf{x}_k, \mathbf{y})$ are jointly typical. If no such index j exists, then an error will be declared.

Let us proceed to construct the receiver's strategy, by generating a partition of the set of outcome sequences $\mathbf{Y} = \{0,1\}^3$. To apply the jointly typical decoding rule, it is needed to calculate the functions[3]:

$$\begin{aligned} \Delta_{\mathbf{x}_0}(\mathbf{y}) &= |-\frac{\log(p(\mathbf{x}_0,\mathbf{y}))}{3} - H(X,Y)| \\ \Delta_{\mathbf{x}_1}(\mathbf{y}) &= |-\frac{\log(p(\mathbf{x}_1,\mathbf{y})}{3} - H(X,Y)| \end{aligned}$$

[3] Notice that only the third condition in the definition of jointly typical sequences is the binding condition to be checked.

which measures the difference between the empirical entropy of each sequence in $\mathbf{Y}$ and the true entropy $H(X,Y)=1,6$.

For example, for $\mathbf{y}=000$, for our specific channel $\nu(0.1,\,0.2)$ and since $\widehat{\theta}=0.5$, then $p(\mathbf{y}=000|\mathbf{x}_0=000)=(p(0|0))^3=(1-\varepsilon_0)^3=0.9^3=0.59$; $p(\mathbf{y}=000|\mathbf{x}_1=111)=(p(0|1))^3=\varepsilon_1^3=0.2^3=0.0003$; $p(\mathbf{x}_0,\mathbf{y})=p(\mathbf{y}|\mathbf{x}_0)p(\mathbf{x}_0)=0.59\times(0.5)^3$, and $p(\mathbf{x}_1,\mathbf{y})=p(\mathbf{y}|\mathbf{x}_1)p(\mathbf{x}_1)=0.0003\times(0.3^3)$, and then:

$$\Delta_{\mathbf{x}_0}(\mathbf{y}=000)=0.485 \text{ and } \Delta_{\mathbf{x}_1}(\mathbf{y}=000)=1.801$$

Now we have to choose an η-approximation in order to partition the output message space. Fix $\eta=0.64$. The reason for such a choice will become clear at the end of the example. Recall that such value is the upper bound of the distance between the empirical entropy and the true entropy to define jointly typical sequences. Then, the jointly typical decoding rule states that a given $\mathbf{y}\in\mathbf{Y}$ is jointly typical with $\mathbf{x}_0=000$, and with $\mathbf{x}_1=111$, respectively, whenever

$$\begin{aligned}\Delta_{\mathbf{x}_0}(\mathbf{y}) &< \eta=0.64\\ \Delta_{\mathbf{x}_1}(\mathbf{y}) &< \eta=0.64,\ \text{ respectively}\end{aligned}$$

The jointly typical decoding rule allows the receiver to define the following subsets of $\mathbf{Y}$,

$$\begin{aligned}P_0^0 &= \{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_0}(\mathbf{y})<\eta\}\\ P_0^{\neg 0} &= \{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_0}(\mathbf{y})\geq\eta\}\\ P_1^{\neg 1} &= \{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_1}(\mathbf{y})\geq\eta\}\\ P_1^1 &= \{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_1}(\mathbf{y})<\eta\}\end{aligned}$$

The first set P_0^0 contains all the sequences in $\mathbf{Y}$ that are probabilistically related to input sequence $\mathbf{x}_0=000$. Conversely, set $P_0^{\neg 0}$ refers to all the sequences of $\mathbf{Y}$ that are not probabilistically related to $\mathbf{x}_0$. Similarly, P_1^1 is the set of sequences in $\mathbf{Y}$ that are probabilistically related to input sequence $\mathbf{x}_1=111$, while $P_1^{\neg 1}$ is the set of sequences in $\mathbf{Y}$ that cannot be related to $\mathbf{x}_1$. These sets are:

$$\begin{aligned}P_0^0 &= \{000,001,010,100\}\\ P_0^{\neg 0} &= \{111,110,101,011\}\\ P_1^{\neg 1} &= \{000,001,010,100\}\\ P_1^1 &= \{111,110,101,011\}\end{aligned}$$

Denote by

$$\begin{aligned}P_0 &= P_0^0\cap P_1^{\neg 1}=\{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_0}(\mathbf{y})<\eta \text{ and } \Delta_{\mathbf{x}_1}(\mathbf{y})\geq\eta\}\\ P_1 &= P_0^{\neg 0}\cap P_1^1=\{\mathbf{y}\in\mathbf{Y}:\Delta_{\mathbf{x}_1}(\mathbf{y})<\eta \text{ and } \Delta_{\mathbf{x}_0}(\mathbf{y})\geq\eta\}.\end{aligned}$$

the set of all sequences of $\mathbf{Y}$ which are *uniquely* related in probability to $\mathbf{x}_0$ and $\mathbf{x}_1$, respectively. Since, $P_0^0 = P_1^{\neg 1}$ this implies that no matters whether $\mathbf{x}_0$ or $\mathbf{x}_1$ has been sent, the receiver univocally assigns $\mathbf{x}_0$ to all sequences in P_0^0 or $P_1^{\neg 1}$. Similarly, $P_0^{\neg 0} = P_1^1$ implies that the receiver decodes all the sequences in either of these sets as corresponding to $\mathbf{x}_1$. Moreover, since $P_0 \cap P_1 = \varnothing$ and $P_0 \cup P_1 = \mathbf{Y}$, then the typical decoding rule generates a true partition. In fact, the jointly typical decoding rule is in this case equivalent to the majority rule decoding. To see this let $\mathbf{y}^k$ be an output sequence with k zeros. Then,

$$p(\mathbf{x}_0 \mid \mathbf{y}^k) = \frac{p(\mathbf{y}^k \mid \mathbf{x}_0)p(\mathbf{x}_0)}{p(\mathbf{y}^k)} = \frac{(1-\varepsilon_0)^k \varepsilon_0^{3-k}}{(1-\varepsilon_0)^k \varepsilon_0^{3-k} + \varepsilon_1^k (1-\varepsilon_1)^{3-k}} \geq \frac{1}{2}$$

if and only if $k \geq 2$.

The jointly typical decoding rule gives rise to the receiver's strategy, for each $\mathbf{y} \in \mathbf{Y}$:

$$\sigma_{\mathbf{y}}^R = a_i, \text{ whenever } \mathbf{y} \in P_i$$

To show that the above strategies are a Nash equilibrium in pure strategies, let us check that both the sender and the receiver's strategies are a best response to each other.

1) The receiver's Nash equilibrium condition translates to her choice of action a_0 whenever $p(\mathbf{y}|\sigma_0^S) \geq p(\mathbf{y}|\sigma_1^S)$, and of action a_1 otherwise. In table 1 below it can be checked that all output sequences $\mathbf{y}$, that satisfy with strict inequality the condition $p(\mathbf{y}|\sigma_0^S) \geq p(\mathbf{y}|\sigma_1^S)$ are exactly those belonging to set P_0, and those for which $p(\mathbf{y}|\sigma_1^S) \geq p(\mathbf{y}|\sigma_0^S)$ with strict inequality are the ones in P_1. Therefore the receiver's jointly typical decoding rule is a best response to the sender's coding strategy.

$\mathbf{y}$	$p(\mathbf{y}\|\mathbf{x}_0)$	$p(\mathbf{y}\|\mathbf{x}_1)$	$\mathbf{y}$
000	0.729	0.008	000
001	0.081	0.032	001
010	0.081	0.032	010
011	0.009	0.128	011
100	0.081	0.032	100
101	0.009	0.128	101
110	0.009	0.128	110
111	0.001	0.512	111

Table 1

2) The sender's Nash equilibrium condition, given the receiver's jointly typical decoding, amounts to choosing input sequences σ_0^S and σ_1^S, in states ω_0 and ω_1,

respectively, such that

$$\begin{aligned}\sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_0^S)u(\sigma_{\mathbf{y}}^R,\omega_0) &= \sum_{\mathbf{y}\in\mathbf{P}_0} p(\mathbf{y}|\sigma_0^S) \geq \sum_{\mathbf{y}\in\mathbf{P}_0} p(\mathbf{y}|\sigma_0'^S) = \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_0'^S)u(\sigma_{\mathbf{y}}^R,\omega_0)\\ \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_1^S)u(\sigma_{\mathbf{y}}^R,\omega_1) &= \sum_{\mathbf{y}\in\mathbf{P}_1} p(\mathbf{y}|\sigma_1^S) \geq \sum_{\mathbf{y}\in\mathbf{P}_1} p(\mathbf{y}|\sigma_1'^S) = \sum_{\mathbf{y}\in\mathbf{Y}} p(\mathbf{y}|\sigma_1'^S)u(\sigma_{\mathbf{y}}^R,\omega_1)\end{aligned}$$

for any other input sequences $\sigma_0'^S$ and $\sigma_1'^S$, respectively.

Let $\sum_{\mathbf{y}\in\mathbf{P}_0} p(\mathbf{y}|\mathbf{x}_0)$ and $\sum_{\mathbf{y}\in\mathbf{P}_1} p(\mathbf{y}|\mathbf{x}_1)$ denote the aggregated probabilities of the sequences in $\mathbf{P}_0$ and $\mathbf{P}_1$ when input sequences $\mathbf{x}_0$ and $\mathbf{x}_1$ are sent. Given the symmetry of the sequences it suffices to check the ones shown in the table 2 below:

$\mathbf{x}_0$	$\sum_{\mathbf{y}\in\mathbf{P}_0} p(\mathbf{y}\|\mathbf{x}_0)$	$\sum_{\mathbf{y}\in\mathbf{P}_1} p(\mathbf{y}\|\mathbf{x}_1)$	$\mathbf{x}_1$
000	0.972	0.028	000
001	0.846	0.154	001
011	0.328	0.672	011
111	0.104	0.896	111

Table 2

Clearly, if the state is ω_0, then obeying the communication protocol and sending $\sigma_0^S = 000$ will be a best reply to the receiver's strategy, since sending instead any other input sequence will only decrease the sender's payoffs, as shown in the left hand side of the above table. Similarly, if the state is ω_1, sending $\sigma_1^S = 111$ will maximize the sender's payoffs against the receiver's strategy, as shown in the right hand side of the above table.

To conclude this example we display in Figure 1 the relationship between the η-approximation and the existence of an output set partition. The horizontal axes represents the output set sequences and the vertical axes are the functions $\Delta_{\mathbf{x}_0}(\mathbf{y})$ (the dotted line) and $\Delta_{\mathbf{x}_1}(\mathbf{y})$ (the continuos line) for the natural coding $\mathbf{x}_0 = 000$ and $\mathbf{x}_1 = 111$. Different values of η have been plotted in the same Figure 1. We obtain the following remarks:

- For an $\eta = 0.9$ and $\mathbf{y} \in \mathbf{Y}$, if the value of $\Delta_{\mathbf{x}_0}(\mathbf{y})$ goes by above of the constant function $\eta = 0.9$, then that of $\Delta_{\mathbf{x}_1}(\mathbf{y})$ will go by below of η, and the same will happen in the other way around. By the Jointly Typical condition every $\mathbf{y}$ is uniquely related in probability to either $\mathbf{x}_0$ or $\mathbf{x}_1$. Therefore for $\eta = 0.9$ a partition of set $\mathbf{Y}$ is easily generated.

- The same reasoning applies to any η in $(0.6, 1.08)$. This is why we have chosen $\eta = 0.64$.

- For $\eta \geq 1.08$ or $\eta \leq 0.6$, there are output sequences belonging to both the output set associated to $\mathbf{x}_0$ *and* that associated to $\mathbf{x}_1$. Hence, there is a need to uniquely reassign those sequences to one of the them.

  In sum, under the natural coding $\mathbf{x}_0 = 000$ and $\mathbf{x}_1 = 111$ it is possible to find a range of η which enables to construct a partition of the output set and therefore support the strategies of the communication protocol as a Nash equilibrium of the extended communication game.

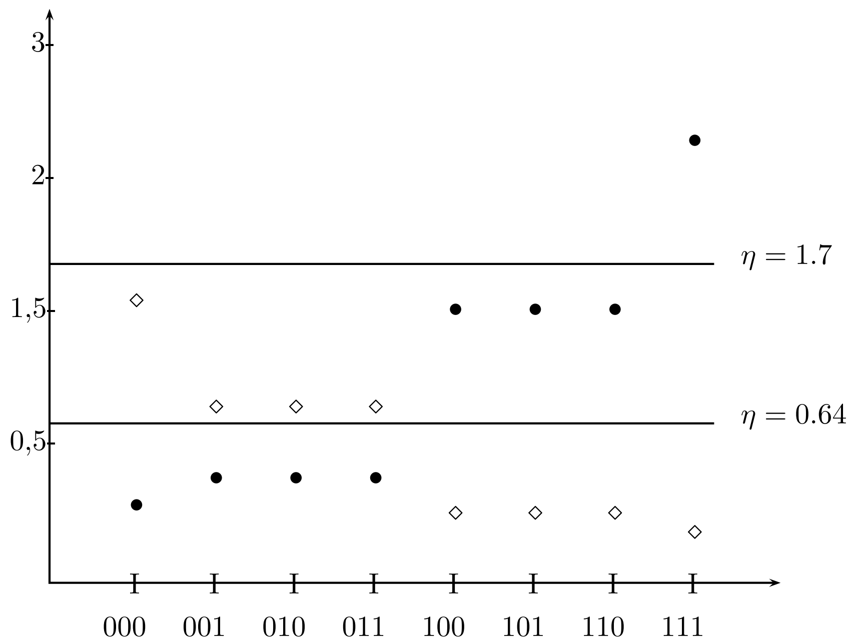

Figure 1: Partition of the output message space around $\mathbf{x}_0 = 000$, $\mathbf{x}_1 = 111$.

However, other realizations of the random code might not guarantee the existence of such an η to construct such partition as the following code realization shows.

4.2 A receiver's deviation

Suppose that a new realization of the code is:

$$\begin{bmatrix} x_1(0) & x_2(0) & x_3(0) \\ x_1(1) & x_2(1) & x_3(1) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_0 = 0,1,0 \\ \mathbf{x}_1 = 0,1,1 \end{bmatrix}$$

where, as above, the channel is $\nu(\varepsilon_0,\varepsilon_1) = \nu(0.1,0.2)$ and Γ^3_ν is the noisy communication extended game. Fix now $\eta = 0.37$.

Let us consider that the receiver observes the output sequence $\mathbf{y} = 010$. Let us calculate $p(\mathbf{y} = 010|\mathbf{x}_0 = 010) = 0.648$ and $p(\mathbf{y} = 010|\mathbf{x}_1 = 011) = 0.144$, and the functions:

$$\begin{aligned}\Delta_{\mathbf{x}_0}(\mathbf{y}) &= |-\frac{\log(p(\mathbf{x}_0,\mathbf{y}))}{3} - H(\mathbf{X},\mathbf{Y})| = 0.40 \\ \Delta_{\mathbf{x}_1}(\mathbf{y}) &= |-\frac{\log(p(\mathbf{x}_1,\mathbf{y})}{3} - H(\mathbf{X},\mathbf{Y})| = 0.36\end{aligned}$$

For $\eta = 0.37$, Shannon protocol dictates that the receiver decodes $\mathbf{y}$ as $\mathbf{x}_1$ and plays action a_1. This situation would correspond with case 3 in subsection 3.1 where the protocol may not order the conditional probabilities. In fact, the Nash equilibrium condition for the receiver when $\mathbf{y} = 010$ translates to choosing action a_0 since, as shown above, the conditional probability of $\mathbf{y}$ given $\mathbf{x}_0 = 010$ (0.648) is bigger than the conditional probability of $\mathbf{y}$ given $\mathbf{x}_1 = 011$ (0.144).

4.3 A sender's deviation

Fix now[4] $n = 5$ and suppose that the specific and common knowledge realization of the random code is the following:

$$\begin{bmatrix} x_1(0) & x_2(0) & \dots & x_5(0) \\ x_1^(1) & x_2^(1) & \dots & x_5^(1) \end{bmatrix} = \begin{bmatrix} \mathbf{x}_0 = 0,0,0,0,0 \\ \mathbf{x}_1 = 0,0,0,1,1 \end{bmatrix}$$

where the two signal strings share the first three digits, and therefore only the last two digits are different.

Then σ_j^S, $j = 0, 1$ is:

$$\begin{aligned}\sigma_0^S &= \mathbf{x}_0 = 00000, \text{ if } \omega = \omega_0 \\ \sigma_1^S &= \mathbf{x}_1 = 00011, \text{ if } \omega = \omega_1\end{aligned}$$

To construct the receiver's strategy, we repeat the above computations of sets P_0^0, $P_0^{\neg 0}$, $P_1^{\neg 1}$, P_1^1, P_0 and P_1 of $\mathbf{Y}$.

Notice that $P_0^0 \neq P_1^{\neg 1}$ implies that the receiver cannot univocally assign some $\mathbf{y}$ in $\mathbf{Y}$ to $\mathbf{x}_0$ no matter whether $\mathbf{x}_0$ or $\mathbf{x}_1$ has been sent. Similarly, $P_0^{\neg 0} \neq P_1^1$ with the same meaning for $\mathbf{x}_1$. Therefore, $P_0 \cup P_1 \subsetneq \mathbf{Y}$. Let us define the set $P_2 = \mathbf{Y} - P_0 \cup P_1$:

$$\begin{aligned}P_2 &= \{\mathbf{y} \in \mathbf{Y} : \Delta_{\mathbf{x}_0}(\mathbf{y}) < \eta \text{ and } \Delta_{\mathbf{x}_1}(\mathbf{y}) < \eta\} \cup \{\mathbf{y} \in \mathbf{Y} : \Delta_{\mathbf{x}_1}(\mathbf{y}) \geq \eta \text{ and } \Delta_{\mathbf{x}_0}(\mathbf{y}) \geq \eta\} \\ &= \{00100, 00111, 01000, 01011, 01100, 01111, 10000, 10011, 10100, 10111, 11000, 11011\}.\end{aligned}$$

This set contains all the sequences in $\mathbf{Y}$, which the receiver is not able to decode, i.e., any $\mathbf{y} \in P_2$ cannot be univocally assigned either to $\mathbf{x}_0$ or $\mathbf{x}_1$: the errors in

[4] We run a systematic search computation for a sender's deviation when $n < 5$ and we concluded that there was none.

Shannon's approach. Therefore, the jointly typical decoding does not generate a partition of $\mathbf{Y}$, and the receiver does not know how to take an action in Γ.

There is a need then to assign the sequences in P_2 to either P_0 or P_1. Consider that the specific rule is to assign each sequences $\mathbf{y} \in P_2$, to that element of the input sequence which is probabilistically closer to them[5], namely

$$\mathbf{y} \in P_0 \text{ if } \Delta_{\mathbf{x}_0}(\mathbf{y}) < \Delta_{\mathbf{x}_1}(\mathbf{y}), \text{ and } \mathbf{y} \in P_1 \text{ otherwise.}$$

Then:

$$\begin{aligned} P_0 &= \{00100, 01000, 01100, 10000, 10100, 11000, 11100\} \\ P_1 &= \{00000, 00001, 00010, 00011, 00101, 00110, 00111, 01001, \\ &\quad 01010, 01011, 01101, 01110, 01111, 10001, 10010, 10011, 10111, \\ &\quad 10011, 10111, 10110, 10111, 11001, 11010, 11011, 11101, 11110, 11111\} \end{aligned}$$

Therefore, $P_0 \cap P_1 = \varnothing$ and $P_0 \cup P_1 = \mathbf{Y}$, and the partition gives rise to the receiver's strategy $\sigma_{\mathbf{y}}^R = a_i$, whenever $\mathbf{y} \in P_i$, and for each $\mathbf{y} \in \mathbf{Y}$.

Recalling that $p(P_0) = \sum_{\mathbf{y} \in \mathbf{P_0}} p(\mathbf{y}|\sigma_0^S)$ and $p(P_1) = \sum_{\mathbf{y} \in \mathbf{P}_1} p(\mathbf{y}|\sigma_1^S)$, then it is easy to calculate that $p(P_0) = 0.729$ and $p(P_1) = 0.271$.

Consider the sender' deviation, i.e.,

$$\begin{aligned} \sigma_0^{dS} &= \mathbf{x}_0^d = 11100, \text{ if } \omega = \omega_0, \text{ instead of } \sigma_0^S = \mathbf{x}_0 = 00000 \\ \sigma_1 &= \mathbf{x}_1 = 00011, \text{ if } \omega = \omega_1 \end{aligned}$$

This deviation does not change the partition but does change the probability associated to sets P_0 and P_1. In particular, $\sum_{\mathbf{y} \in \mathbf{P}_1} p(\mathbf{y}|\mathbf{x}_0 = 00000) = 0.21951$ and $\sum_{\mathbf{y} \in \mathbf{P}_1} p(\mathbf{y}|\mathbf{x}_0^d = 00011) = 0.98916$.

Suppose that $\omega = \omega_0$ and let σ_0^S and $\sigma_{\mathbf{y}}^R$ be the strategies of following faithfully the protocol in Γ_ν^5, for each $\mathbf{y} \in \mathbf{Y}$. Then, the sender's expected payoffs are

$$\begin{aligned} \pi_0^S &= \pi_0^S(\sigma_0^S, \{\sigma_{\mathbf{y}}^R\}_{\mathbf{y}}) = \sum_{\mathbf{y} \in \mathbf{P}_0} p(\mathbf{y}|\sigma_0^S)1 = 0.21951 \\ \pi_0^{dS} &= \pi_0^S(\sigma_0^{dS}, \{\sigma_{\mathbf{y}}^R\}_{\mathbf{y}}) = \sum_{\mathbf{y} \in \mathbf{P}_0} p(\mathbf{y}|\sigma_0^{dS})1 = 0.80352 \end{aligned}$$

and the sender will then deviate.

[5] This rule is in the spirit of the maximum likelihood criterion.

5 Concluding remarks

Information Theory tells us that whatever the probability of error in information transmission, it is possible to construct error-correcting codes in which the likelihood of failure is arbitrarily low. In this framework, error detection is the ability to detect the presence of errors caused by noise, while error correction is the additional ability to reconstruct the original error-free data. Detection is much simpler than correction, and the basic idea is to add one or more "check" digits to the transmitted information (e.g., some digits are commonly embedded in credit card numbers in order to detect mistakes). As is common in Information Theory protocols, both the sender and the receiver are committed to use specific rules in order to construct error correcting/detecting codes.

Shannon's theorem is an important theorem in error correction which describes the maximum attainable efficiency of an error-correcting scheme for expected levels of noise interference. Namely, Shannon's Theorem is an asymptotic result and establishes that for all small tolerance it is possible to construct error-correcting codes in which the likelihood of failure is arbitrarily low, thus providing necessary and sufficient conditions to achieve a good information transmission. Nevertheless, the asymptotic nature of such a protocol masks the difficulties to apply information theory protocols to finite communication schemes in strategic sender-receiver games.

In this paper we consider a game-theoretical model where a sender and a receiver are players trying to coordinate their actions through a finite time communication protocol *à la* Shannon. Firstly, given a common knowledge coding rule and an output message, we offer the Nash equilibrium condition for the extended communication game. Specifically, the receiver's equilibrium conditions are summarized by choosing the action corresponding to that state of nature for which the conditional probability of the received message is higher. This implies an ordering of the probability of receiving a message conditional to any possible input message. On the other hand, given the realized state of nature and the receiver's partition of the output message space generated by the coding and the decoding rules, the sender's equilibrium conditions are specified by choosing the input message maximizing the sum of the above conditional probabilities over all output messages belonging to the partition corresponding to that state of nature. Secondly, we relate the Nash equilibrium strategies to those of Shannon's coding and decoding scheme. Particularly, we rewrite the receiver's Nash constraint in terms of the entropy condition of the Jointly Typical Set, pointing out that such entropy condition may not be enough to guarantee the partition of the output space. Finally, we provide two counterexamples to illustrate our findings.

Consequently, coding and decoding rules under Information Theory satisfy a set of information transmission constraints, but they may fail to be Nash equilibrium strategies.

References

[1] Blume, A., O. J. Board and K. Kawamura (2007): "Noisy Talk", Theoretical Economics, Vol. 2, 395–440.

[2] Cover, T. M. and J. A. Thomas (1991): *Elements of information theory.* Wiley Series in Telecomunications. Wiley.

[3] Crawford, V. and J. Sobel (1982): "Strategic Information Transmission", Econometrica, Vol. 50, 1431–1451.

[4] Gossner, O., P. Hernández and A. Neyman (2006): "Optimal use of communication resources", Econometrica, Vol. 74, 1603–1636.

[5] Gossner, O. and T. Tomala (2006): "Empirical Distributions of beliefs under imperfect monitoring", Mathematics of Operations Research, Vol. 31, 13–31.

[6] Gossner, O. and T. Tomala (2007): "Secret Correlation in Repeated Games with Imperfect Monitoring", Mathematics of Operations Research, Vol. 32, 413–424.

[7] Gossner, O. and T. Tomala (2008): "Entropy bounds on Bayesian Learning", Journal of Mathematical Economics, Vol. 44, 24–32.

[8] Hernández, P. and A. Urbano. (2008): "Codification Schemes and Finite Automata, Mathematical Social Sciences Vol. 56, 3, 395–409.

[9] Hernández, P., A. Urbano and J. Vila (2010): "Grammar and Language: An Equilibrium Approach", Working Paper ERI-CES 01/2010.

[10] Koessler, F. (2001): "Common Knowledge and Consensus with Noisy Communication", Mathematical Social Sciences, 42(2), 139–159.

[11] Mitusch, K. and R. Strausz (2005): "Mediation in Situations of Conflict and Limmited Commitment", Journal of Law, Economics and Organization, vol. 21(2), 467–500.

[12] Myerson, R. (1991): Game Theory. Analysis of conflict. Harvard University Press, Cambridge, Massachusetts, London, England.

[13] Rubinstein, A. (1989): "The Electronic Mail Game: A Game with Almost Common Knowledge", American Economic Review 79, 385–391.

[14] Shannon, C.E. (1948): "A Mathematical Theory of Communication," Bell System Technical Journal, 27, 379–423; 623–656.