Pragmatic Languages with Universal Grammars: An Equilibrium Approach

Penélope Hernández, Amparo Urbano and José E.Vila*

July 5, 2009

Abstract

The aim of this paper is to explore the role of a *pragmatic* Language with a *universal* grammar as a coordination device under communication misunderstandings. Such a language plays a key role in achieving efficient outcomes in those sender-receiver games, where there may exist noisy information tranmission. The Language is pragmatic in the sense that the Receiver' best response depends on the *context*, i.e, on the payoffs and on the initial probability distribution of the states of nature of Γ. The Language has a universal grammar because the Sender's coding rule does not depend on such specific parameters of Γ and can then be applied to any sender-receiver game with noisy communication.

The common knowledge "corpus" or set of standard prototypes designed by the Sender and the Receiver's "pragmatic variations" around the standard prototypes, generate an equilibrium pragmatic Language. Furthermore, such a Language is efficient: in spite of initial misunderstandings, the Receiver is able to infer with a high probability the Sender's meaning and thus expected payoffs are close to those of communication without noise.

1 Introduction

The aim of this paper is to explore the role of a *pragmatic* Language with a *universal* grammar as an equilibrium coordination device under communication misunderstandings. Such a language plays a key role in achieving efficient outcomes in those sender-receiver games, where there may exist noisy information tranmission and the length of communication is finite. The Language is pragmatic in the sense that the Receiver' best response depends on the *context*, i.e, on the payoffs and on the initial

*Departament of Economic Analysis and ERI-CES. University of Valencia. Facultad de Economia, Campus dels Tarongers. 46022 Valencia (Spain). e-mails: penelope.hernandez@uv.es; amparo.urbano@uv.es and jose.e.vila@uv.es. The authors thank both the Spanish Ministry of Science and Technology and the European Feder Founds for financial support under project SEJ2007-66581.

probability distribution of the states of nature of Γ. The Language has a universal grammar because the Sender's grammar does not depend on such specific parameters of Γ and can then be applied to any sender-receiver game with noisy communication. The common knowledge "corpus" or set of "standard prototypes" designed by the Sender and the Receiver's "pragmatic variations" around the standard prototypes, generate an equilibrium pragmatic Language.

One of the advantages of human language is that it conveys information and makes it possible to cooperate about future goals (Gärdenfors, 2002). When communication first appeared, the most important issue was the communicative act in itself and the context it occurred in, not the expressive form of the act. When communicative acts (later speech acts) in due time became more varied and eventually conventionalized and detached from the immediate context, one could start analyzing the different *meanings* of the acts. Thus, when we nowadays communicate by language, our *utterances* (signals conveying information whithin a context) have meaning. Language is therefore a symbolic system of communication making it possible the inference of meaning. In fact, the meaning of a linguistic utterance is not transmitted directly, but is inferred indirectly by the hearer, through pragmatic insights and the social context in which the utterance is received. Furthermore, when linguistic communication becomes even more conventionalized and combinatorially richer, certain markers, alias *syntax*, are used to disambiguate the communicative context.

These considerations lead naturally to queries about the efficiency of language to communicate and learn, inference models of creation of meanings and the role of grammar and categorization in linguistic structures. Several answers have come from Linguistics (Grice, 1969,1975; Gärdenfors, 2000, 2002; Azrieli and Lehrer, 2007; Jäger, 2007, etc.), Mathematics and Computer Science (Batali, 1998; Nowak and Krakauer, 1999; Smith, 2003, 2003a; Kirby, 2001, 2002,2007; Voght, 2000, 2005, etc). Nevertheless the importance of language in Economics has not been yet sufficiently stressed with the exception of a handful papers such as Rubinstein (1996, 2000), Blume (2000, 2005), Blume and Board (2008), Segal (2001), Selten and Warglien (2007), Balinski and Laraki (2007a,b) among others.

Coordination takes place many times in different environments and in diverse contexts and here human "natural" languages may experience some difficulties in trying to reduce inefficient outcomes. For instance, empirical analyses have revealed that the use of non-standard or non-native grammatical variants only rarely leads to any communication breakdown, whereas most breakdowns occur due to lexical or phonetic obstacles[1]. To explore the role of languages as a coordination devices under communication misunderstandings, we study the design and use of a symbolic language leading to coordination. The language that we propose is *pragmatic*

[1] In fact, as reported in Reiter and Sripada (2002), linguists have acknowledged that people may associate different meanings with the same word.

in the sense of Grice (1975). Pragmatics examines the influence of *context*[2] on the interpretation of an utterance. Our context is a sender-receiver game with noisy information transmission. We find that communication by such a pragmatic language is an equilibrium outcome of the noisy communication game. Furthermore, the grammar of our language is *universal* because it only depends on the number of the meanings and the length of the communication episode, and not on the rewards of each meaning. Our encoding rule (the syntax language) ensures the successful transmission of language itself and universality guarantees the language implementation in very different contexts. Thus, our work stresses that of Nowak and Krakauer (1999) who argue that grammar can be seen as a simplified rule system that reduces the chances of mistakes in implementation and comprehension and it is therefore favored by natural selection in a world where mistakes are possible. Our results also theoretically match those of Selten and Warglien (2007), who show in a series of laboratory experiments that in an environment with novelty compositional grammar offer considerable coordination advantages and therefore is more likely to arise.

It is assumed that the context, the set of public signals, the grammar rules (encoding rule) and the length of the communication are common knowledge but there may be communication misunderstandings. For instance, though native English speakers may remember 80.000 words, very few of them will use more than 7500 English words in their communication and even in this case communication misunderstandings may appear[3]. To develop this idea formally, we consider a sender-receiver game with aligned interests, where the costs of miscoordination are different in distinct states of nature. When an individual communicates with another one, the former has a precise purpose in mind ("a meaning"), that has to be expressed by a common language. Thus, the informed Sender, has to tell the uninformed one, the Receiver, which action to choose. The set of the Sender's meanings is the set of the Receiver's actions. To communicate the Sender encodes her meanings in a set of public signals which are sent to the Receiver. Each signal can, in principle, be anything, for example, letters or merely symbols in an alphabet of salient features which is used to create signals. Signals could be subject to extraneous factors that would distort or interfere with its reception. This unplanned distortion or interference is known as *noise*[4]. The noise we are concerned with is such that the messages are

[2]In Linguistics, a context comprises the speaker, the hearer, the place, the time and so forth. How the hearer views the intentions of the speaker and how the speaker views the presuppositions of the hearer are relevant to the understanding of an utterance.

[3]These misunderstandings can even be worse due to the fact that English is the most frequent language in business conversations between people with different mother tongues. Most of the times, non-native are faced with non-native English (in conversation with non-natives) or non-standard English (e.g. when reading CNN headlines).

[4]Noise refers to anything introduced into messages that is not included in them by the Sender. Noise may range from mechanical noise, such as the distortion of a voice in the telephone or the interference with a television signal producing "snow" on the TV screen, to any noise generated in human communications. A more general and basic situation is where the noise underlies the

always received by the Receiver but they can differ from those sent by the Sender, that is, signals may be distorted in the communication process. We model this noise by assuming that each signal can randomly be mapped to the whole set of possible signals[5]. We assume that the set of signals is a discrete alphabet and that the language dictionary is a combination of elements of this set. More precisely, given a communication length, an input sequence is a concatenation of signals. The Sender utters one of these sequences and the Receiver hears an output sequence, which is a probabilistic transformation of the signal string.

The encoding rule or grammar is used to design a communication device between the Sender and the Receiver. From a given and common knowledge set of public signals and a communication length, the Sender defines a dictionary or "corpus" of sequences which generates the set of *standard prototypes* which are a one-to one mapping into the set of Receiver's actions. The structure or grammar specifies that each prototype sequence is positionally arranged and maximally separated from any other sequence. The Receiver has to infer a meaning from each heard sequence or, in other words, to assign an action to each received sequence. Without any noise, the Receiver would accurately infer the action to play from any received prototype. With noise each received sequence belongs to the whole set of possible language sequences and could have been generated from any prototype. Then, the Receiver's criterion is a best-response decoding which partitions the set of possible language sequences into subsets, with a unique action assignment to each of them. In linguistic terms, each of these subset is called the *pragmatic variation* of a given prototype. The way in which each sequence is assigned to a particular pragmatic variation instead of the others, is given by some measure of vicinity or "distance" between the sequence and each one of the other standard prototypes. The bound on such distance from a given prototype is called "*the vicinity bound*".

The partition of the output space in the pragmatic variations of the standard prototypes is related to the work on categorization based on prototypes (see Azrieli and Lehrer, 2007, Jäger, 2007 and references herein). In particular, Jäger investigates communication in a partnership signaling game where the set of meanings is equipped with a Euclidean geometrical structure. Perfect communication is not possible because the number of meanings exceeds the number of signals. Under an

---

information transmission technology. This setting refers to information transmission situations where the signals can be garbled or corrupted in their transmission

[5]In economics there are many situations where "rational" agents have erroneous perceptions, there are signaling models with noise, and, in general, information transmission models with incomplete information such as those of Crawford and Sobel (1982), Lipman and Seppi (1995), Koessler (2000, 2001, 2004), among others. In many of these models the noise mainly refers to the strategic uncertainty of the agents about the relevant parameters of the strategic situation under study rather than communication misunderstandings. Blume, Board and Kawamura (2007) examines the possibilities for communication in the Crawford and Sobel's model in a noisy environment. In linguistics, Nowak, Krakauer and Dress (1999) investigate the evolution of communication in the presence of noise: individual may mistake one signal for another

evolutionary approach, he shows that the sender's strategy partitions the meaning space into quasi-convex categories. In our noisy communication game, the receiver's best reply to any pure sender's strategy induces a categorization of the underlying set of public signals around the standard protoypes. This partition is due to an inference process rather than to an evolutionary dynamics.

On the other hand, the structure of a language helps to enhance efficiency in communication. In Linguistics Kirby (2001, 2002, 2007) focuses on the emergence of composition and recursion in languages. In Economics, Rubinstein (1996, 2000) is concerned with the structure of binary relations appearing in natural language and Blume (2000, 2005) explores the use of structure in languages and how such efficient structures facilitate coordination and learning in repeated coordination games. Furthermore, Blume and Board (2008) take it as given that language is an imperfect technology that leaves messages subject to interpretation. Contrary to us, they investigate the strategic use of interpretable messages. Crémer, Garicano and Prat (2007), characterize efficient technical languages and study their interaction with the scope and structure of organizations. A recent paper on the evolution of language is Demichelis and Weibull (2008), were two parties have a common language and agree on its meaning. They show that such a shared culture -language and honesty code- facilitates coordination on socially efficient equilibrium outcomes in strategic interactions.

A *Language* is defined as the pair corpus and pragmatic variations: the way in which the Sender transmits the meanings and that of the Receiver's understandings. An equilibrium pragmatic Language is that for which the players' strategies are a best response to each other. We show that to communicate by such a Language is an equilibrium of the noisy communication game: the pragmatic Language is a coordination device under interpretation failures. In fact, it performs quite well as an inference of meaning model, that is, in spite of initial misunderstandings, the Receiver is able to infer with a high probability the Sender's meaning. This result guarantees expected payoffs close to those of communication without noise. Alternatively, we characterize the time needed to span the pragmatic variations in order to reduce the chances of misunderstandings and increase expected payoffs. The pragmatic Language facilitating coordination shares the spirit of Balinski and Laraki (2007)'s work in the Theory of Social Choice. They show that a more "realistic" model in this field, in the sense that messages are grades expressed in a common language, allows preferences to be aggregated.

The paper is organized as follows. Section 2 presents the one-shot game, and the extended communication game. The existence of a language supporting players' coordination is presented in section 3, where the equilibrium pure (separating) strategies for the Sender and the Receiver are constructed. To highlight the main features of our construction we offer some examples in section 4. In section 5, the efficiency of our equilibrium for finite communication length is measured and the speed of convergence of ex-ante payoffs is characterized. Some concluding remarks close the paper.

2 The Model

2.1 The coordination game

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 $\Omega = \{\omega_0, \ldots, \omega_{|\Omega|-1}\}$. Nature chooses first randomly $\omega_j \in \Omega$ with probability q_j and then the sender is informed of such 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 $u : A \times \Omega \rightarrow R$ be the (common) players' payoff function, i.e., $u(a_t, \omega_j)$, $j = 0, 1, ..., |\Omega| - 1$. 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:

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

The sender observes the value of ω and then sends a message or string of signals from some message space[6]. In sender-receiver games, players try to share their private information to achieve coordination. Hence, they usually communicate using a human or an artificially constructed language. More precisely, the Sender has a precise purpose in mind or meaning, that has to be expressed in the form of a signal or a message. The Receiver has to decode the message to infer the original Sender's purpose, jointly with the context. In our setting the context is the coordination game and the Sender meanings are the Receiver actions. To communicate the Sender encodes the meanings to be transmitted in a set of public signals, from the underlying common language, which are sent to the Receiver. Each signal or utterance can, in principle, be represented in any form, for example, letters or merely symbols in an alphabet of salient features which is used to create signals. To simplify the model it is assumed that the set of basic signals is the binary alphabet and that the Sender combines elements of this set to communicate. Signals may be distorted in the communication process (the Receiver may misunderstand phonemes, emphasize the

[6] Suppose that R plays according to the mixed strategy $\alpha = (\alpha_1, \ldots, \alpha_{|A|})$ that assigns probability α_t to action a_t, then the payoffs obtained by both players are $\sum_{\omega_j \in \Omega} q_j \alpha M_j$. Notice that, since this expression is linear on α, then the optimal election of the probabilities of α (from the viewpoint of R) corresponds to a pure strategy (the one corresponding to the vertex j^* such that $j^* = \arg\max\{q_j M_j\}_j$). Therefore, the most that players can get without any communication is the $\max\{q_j M_j\}_j$. On the other hand, if noiseless communication were possible, then both players would achieve at each state ω_j the corresponding payoff M_j.

wrong part of the message, among others). This distortion or interference is known as *noise*. The noise is such that while messages are always received by the Receiver, they may differ from those sent by the Sender. We model this noise by assuming that each signal can randomly be mapped to the whole set of possible signals.

We follow a unifying approach to this noisy information transmission and consider that agents communicate through a *discrete noisy channel:* a system consisting of input and output alphabets, and a probability transition matrix[7]. Formally, a noisy binary channel υ is defined by:

- Two sets $X = Y = \{0, 1\}$ as the input and the output basic signal sets respectively
- A transition probability p: an input signal $s \in X$ is transformed by the channel in an output signal $r \in Y$ with a probability $p(r|s)$. Let ε_l be the probability of a mistransmission of input signal l, then since the channel is binary $p(1|0) = \varepsilon_0$ and $p(0|1) = \varepsilon_1$ and the noisy communication channel is denoted by $\upsilon(\varepsilon_0, \varepsilon_1)$.

A procedure of meaning inference allows, in general, a partial probabilistic recovering of the original meaning associated to the prototype sent by the Sender. However, in some specific situations, communication episodes are not informative and no meaning at all can be recovered from the received signal. For instance, if the Receiver cannot probabilistically decide that two signals are different, and every signal is essentially the same as the next, then there is no way a meaning (an action) can emerge. In this section we define what is an informative noisy channel. The key point is that the 'informativeness' of the channel is not related to the probability of properly understanding each basic signal 0 and 1, but the relation between these two probabilities, i. e. to the probability of discriminating between input basic signals once an output basic signal is observed.

Let s be an input signal belonging to some space X, and r a realized output signal in space Y. Let $p(s)$ be the a prior probability that signal s and is delivered through the channel. The channel transforms s into r according to $p(r|s)$. From the observed r, any input signal s is updated by Bayes' rule yielding the posterior probability of each s given r. Given two output signals r and $\widehat{r}$ and two input signals s and $\widehat{s}$, it is said that r is *more favorable than* $\widehat{r}$ for s, whenever the posterior odds of inputs s and $\widehat{s}$ given the output r are at least as high as those of inputs s and $\widehat{s}$ given the output $\widehat{r}$. Therefore, the noisy channel $\upsilon(\varepsilon_0, \varepsilon_1)$ is *informative* whenever for any realized output signal r and any pair of input signals s, $\widehat{s}$: $\frac{p(s|r)}{p(\widehat{s}|r)} = \frac{p(r|s)p(s)}{p(r|\widehat{s})p(\widehat{s})} \neq \frac{p(s)}{p(\widehat{s})}$.

[7] The introduction of noise into a defined channel is well understood in another strand of literature. Such information transmission has been mainly tackled by Information Theory tools. Traditional Information Theory, pioneered by Shannon (1948) abstracts away from equilibrium queries, and focuses on the process of information transmission itself.

How informative is the channel and which output signals are more favorable than others are crucial to design the Sender's set of input sequences and the Receiver's decodification procedure. Applying the above concepts to channel $\upsilon(\varepsilon_0, \varepsilon_1)$, where $s \in \{0,1\}$ and $r \in \{0,1\}$ it is obtained[8]:

Lemma 1 *If $\varepsilon_0 + \varepsilon_1 \neq 1$, then $\upsilon(\varepsilon_0, \varepsilon_1)$ is informative. Moreover:*

- *If $\varepsilon_0 + \varepsilon_1 < 1$, then for input signal 0, output signal 0 is more favorable than output signal 1, and for input signal 1, output signal 1 is more favorable than output signal 0.*

- *If $\varepsilon_0 + \varepsilon_1 > 1$, then for input signal 1 output signal 0 is more favorable than output signal 1, and for input signal 0, output signal 1 is more favorable than output signal 0.*

The proof is given in Appendix A.

Communication goes on for n periods. It is also assumed that the channel is memoryless, i.e., the probability distribution of the output depends only on the input at that time and is conditionally independent of previous channel inputs or outputs. Then, given a communication length n, the Sender utters to the channel an input sequence, $x \in X^n = \{0,1\}^n$, which is a concatenation of basic binary signals and the Receiver hears an output sequence, $y \in Y^n = \{0,1\}^n$, which is an independent probabilistic transformation of the signal string. Thus Γ is extended by a pre-play phase of communication where the Sender uses n times the channel $\upsilon(\varepsilon_0, \varepsilon_1)$. Let Γ_υ^n denote this extended communication game, where after the communication episode S chooses an action and the uninformed player R associates an action (infers a meaning from y and the game context) to the received sequence and payoffs are realized.

Since the Sender encodes the meanings in a set of public signals, his strategy in Γ_υ^n is a rule suggesting the message to be sent at each ω_j: a sequence $\sigma_j^S \in X^n$ sent

[8] Noisy channel are only characterized by their levels of aggregate noise. Firstly, if channel $\upsilon(\varepsilon_0, \varepsilon_1)$ is *informative with low levels of aggregate noise* ($\varepsilon_0 + \varepsilon_1 < 1$) and output signal 0 is observed, then it will be more likely that it comes from input signal 0 than from input signal 1 and viceversa. Thus, the matching between the output and the input signal yields a more accurate posterior odds. Notice that for symmetric channels, i.e. $\varepsilon_0 = \varepsilon_1 = \varepsilon$, this condition implies that the misunderstandings are not too high, i.e. $\varepsilon < \frac{1}{2}$. Nevertheless, $\varepsilon_0 = 0.2$ and $\varepsilon_1 = 0.7$ in asymmetric channels also fulfil the condition. Secondly, when channel $\upsilon(\varepsilon_0, \varepsilon_1)$ is *informative with high levels of aggregate noise* ($\varepsilon_0 + \varepsilon_1 > 1$) and the output signal 1 is observed, then it will be more likely that it comes from input signal 0 than from input signal 1 and viceversa. Here, the unmatching between the output and the input signal yields a more accurate posterior odds. For symmetric channels the above condition is equivalent to $\varepsilon > \frac{1}{2}$, or for instance to $\varepsilon_0 = 0.3$ and $\varepsilon_1 = 0.9$ for asymmetric ones. Finally, when $\varepsilon_0 + \varepsilon_1 = 1$, there is no way to discriminate between input signals once an output signal is observed and the channel is not-informative. This happens when $\varepsilon = \frac{1}{2}$ in symmetric channels, or, for instance, when $\varepsilon_0 = 0.1$ and $\varepsilon_1 = 0.9$ in asymmetric ones.

by S given that the true state of nature is ω_j. Each sequence σ_j^S is called a *standard prototype*. The set of standard prototypes $\{\sigma_j^S\}_j$ is the *corpus.*

A strategy of R is a 2^n-tuple $\{\sigma_y^R\}_y$, where σ_y^R specifies an action choice as a response to the realized output sequence $y \in \{0,1\}^n$. The action associated by R to each sequence, jointly with the context, is the *meaning decodification* . Then, a Receiver's strategy is the inference of a meaning for any sequence in the language, even for those not included in the corpus. An univocal construction of meanings partitions the set Y^n into $|\Omega|$ subsets, each of them bringing together all the sequences whose meaning is an action $\widehat{a}_j$, $j = 1,.., |\Omega|$. Each subset of this partition represents the *pragmatic variations* associated to a particular standard prototype.

A *Language* was defined as the pair corpus and pragmatic variations: the way in which the Sender transmits the meanings and that of the Receiver's understandings. An equilibrium Language is that for which the players' strategies are a best response to each other.

Expected payoffs in Γ_v^n are defined in the usual way. Let the tuple of the Sender's payoffs be denoted by $\{\pi_j^S\}_j = \{\pi_j^S(\sigma_j^S, \{\sigma_y^R\}_y)\}_j$, where for each ω_j,

$$\pi_j^S = \pi_j^S(\sigma_j^S, \{\sigma_y^R\}_y) = \sum_{y \in Y^n} p(y|\sigma_j^S) u(\sigma_y^R, \omega_j)$$

and where $p(y|\sigma_j^S)$ is the Sender's probability about the realization of the output sequence $y \in \{0,1\}^n$ conditional on having sent sequence σ_j^S in state ω_j.

Let the tuple of the Receiver's payoffs be denoted by $\{\pi_y^R\}_y = \{\pi_y^R(\{\sigma_j^S\}_j, \sigma_y^R)\}_y$, where for each output sequence $y \in \{0,1\}^n$,

$$\pi_y^R = \pi_y^R(\{\sigma_j^S\}_j, \sigma_y^R) = \sum_{j=1}^{|\Omega|} p(\sigma_j^S|y) u(\sigma_y^R, \omega_j)$$

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

A pure strategy perfect Bayesian equilibrium of the communication game is a pair of tuples $(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_y^R\}_y)$, i. e. a Sender's corpus and a Receiver's sets of pragmatic variations, and a set of probabilities $\{p(\sigma_j^S|y)\}_j$ for the Receiver such that for each ω_j, and for any other strategy $\widetilde{\sigma}_j^S$ of the Sender,

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

and for each $y \in \{0,1\}^n$ and for any other Receiver's strategy $\widetilde{\sigma}_y^R$,

$\widehat{\pi}_y^R = \pi_y^R(\{\widehat{\sigma}_j^S\}_j, \widehat{\sigma}_y^R) \geq \pi_y^R(\{\widehat{\sigma}_j^S\}_j, \widetilde{\sigma}_y^R)$ where by Bayes rule each $p(\sigma_j^S|y)$ is given by: $p(\sigma_j^S|y) = \frac{p(y|\sigma_j^S)p(\sigma_j^S)}{p(y)}$.

The ex-ante payoffs of this communication game are given by

$$\Pi^S(\{\sigma_j^S\}_j, \{\sigma_y^R\}_y) = \sum_{j=1}^{|\Omega|} q_j \pi_j^S(\sigma_j^S, \{\sigma_y^R\}_y) = \sum_{j=1}^{|\Omega|} q_j \sum_{y \in Y^n} p(y|\sigma_j^S) u(\sigma_y^R, \omega_j)$$

for the Sender an those of the Receiver are defined by,

$$\Pi^R(\{\sigma_j^S\}_j, \{\sigma_y^R\}_y) = \sum_{y \in Y^n} p(y) \pi_y^R(\{\sigma_j^S\}_j, \sigma_y^R) = \sum_{y \in Y^n} p(y) \sum_{j=1}^{|\Omega|} p(\sigma_j^S|y) u(\sigma_y^R, \omega_j)$$

Notice that $\Pi^S(\{\sigma_j^S\}_j, \{\sigma_y^R\}_y) = \Pi^R(\{\sigma_j^S\}_j, \{\sigma_y^R\}_y)$, since Γ is symmetric and $\sum_{j=1}^{|\Omega|} \sum_{y \in Y^n} q_j p(y|\sigma_j^S) = \sum_{y \in Y^n} \sum_{j=1}^{|\Omega|} p(y) p(\sigma_j^S|y)$. Denote this common ex-ante payoffs by Π_v.

In order to distinguish among sequences, a distance function among them has to be defined. A natural and intuitive function is the *Hamming distance*. Formally, consider two n-dimensional sequence $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$. Let I stands for the indicator function. The Hamming distance between two sequences x, y, denoted $h(x,y)$, is defined as $h(x,y) = \sum_{t=1}^n I_{x^t \neq y^t}$.

Suppose that $n = m\ |\Omega|$. For $1 \leq l \leq |\Omega|$, the $l-$block of length m of $x = (x^1, \ldots, x^n)$ is the subsequence $(x^{lm+1}, \ldots, x^{(l+1)m})$. The Hamming distance in the l-th block between x and y, denoted by $h_l(x,y)$ is equal to the Hamming distance between the l-th block of x and l-th block of y. Formally,

$$h_l(x,y) = h((x^{lm+1}, \ldots, x^{(l+1)m}), (y^{lm+1}, \ldots, y^{(l+1)m})) = \sum_{t=1}^{m} I_{x^{lm+t} \neq y^{lm+t}}$$

By additivity, the Hamming distance between x and y coincides with the sum of the Hamming distances of the $|\Omega|-$blocks between x and y: $h(x,y) = \sum_{l=1}^{|\Omega|} h_l(x,y)$.

3 Pure equilibrium strategies under informative noisy channels.

Our main finding shows the existence of a *pragmatic* equilibrium Language, with a *universal* structure or grammar. In fact, we show how to construct such a Language. Language is pragmatic in the sense that the Receiver' decoding rule depends on the *context*, i.e, on the payoffs and on the initial probability distribution of the states of nature of Γ. Language has a *universal* grammar because the Sender's corpus coding rule does not depend on such specific parameters of Γ and can then be applied to any sender-receiver game. Both rules are a best response to each other, generating an equilibrim Language

In the noisy communication game Γ_v^n, the cardinality of the set of communication sequences exceeds that of the set of states of nature. Then, given the set of basic signals $\{0,1\}$ and n, the Sender constructs a dictionary or "*corpus*" of sequences from $\{0,1\}^n$, the underlying set of meanings, by selecting $|\Omega|$ of them, one for each state of nature. Each selected sequence is a *standard prototype.* For each realized state of nature, the Sender utters to the channel a standard prototype. A Sender's pure strategy is then the standard prototype to be sent at each ω_j. Recall that $\{\sigma_j^S\}_j$ is the set of the Sender's pure strategies. We propose pure strategies that divide any input sequence $(x_j^1, \ldots, x_j^n)$ in $|\Omega|$ blocks of length m in such a way that all blocks but the j-th consist of repetitions of signal 1 and the j-th block is composed of m repetitions of signal 0.

For each prototype, the noise induces any output sequence in $Y^n = \{0,1\}^n$, say y. Once y is observed, the Receiver chooses her best response. Her choice is based on the following meaning inference procedure: given the noisy information transmission, she partitions the set of all possible received sequences $\{0,1\}^n$ in a collection of subsets which are called "*the pragmatic variation classes*", denoted by $\{\sigma_y^R\}_y$. Each pragmatic variation class is associated to a particular standard prototype and hence to a particular action. Therefore, the Receiver will play the action dictated by the pragmatic variation including output sequence y. Since the Receiver maps the observed output sequence into prototypes, his best reply to any pure Sender's strategy induces a categorization[9] of the output space, $Y^n = \{0,1\}^n$, around the standard protoypes. Thus, at equilibrium the output space is partitioned in a finite number of sets.

The main result states that the transmission of the corresponding standard prototype (for a given state of nature) by the Sender, and the choice of the action suggested by the classes of pragmatic variations (for a realized output sequence) by the Receiver are a pure strategy Bayesian Nash equilibrium of Γ_v^n. Firstly, we present the Theorem, proven in Appendix A, and then we show how to construct such strategies.

Theorem 1 *There exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, the pair of tuples $(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_y^R\}_y)$ and the set of probabilities $\{p(\sigma_j^S|y)\}_j$, where* $p(\widehat{\sigma}_j^S|y) = \frac{p(y|\widehat{\sigma}_j^S)p(\widehat{\sigma}_j^S)}{p(y)}$,

[9] Categorization is the way by which a set of entities, identified with some finite dimensional Euclidean space, is partitioned into a finite number of categories. Categories are sets of entities to which we react in an identical or similar way. Partitions of the space that are generated by a set of center points are well known as Voronoi diagrams or Dirichlet tessellations. In its simplest form, a Voronoi diagram is a partition of some Euclidean space into a finite number of sets.

for each $\widehat{\sigma}_j^S$, $j = 1, \ldots, |\Omega|$, *is a pure strategy Bayesian Nash equilibrium of* Γ_υ^n.

The partition of the output space in the pragmatic variations of the standard prototypes is related to the work on categorization based on prototypes (see Azrieli and Lehrer, 2007, Jäger, 2007 and references herein). Azrieli and Lehrer(2007), suggest a categorization model based on prototypes. They formulate the notion of categorization by a geometrical notion of partitions of a finite dimensional Euclidean space. They study a special kind of such categorizations: those that are generated by *extended prototypes*.

Jäger (2007) analyzes the class of sender-receiver games, where the cardinality of the set of meanings exceeds the size of the set of signals by several orders of magnitude. Under these conditions, communication cannot be guaranteed to be perfect because the number of signals provides an upper bound to the number of meanings that can be communicated in a precise and unambiguous way. The players should still strive to maximize the *similarity* between the meaning that the sender wants to express and the interpretation that the receiver assigns to the transmitted signal. The similarity function is a two-place function that measures the similarity between two points of the meaning space and only depends on the Euclidean distance (the maximal similarity is 1 and each point is maximally similar to itself). The utility function is the expected similarity of the meanings and therefore it is inversely related to the distance between the meaning that the sender tries to communicate and the interpretation that the receiver assigns to the transmitted message. He shows that under the replicator dynamics, a strict equilibrium set (the static characterization of asymptotically stable sets of rest points for the asymmetric replicator dynamics) is such that for each receiver's pure strategy r, the inverse image of any s in the sender's best reply to r is consistent with the Voronoi tesselation of the meaning space that is induced by the image of r.

On the contrary, in our noisy communication game, the receiver's best reply to any pure sender's strategy induces a categorization of the underlying meaning space around the standard protoypes. Our similarity function is the Hammig distance between sequences. This partition is due to an inference process rather than to an evolutionary dynamics. Thus, noisy communication processes induce pragmatic categorizations of the meaning space. Unlike Jäger, the players' utility function reflects different payoffs under different standard prototypes, this meaning that it is the weighted (by expected payoffs) distance what matters for categorization. Our categorization takes then into account not only the pure similarity or distance to the standard prototypes but also the expected payoffs of game Γ.

In the sequel, we offer the construction of the pure equilibrium strategies for the Sender and the Receiver, i.e. $\{\widehat{\sigma}_j^S\}_j$ and $\{\widehat{\sigma}_y^R\}_y$, respectively, in Γ_υ^n when the noisy channel $\upsilon(\varepsilon_0, \varepsilon_1)$ is informative with low levels of aggregate noise, i.e., $\varepsilon_0 + \varepsilon_1 < 1$. The remaining case is similar and we omit it.

3.1 The corpus and the pragmatic variations of the standard prototypes.

One is tempted to look at Information Theory[10] to design the players' coding and decoding strategies in our noisy communication game Γ_v^n. More specifically, Coding Theory is concerned with the design of practical encoding and decoding systems to achieve reliable communication over a noisy channel. The cardinality of the set of messages is here the same than that of the original source vector. The general idea is that the encoding system introduces systematic redundancy into the transmitted message, while the decoding system uses this known redundancy to deduce from the received message both the original source vector and the noise introduced by the channel[11]. The basic Theorem of Information Theory is then the achievability of the channel capacity by a communication protocol (based on encoding and decoding rules) under the implicit assumption that the two communicating agents commit ex-ante to following a particular encoding and decoding strategies before the communication stage.

In game theoretical analysis players are required to take actions when they are called upon to do so, therefore given a common knowledge encoding rule and an output message, the receiver's equilibrium conditions summarizes to choosing the action corresponding to that state of nature for which expected payoffs are higher. Furthermore, the cardinality of the set of messages exceeds in our case that of the set of the original source. Thus, unlike Information Theory, the role of a decoding rule in our problem is not that of recovering a string potentially perturbed by the noise channel but instead that of inferring at equilibrium which of the actual $|\Omega|$ valid messages was actually sent through the channel. Since we are interested in encoding systems supporting equilibria, there is only one feasible decoding rule which is given by the '*best response*' decoding rule[12]. Obviuosly, not all coding and decoding rules from Information Theory can generate the conditions to satisfy the Nash equilibrium conditions. This is so even when players' strategies come from a well-established theory guaranteeing a good rate of information transmission (see, Hernandez, Urbano and Vila, 2009. By this reason the application of standard encoding systems, which are more efficient in terms of transmission rates for blocks

[10] Information Theory is concerned with the theoretical limitations and potential of noisy communication systems. It deals with the problem of transmitting a block of bits of a given length k over a binary noisy channel. To this end, the encoding rule is a rule that transforms each of the 2^k vectors of bits of some original source into a new vector of length $k + l$ to be sent over the channel. The extra l bits generate redundancies that will help the receiver to recover the original vector. The key question is then the analysis of the trade off between the probability or error and the information transmission rate that is achieved over the channel by the use of a specific coding system.

[11] There are mainly two different families of encoding rules for binary noisy channels: repetition codes and linear block codes (where the most known are the Hamming codes).

[12] Decoding rules associated to standard code systems as Hamming (7,4) or random codes are not, in general, best responses to the received string.

of k bits, may not be too apropriate when designing simple and universal encoding systems supporting equilibria.

The next step is to design a Sender's encoding[13] rule (the corpus of the standard prototypes) which is a best response to the Receiver's best response decoding. A coding rule for our problem is a rule assigning a string of n symbols in $\{0,1\}$ to each state of nature. We will use a variation of a repetition[14] code: *the block coding rule*. In the Appendix B, we offer a general characterization of the Receiver's best response decoding rule, for any feasible Sender's corpus. Obviuosly, it depends of the game parameters, the noise and the encoding parameters. To leave apart the encoding parameters, and focus on the game theoretical aspects of the problem, we construct an easy and *universal* encoding rule allowing a simple characterization (for instance, in terms of the Hamming distance) of the best response decoding rule, only depending on both the game and the noise parameters of any sender-receiver game with noisy communication. As it will become clear in the sequel the block coding rule will be independent of the game payoffs and of the initial probabilities of the sates of nature.

3.1.1 The Corpus: Block Coding

The Sender construct the grammar of the Language from the set of basic signals $\{0,1\}$ and the communication length. For a given length n, the set of possible utterances is then $\{0,1\}^n$. Since each state ω_j is associated with a receiver's optimal action $\widehat{a_j}$, then, the corpus consists of the $|\Omega|$ sequences in $\{0,1\}^n$ given by $\{\widehat{\sigma}_1^S, .., \widehat{\sigma}_{|\Omega|}^S\}$ and each of the $|\Omega|$ sequences $\widehat{\sigma}_j^S$ is the standard prototype encoding the meaning "take the action $\widehat{a_j}$". The Sender's pure strategy assigns to each state ω_j a tuple $\widehat{\sigma}_j^S = (x_j^1, \ldots, x_j^n)$. Assume that the number of states of nature $|\Omega|$ is a multiple[15] of n, i.e., there exists an integer m such that $n = m\ |\Omega|$.

Since many sequences in $\{0,1\}^n$ are possible, some grammar is needed to isolate structural regularities. In particular, our language grammar is based on a block structure which allows us to construct a corpus as follows: each $\widehat{\sigma}_j^S \in X^n = (x_j^1, \ldots, x_j^n)$ where the element

$$x_j^i = \begin{cases} 0 & if\ (j-1)m - 1 \le i \le jm \\ 1 & otherwise \end{cases}$$

[13]The framework in our paper is quite different from that of Coding Theory. In particular, the number of states of nature (the cardinality of Ω), from which one has to be transmitted, is fixed and usually small.

[14]Repetition codes generate redundancy by the repetition of every bit of the message a prearranged number of times. This family of codes can achieve arbitrarily small probability of error only by decreasing the rate of transmission. However, they are useful for many practical purposes as, for instance, when universality is required.

[15]If n is not a multiple of m, one may consider m as the integer part of $\frac{n}{[\Omega]}$. The remainder elements would be considered without meaning.

In other words, the input sequence $(x_j^1, \ldots, x_j^n)$ is divided in $|\Omega|$ blocks of length m in such a way that all blocks but the j-th consist of repetitions of signal 1 and the j-th block is composed of m repetitions of signal 0.

Each standard prototype sequence is mapped one-to-one into one of the Receiver's set of actions. Thus, the structure or grammar specifies that each prototype sequence is positionally arranged, that is, in blocks. A first property of this grammar is that prototypes have the maximal separation among them.

This property of the grammar is related to the way to compare any output sequence y to all prototypes. The block structure permits us to compare any sequence y block by block. Thus, the relevant information when comparing y with prototype $\widehat{\sigma}_l^S$ is only contained in the corresponding block l. Moreover, in this block l all the remaining prototypes give the same information with respect to σ_l^S. The following lemma formalizes this property, where the block Hamming distance between sequences is the measure distance.

Lemma 2 *For all* $k, k' = 1, 2, \ldots, |\Omega|,\ \ k \neq k'$ *we have that*

1. $h_l(\widehat{\sigma}_k^S, y) = h_l(\widehat{\sigma}_{k'}^S, y)$ *if* $k \neq l \neq k'$
2. $h_k(\widehat{\sigma}_k^S, y) + h_k(\widehat{\sigma}_{k'}^S, y) = m$ *if* $k \neq k'$

The players' strategies can be understood as a communication protocol. One of the most desired properties in communication protocol design is universality. The corpus satisfies this property since it does not depend on the specific parameters of Γ, that is, on the payoffs and the initial probability distribution of the states of nature.

In summary, the corpus consists of $|\Omega|$ different standard prototypes such that each prototype reserves m locations for the signal 0 and the remainder locations contain the signal 1. Thus, the Sender's pure strategy assigns to each state ω_j the standard prototype $\widehat{\sigma}_j^S = (1, 1, \ldots, 1, \ldots, \underbrace{0, 0, \ldots, 0}_{j}, \ldots, \underbrace{1, 1, \ldots, 1}_{|\Omega|})$.

3.1.2 The Receiver's best response: The pragmatic variations of the standard prototypes.

The Receiver has to take an action in Γ after hearing an output sequence y maximizing her expected payoffs. Equivalently, for each y she chooses the action $\widehat{a}_l(y)$ such that

$$\sum_{j=1}^{|\Omega|} p(\widehat{\sigma}_j^S)|y)u(\widehat{a}_l|\omega_j) \geq \sum_{j=1}^{|\Omega|} p(\widehat{\sigma}_j^S)|y)u(a_k|\omega_j),$$

for any other $k \neq l$ which, given both the linearity of the Receiver's payoff functions in probabilities $\{p(\sigma_l^S|y)\}_l$, $l = 1, ..., |\Omega|$ and the matrix payoffs, is equal to, $p(\widehat{\sigma}_l^S|y)M_l \geq p(\widehat{\sigma}_k^S|y)M_k$, or $\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} \geq \frac{M_k}{M_l}$.

To solve this problem the Receiver computes the corresponding $p(\widehat{\sigma}_j^S|y)$. By Bayes' rule, such conditional probability is $p(\sigma_j^S|y) = \frac{p(y|\sigma_j^S)p(\sigma_j^S)}{p(y)}$ and the likelihood ratio of messages $\widehat{\sigma}_j^S$ and $\widehat{\sigma}_k^S$, conditional to the observed y is

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{\frac{p(y|\widehat{\sigma}_l^S)p(\widehat{\sigma}_l^S)}{p(y)}}{\frac{p(y|\widehat{\sigma}_k^S)p(\widehat{\sigma}_k^S)}{p(y)}} = \frac{q_l}{q_k}\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)}$$

where $p(y|\widehat{\sigma}_j^S)$ is given by the channel's errors probabilities and by the Sender's standard prototypes.

The next proposition states that given a noisy communication channel $\upsilon = \{\varepsilon_0, \varepsilon_1\}$ and the set of standard prototypes, the likelihood ratio of input signals $\widehat{\sigma}_j^S$ and $\widehat{\sigma}_k^S$ can be written in terms of the noisy parameters and the block hamming distance between the output signal y and the specific Sender's strategies. In Appendix A some easy but cumbersome calculations show that.

Proposition 1 *For all $k, l = 1, \ldots, |\Omega|$, $k \neq l$ and for all $y \in Y$,*

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{q_l}{q_k}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m} \tag{1}$$

The Receiver has to infer a meaning (to take an action) from the received sequence. To do so she has to assign the received sequence to one of the standard prototypes. This assignment is based on both the number of different elements (errors) that each two standard prototype sequences $\widehat{\sigma}_l^S$ and $\widehat{\sigma}_k^S$ may have with respect to the observed output sequence y and on the ratio of expected payoffs $\frac{M_k q_k}{M_l q_l}$. More precisely, the Receiver's pure equilibrium strategy generates a partition of output set Y^n based on both the above likelihood ratio and expected payoffs.

For each $l \in \{1, \ldots, |\Omega|\}$, compute first the parameters $\{C_{lk}\}_{l \neq k}$ as: $C_{lk} = \frac{\ln \frac{M_k q_k}{M_l q_l}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m$, and denoted "*vicinity bounds*", where we take the integer approximation of the numbers. There are $|\Omega| \times (|\Omega| - 1)$ of such parameters C_{lk} that can be arranged in the following matrix,

$$\begin{pmatrix} * & C_{21} & C_{31} & \cdots & C_{l1} & \cdots & C_{|\Omega|1} \\ C_{12} & * & C_{32} & \cdots & C_{l2} & \cdots & C_{|\Omega|2} \\ C_{13} & C_{23} & * & \cdots & C_{l3} & \cdots & C_{|\Omega|3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ C_{1l} & C_{2l} & C_{3l} & \cdots & * & \cdots & C_{|\Omega|l} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ C_{1|\Omega|} & C_{2|\Omega|} & C_{3|\Omega|} & \cdots & C_{l|\Omega|} & \cdots & * \end{pmatrix}$$

where each column gives the constraints defining subsets of Y^n and with typical element C_{lk}. This parameter is an upper bound on the distance between blocks l and k in y and the corresponding ones in $\widehat{\sigma}_l^S$. Thus, C_{lk} bounds the number of permitted mistakes to ensure that output sequence y comes from $\widehat{\sigma}_l^S$ instead of coming from $\widehat{\sigma}_k^S$.

The properties[16] of the vicinity bounds C_{lk} are the key to characterize the Receiver's strategy. Namely, the sign of the C_{lk} establishes when an action will be played. Secondly, the pair C_{lk} and C_{kl} determines the condition under which a partition is generated. Finally, parameters C_{lk} are an increasing function of n. Formally,

1. The expression $\ln(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1})$ is negative for informative channels with low levels of aggregate noise and since m is the integer part of $\frac{n}{|\Omega|}$ we may have negative C_{lk}. This will give rise to either a degenerate partition or, in case that a complete column is negative, to the not playing at all the corresponding action. We will come back to this case when tackling efficiency.

2. Since $C_{lk} = \frac{\ln \frac{M_k q_k}{M_l q_l}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m$ and $C_{kl} = \frac{\ln \frac{M_l q_l}{M_k q_k}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m$, and both are integer approximations, then either $C_{lk}+C_{kl} = 2m-1$ whenever $q_l M_l \neq q_k M_k$ or $C_{lk}+C_{kl} = 2m$ whenever $q_l M_l = q_k M_k$.

3. Let us specify the size of the standard prototype block, m, in parameters $\{C_{lk}\}_{l,k}$ by denoting any of them as C_{lk}^m. If the length of the input sequence increases up to $n+|\Omega|$, then $m = \frac{n+|\Omega|}{|\Omega|} = m+1$, and $C_{lk}^{m+1} = \frac{\ln \frac{M_k q_k}{M_l q_l}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m + 1$, and thus, $C_{lk}^{m+1} = C_{lk}^m + 1$.

Output set Y^n is next partitioned in subsets Y_l with $l = \{1, \dots, |\Omega|\}$, the pragmatic variation of σ_l^S, whose specific expression depends on the noisy nature of the communication channel, the odds of expected payoff and the length of the communication. To define each subset Y_l, we divide the set of states of nature in states

[16] Sets Y_l, $l = \{1, \dots, |\Omega|\}$ do not merely recover the information sent by the Sender. Notice that by Bayesian updating, whenever input sequence σ_l^S is more likely than input sequence σ_k^S for any observed output signal y, then the conditional probability ratio $\frac{p(\sigma_l^S|y)}{p(\sigma_k^S|y)}$ is bigger than 1. Take for instance $\varepsilon_0 + \varepsilon_1 < 1$ and let $\widehat{Y}_l$ denote all output sequences $y \in Y^n$ such that σ_l^S is more likely than σ_k^S, then by expression (2) of Proposition 1, this implies that

$$\widehat{Y}_l = \{y \in Y^n | h_l(\sigma_l^S, y)) + h_k(\sigma_l^S, y)) < \frac{\ln \frac{q_k}{q_l}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m, \text{ for all } k \neq l, \text{ and } k = \{1, \dots, |\Omega|\}\}$$

which is different from the above partition Y_l. Moreover when input sequences are uniformly distributed, i.e., $q_l = q_k$, for all $k \neq l$, and $k = \{1, \dots, |\Omega|\}$, then

$$\widehat{Y}_l = \{y \in Y^n | h_l(\sigma_l^S, y)) + h_k(\sigma_l^S, y)) < m, \text{ for all } k \neq l, \text{ and } k = \{1, \dots, |\Omega|\}\}$$

k, different from states l in expected payoffs, that is, $q_l M_l \neq q_k M_k$, and the other states where their expected payoffs coincide. In this (symmetric) later case, a rule to break tyings is needed. Our rule is the same for every pragmatic variation, coincides with the length of the block and is independent of both the noise and expected payoffs payoffs. Let $\tilde{\Omega}_l = \{k \in \{1, \ldots, |\Omega|\}$ such that $q_l M_l \neq q_k M_k\}$. Set $\tilde{\Omega}_l$ could be empty when, for instance, Γ had the same payoffs at each state and the priors were uniformly distributed.

As already said, we are considering informative channel with low levels of aggregate noise, $\varepsilon_0 + \varepsilon_1 < 1$. Here, the matching between the output and the input signal yields a more accurate posterior odds. Then, the elements of the partition are determined by a vicinity bound by above on the number of the permitted block hamming distances (errors) between a standard prototype $\widehat{\sigma}_l^S$ and the realized output sequence y.

$$
\begin{aligned}
Y_l = \{y \in Y^n | \quad & h_l(\widehat{\sigma}_l^S, y)) + h_k(\widehat{\sigma}_l^S, y)) \leq C_{lk}, \text{ for all } k \in \tilde{\Omega}_l \\
& h_l(\widehat{\sigma}_l^S, y)) + h_{k'}(\widehat{\sigma}_l^S, y)) \leq C_{lk'}, \text{ for all } k' \epsilon \notin \tilde{\Omega}_l \ , k' < l \\
& h_l(\widehat{\sigma}_l^S, y)) + h_{k'}(\widehat{\sigma}_l^S, y)) < C_{lk'}, \text{ for all } k' \notin \tilde{\Omega}_l \ , k' > l\}
\end{aligned}
$$

Finally, for each realized y, the Receiver's pure equilibrium strategy is

$$\widehat{\sigma}_y^R = \widehat{a}_l \Leftrightarrow y \in Y_l$$

Given the above description of Y_l, and as shown in the Appendix A, sets Y_l, $l = \{1, \ldots, |\Omega|\}$, are a true partition of Y^n. Therefore, the Receiver's best reply to any pure Sender's strategy induces a categorization of the potential meaning space, $Y^n = \{0,1\}^n$, around the standard protoypes.

The process to construct the specific sequences belonging to each Y_l is a little cumbersome but the following example nicely illustrate the whole construction.

4 Examples

Example 1. Consider the incomplete information two-player game Γ with three states of nature where nature chooses ω_j, $j = 1, 2, 3$ according to law $q = (q_1, q_2, q_3) = (0.5, 0.25, 0.25)$. The set of actions for player R is $A = \{a_1, a_2, a_3\}$, and payoffs for the three states of nature are $M_1 = 1$, $M_2 = 7$ and $M_3 = 43$, or in matrix form:

$$
\begin{array}{c}
\\ \omega_1 \\ \omega_2 \\ \omega_3
\end{array}
\begin{array}{c}
\begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\
\left(\begin{array}{ccc}
(1,1) & (0,0) & (0,0) \\
(0,0) & (7,7) & (0,0) \\
(0,0) & (0,0) & (43,43)
\end{array}\right)
\end{array}
$$

Suppose that the players communicate through the noisy channel $\upsilon(\varepsilon_0, \varepsilon_1)$ with associated transition probabilities $p(1 \mid 0) = \varepsilon_0 = 0.1$ and $p(0 \mid 1) = \varepsilon_1 = 0.6$, and that players can only communicate up to n times. Thus, the communication channel is $\upsilon = \{0.1,\, 0.6\}$ and Γ_υ^n is the associated extended communication game.

Suppose that $n = 6$, then for each state ω_j, $j = 1, 2, 3$, the Sender divides the standard prototype sequences $(x_j^1, \ldots, x_j^6) \in X^6$, in 3 blocks of length $m = \frac{n}{|\Omega|} = 2$, where the j-th block, consists of two consecutive 0's and the other blocks of two consecutive 1's. Thus, then the corpus consists of the Sender's 3-tuple of standard prototypes: $\widehat{\sigma}_1^S = 001111$, $\widehat{\sigma}_2^S = 110011$ and $\widehat{\sigma}_3^S = 111100$.

To construct the matrix of parameters C_{lk}, the Receiver considers first the matrix of elements

$$\frac{\ln \frac{M_k q_k}{M_l q_l}}{\ln \frac{\varepsilon_0}{1-\varepsilon_0} \frac{\varepsilon_1}{1-\varepsilon_1}} = C_{lk} - m$$

$$\begin{pmatrix} * & 0.70 & 1.71 \\ -0.70 & * & 1.01 \\ -1.71 & -1.01 & * \end{pmatrix}$$

Next, the output set $Y = \{0, 1\}^6$ is partitioned by the Receiver in subsets $Y_l = \{y | h_l(\widehat{\sigma}_l^S, y)) + h_k(\widehat{\sigma}_l^S, y)) \leq C_{lk}, \forall k, l = 1, 2, k \neq l\}$, where $\{Y_1, Y_2, Y_3\}$ are defined by the above matrix as follows:

$$\begin{aligned}
Y_1 = \quad & \{y \in \{0,1\}^6 \quad |h_1(001111, y) + h_2(001111, y) \leq 1 = C_{12} \\
& \qquad\qquad\qquad |h_1(001111, y) + h_3(001111, y) = 0 = C_{13}\ \} \\
Y_2 = \quad & \{y \in \{0,1\}^6 \quad |h_2(110011, y) + h_1(110011, y) \leq 2 = C_{21} \\
& \qquad\qquad\qquad |h_2(110011, y) + h_3(110011, y) = 0 = C_{23}\ \} \\
Y_3 = \quad & \{y \in \{0,1\}^6 \quad |h_3(111100, y) + h_1(111100, y) \leq 3 = C_{31} \\
& \qquad\qquad\qquad |h_3(111100, y) + h_2(111100, y) \leq 3 = C_{32}\ \}
\end{aligned}$$

Note that the worse expected payoffs in states 1 and 2 as compared to those in state 3, makes that both $C_{13} = C_{23} = 0$, i.e., no Hamming distance (mistake) will be permitted between the observed y and the standard prototype $\widehat{\sigma}_1^S$, if the Receiver has to assess that y comes from $\widehat{\sigma}_1^S$ instead of coming from $\widehat{\sigma}_3^S$, and similarly for the standard prototype $\widehat{\sigma}_2^S$.

How to construct the pragmatic variations around the standard prototypes? Take the minimum of the C_{lk}'s. In our example consider, for instance, Y_1, where this minimum is given by the Hamming distance of any output sequence y to blocks 1 and 3, i.e. $h_1(001111, y) + h_3(001111, y) = 0$. Parameter $C_{13} = 0$ implies that no error is permitted in blocks 1 *and* 3 together and hence these blocks in all sequences belonging to Y_1 have to be equal to the first and third blocks, respectively, of $\widehat{\sigma}_1^S = 001111$, i.e. sequences of the form $\{00 * *11\}$. This means that the pragmatic variation associated to σ_1^S does not permit any variation between blocks 1 and 3, thus fixing the elements of these two blocks in Y_1. Now, let us consider the elements of block 2 of Y_1, where $C_{12} = 1$. This implies that at most one error is permitted in blocks 1 *and* 2 together, but since $C_{13} = 0$, this error can only be in block 2. Hence block 2 in all sequences in Y_1 is composed of the sequences $\{(1,1), (0,1), (1,0)\}$. Thus, block 2 permits all the variations around $\widehat{\sigma}_1^S$. The distance asymmetries among blocks reflects the expected payoffs asymmetries of Γ. Similar reasoning will give us the set of sequences in $Y = \{0,1\}^6$ belonging to Y_2. Finally, notice that by lemma 2 and by the integer approximation $C_{32} + C_{23} = 2m - 1 \leq 2m$ and $C_{31} + C_{13} = 2m - 1 \leq 2m$, and the sequences belonging to Y_3 can be easily characterized.

Thus, the set[17] $Y = \{0,1\}^6$ is partitioned in three sets of sequences[18] Y_1, Y_2 and Y_3. where

$$Y_1 = \left\{ \begin{array}{ccc} \{(0,0) & (1,1) & (1,1)\} \\ \{(0,0) & (0,1) & (1,1)\} \\ \{(0,0) & (1,0) & (1,1)\} \end{array} \right\} \quad Y_2 = \left\{ \begin{array}{ccc} \{(1,1) & (0,0) & (1,1)\} \\ \{(1,0) & (0,0) & (1,1)\} \\ \{(0,1) & (0,0) & (1,1)\} \\ \{(0,0) & (0,0) & (1,1)\} \end{array} \right\}$$

and $Y_3 = \{0,1\}^6 - Y_1 - Y_2$.

Then, for each y the Receiver's pure equilibrium strategy is:

$$\widehat{\sigma}_y^R = \widehat{a}_j \Longleftrightarrow y \in Y_j,\ j = 1,2,3$$

and equilibrium expected payoff are:

$$\begin{aligned} \Pi^S(\{\widehat{\sigma}_j^S\}_j, \left\{\widehat{\sigma}_y^R\right\}_y) &= \Pi^R(\{\widehat{\sigma}_j^S\}_j, \left\{\widehat{\sigma}_y^R\right\}_y) = \Pi_v = \sum_{j=1}^{3} \sum_{y \in Y^n} q_j p(y|\widehat{\sigma}_j^S) u(\widehat{\sigma}^R(y), \omega_j) \\ &= \sum_{j=1}^{3} q_j M_j \sum_{y \in Y^n} q_j p(y|\widehat{\sigma}_j^S) = \sum_{j=1}^{3} q_j p(Y_j|\widehat{\sigma}_j^S) M_j \end{aligned}$$

---

[17] Notice also that the corresponding sets $\widehat{Y_l}$ only coming from Bayesian updating are by point 2) in the main properties of Y_l:

$$\begin{aligned} \widehat{Y_1} = \quad & \{y \in \{0,1\}^6 \quad |h_1(001111, y) + h_2(001111, y) \leq 2 = m \\ & \qquad\qquad\qquad |h_1(001111, y) + h_3(001111, y) \leq 2 = m \\ \widehat{Y_2} = \quad & \{y \in \{0,1\}^6 \quad |h_2(110011, y) + h_1(110011, y) \leq 1 = m \\ & \qquad\qquad\qquad |h_2(110011, y) + h_3(110011, y) \leq 0 < m \\ \widehat{Y_3} = \quad & \{y \in \{0,1\}^6 \quad |h_3(111100, y) + h_1(111100, y) \leq 1 = m \\ & \qquad\qquad\qquad |h_3(111100, y) + h_2(111100, y) \leq 0 = m \end{aligned}$$

Take for instance the sequence $y = 111111$, $y \in \widehat{Y_1}$ but $y \notin Y_1$.

[18] Notice that sets Y_1,Y_2 and Y_3 are a partition of Y. For in order a sequence of Y_3 is also in Y_1, it is needed a Hamming distance of $2m$. But the maximum distance is $2m - 1$.

where $p(Y_j|\widehat{\sigma}_j^S) = \sum_{y \in Y_j} p(y|\widehat{\sigma}_j^S)$. The value of these probabilities in our example are given by: $p(Y_1|\widehat{\sigma}_1^S) = 0.083$; $p(Y_2|\widehat{\sigma}_2^S) = 0.130$ and $p(Y_3|\widehat{\sigma}_3^S) = 0.994$, and then $\Pi_v = 10.96$. The ex-ante payoffs of noiseless communication are $\sum_{j=1}^{3} q_j M_j = 13$ and the maximum payoffs of the silent game are $q_3 M_3 = 10,75$. Then, Π_v is between them.

Example 2. A symmetric game: Consider the game of example 1, but with $q_1 M_1 = q_2 M_2 = q_3 M_3 = \frac{1}{3}$ instead. This game is equivalent to one with $q_1 = q_2 = q_3 = \frac{1}{3}$ and payoffs $M_1 = M_2 = M_3 = 1$. All the pragmatic variations are now,

$$\begin{array}{rll} Y_1 = & \{y \in \{0,1\}^6 & |h_1(001111, y) + h_2(001111, y) < 2 = m \\ & & |h_1(001111, y) + h_3(001111, y) < 2 = m \end{array}$$

$$\begin{array}{rll} Y_2 = & \{y \in \{0,1\}^6 & |h_2(110011, y) + h_1(110011, y) \leq 2 = m \\ & & |h_2(110011, y) + h_3(110011, y) < 2 = m \; \} \end{array}$$

$$\begin{array}{rll} Y_3 = & \{y \in \{0,1\}^6 & |h_3(111100, y) + h_1(111100, y) \leq 2 = m \\ & & |h_3(111100, y) + h_2(111100, y) \leq 2 = m \; \} \end{array}$$

In this symmetric case, the rule to break tyings is the same for every pragmatic variation, coincides with the length of the blocks and is independent of both the noise and payoffs.

Example 3. Let us come back to the particular case $|\Omega| = 2$, with $n = 4$, where an alternative corpus and standard could be $\widehat{\sigma}_1^S = 0000$ and $\widehat{\sigma}_2^S = 1111$. Further assume that $\varepsilon_0 = \varepsilon_1 = \varepsilon$.

Applying our best-reply reasoning to the above corpus, the Receiver's pragmatic variations for any received sequence y, when $q_1 M_1 \neq q_2 M_2$ are:

$$\begin{array}{rll} Y_1 = & \{y \in \{0,1\}^4 & |h_1(0000, y) + h_2(0000, y) \leq C_{12} \; \} \\ Y_2 = & \{y \in \{0,1\}^4 & |h_2(1111, y) + h_1(1111, y) \leq C_{21} \; \} \end{array}$$

and the Receiver's pure strategy is as before:

$$\widehat{\sigma}_y^R = \widehat{a}_j \Longleftrightarrow y \in Y_j, \; j = 1, 2$$

Notice that: 1) If Γ has symmetric expected payoffs, i.e., $q_1 M_1 = q_2 M_2$, then $C_{12} = C_{21} = m = \frac{n}{2}$, and the Receiver's best response when she hears a y is the well-known *majority rule*: playing $\widehat{a}_1$ whenever the number of 0ś is strictly greater than the number of 1ś and $\widehat{a}_2$ whenever the number of 1ś is greater than or equal to the number of 0ś. Nevertheless, when $q_1 M_1 \neq q_2 M_2$, the majority rule is not a best-response. 2) The set of pragmatic variations has the same structure than ours, bur referred to the alternative standard prototypes.

5 Efficiency Analysis

This section analyzes the power of our pragmatic Language as a coordination device under noisy communication. This analysis entails to first assessing its performance as a meaning inference model, i.e., to bound the size of the potential wrong inferences of meaning as a function of the parameters ε and n. Then, efficiency is analyzed by comparing, for each communication length n, how close ex-ante payoffs are to those of reliable communication, thus providing, for a given payoff-approximation parameter, the communication threshold length.

Let Γ_{υ_0} be the game where the Sender communicates the realized state of nature with no mistake, i.e., $\varepsilon_0 = \varepsilon_1 = 0$, and let Π_{υ_0} be the associated Nash equilibrium ex-ante payoffs where agents play the action pair with positive payoffs, at each state of nature,

$$\Pi_{\upsilon_0} = \sum_{j=1}^{|\Omega|} q_j M_j$$

Alternatively, the common ex-ante expected payoffs of our extended communication game Γ_υ^n were denoted by Π_υ. To stress the dependence of such payoffs on the communication length n, let us denote them as Π_υ^n. Then,

$$\Pi_\upsilon^n = \Pi_\upsilon^n(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_y^R\}_y) = \sum_{j=1}^{|\Omega|} q_j M_j p(Y_j^n|\widehat{\sigma}_j^S)$$

where $p(Y_j^n|\widehat{\sigma}_j^S) = \sum_{y \in Y_j^n} p(y|\widehat{\sigma}_j^S)$.

Then

$$\Pi_{\upsilon_0} - \Pi_\upsilon^n = \sum_{j=1}^{|\Omega|} q_j M_j (1 - p(Y_j^n|\widehat{\sigma}_j^S))$$

The difference between the above expected payoffs depends on probabilities $p(Y_j^n|\widehat{\sigma}_j^S)$. Each of this quantities measures the probability mass of the pragmatic variation of each standard prototype $\widehat{\sigma}_j^S$: the Receiver's probability of playing $\widehat{a}_j^R$ when the Sender utters $\widehat{\sigma}_j^S$ in a communication episode of length n. Thus, a probability $p(Y_j^n|\widehat{\sigma}_j^S)$ close to one means that in spite of initial misunderstandings, the Receiver is able to properly infer the Sender's meaning from $\widehat{\sigma}_j^S$.

The first finding is that our pragmatic Language performs (probabilistically) quite well as an inference meaning device, under noisy communication. To show this, we construct an upper bound on $1 - p(Y_j^n|\widehat{\sigma}_j^S) = p(Y^n - Y_j^n|\widehat{\sigma}_j^S)$, i.e. the probability of not inferring the action $\widehat{a}_j^R$ by the Receiver when the Sender utters sequence $\widehat{\sigma}_j^S$. By definition of each Y_j, this wrong inference takes place whenever some vicinity bounds are not fulfilled, i.e., whenever $h_j(\widehat{\sigma}_j^S, y) + h_k(\widehat{\sigma}_k^S, y) > C_{jk}$ for $k \in K$, where K is any non-empty subset in $\{1, 2, .., j-1, j+1, ...|\Omega|\}$. Then, we partition the event $Y^n - Y_j^n$ into a series of disjoint events E_K, where each of them is formed by the output sequences not satisfying the corresponding vicinity bounds C_{jk} for $k \in K$. In Appendix A it is proven that the probability of such events, $p(E_K|\widehat{\sigma}_j^S)$, and then $p(Y^n - Y_j^n|\widehat{\sigma}_j^S)$, can be written as a polynomial of the bigger noise parameter, say ε_1, and the smallest vicinity bound (which depends on n). The next Proposition states this result:

Proposition 2 *Given a noisy communication channel $\upsilon(\varepsilon_0, \varepsilon_1)$ with $\varepsilon_0 < \varepsilon_1$ and game Γ_υ^n, for any $n \geq |\Omega|$, then*

$$p(Y^n - Y_j^n|\widehat{\sigma}_j^S) = 1 - p(Y_j|\widehat{\sigma}_j^S) \leq \varepsilon_1^{c_j+1}\phi_j(\varepsilon_0, \varepsilon_1)$$

where $C_j = \min\{C_{jl}|l = 1, ..., |\Omega|; j \neq l\}$ and $\phi(\varepsilon_0, \varepsilon_1)$ is a function on ε_0 and ε_1 such that $0 \leq \phi_j(\varepsilon_0, \varepsilon_1) \leq 1$, for each j.

The above Proposition says that there is an (small) upper bound on the probability of wrong inferences. The formula precisely measures such a bound[19].

We turn next to the efficiency issue. For a fixed communication length n, a pair of equilibrium strategies $(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_y^R\}_y)$ is η-efficient if

$$\Pi_{\upsilon_0} - \Pi_\upsilon^n < \eta$$

For any $\eta > 0$, we offer a threshold length $\widehat{n}$ such that both the associated corpus and set of pragmatic variations support η-efficient equilibrium strategies.

By Proposition 2 and since $0 \leq \phi_j(\varepsilon_0, \varepsilon_1) \leq 1$, the difference between reliable and noisy communication expected payoffs is given by

$$\Pi_{\upsilon_0} - \Pi_\upsilon^n = \sum_{j=1}^{|\Omega|}(1 - p(Y_j^n|\widehat{\sigma}_j^S))q_j M_j = \sum_{j=1}^{|\Omega|}\varepsilon_1^{\widetilde{c}_j+1}\phi_j(\varepsilon_0, \varepsilon_1)q_j M_j \leq \sum_{j=1}^{|\Omega|}\varepsilon_1^{\widetilde{c}_j+1}q_j M_j$$

---

[19] This bound allow us to obtain asymptotic properties of the proposed equilibrium strategies. By the above proposition, $1 \geq p(Y_j|\widehat{\sigma}_j^S) \geq 1 - \varepsilon_1^{c_j^m+1}\phi_n(\varepsilon_0, \varepsilon_1)$, where c_j^m is a function of n, since $m = \frac{n}{|\Omega|}$ and polynomials $\phi_n(\varepsilon_0, \varepsilon_1)$ also depend on n. By the properties of the C_{lk} parameters, c_j^m grows by a unit whenever the block length $m = \frac{n}{|\Omega|}$ increases by one. Since $\varepsilon_0, \varepsilon_1 < 1$, then $\lim_{n\to\infty}\varepsilon_1^{c_j+1} = 0$ and $\lim_{n\to\infty}\phi_n(\varepsilon_0, \varepsilon_1)$ is a constant, hence:

$$\lim_{n\to\infty} p(Y_j^n|\widehat{\sigma}_j^S) \geq 1 - \lim_{n\to\infty}\varepsilon_1^{c_j^m+1}\phi_n(\varepsilon_0, \varepsilon_1) = 1$$

Therefore, the limit of $(\Pi_{\upsilon_0} - \{\Pi_\upsilon^n\}_n)$ is 0 when n goes to infinity, since this limit is $q_j M_j - \lim_{n\to\infty} q_j M_j\{p(Y_j^n|\widehat{\sigma}_j^S)\}_n$ and it is zero, for all $j = 1, \ldots, |\Omega|$, whenever $\lim_{n\to\infty}\{p(Y_j^n|\widehat{\sigma}_j^S)\}_n = 1$

where $\widetilde{c}_j = \min\{C_{jl} | l = 1, ..., |\Omega|; j \neq l\}$. Denote $\widetilde{c} = \min\{\widetilde{c}_j | j = 1, ..., |\Omega|\}$, and assuming, without loss of generality, that $q_1 M_1 \leq ... \leq q_{|\Omega|} M_{|\Omega|}$, then the vicinity bound $\widetilde{c}$ is

$$\widetilde{c} = \frac{\ln \frac{q_{|\Omega|} M_{|\Omega|}}{q_1 M_1}}{\ln \frac{\varepsilon_0 \varepsilon_1}{(1-\varepsilon_0)(1-\varepsilon_1)}} + \frac{n}{|\Omega|}$$

Therefore, to get an η-approximation,

$$\Pi_{v_0} - \Pi_v^n \leq \varepsilon_1^{\widetilde{c}+1} \sum_{j=1}^{|\Omega|} q_j M_j < \eta$$

Or, in other words,

$$\varepsilon_1^{\widetilde{c}+1} < \frac{\eta}{\sum_{j=1}^{|\Omega|} q_j M_j}$$

And, since $\ln \varepsilon_1 < 0$,

$$\widetilde{c} > \frac{1}{\ln \varepsilon_1} \ln \frac{\eta}{\sum_{j=1}^{|\Omega|} q_j M_j} > \frac{1}{\ln \varepsilon_1} \ln \frac{\eta}{\sum_{j=1}^{|\Omega|} q_j M_j} - 1$$

and $\widehat{n}$ is bounded by the expression

$$\widehat{n} > |\Omega| \left(\frac{1}{\ln \varepsilon_1} \ln \frac{\eta}{\sum_{j=1}^{|\Omega|} q_j M_j} - \frac{\ln \frac{q_{|\Omega|} M_{|\Omega|}}{q_1 M_1}}{\ln \frac{\varepsilon_0 \varepsilon_1}{(1-\varepsilon_0)(1-\varepsilon_1)}} \right)$$

The minimum length of the communication episode that allows η-efficiency depends on the relative approximation level, the biggest amount of noise and the maximum payoff-range. Notice that if the first term of the right hand side of the above equation were negligible, then the length $\widehat{n}$ would coincide with the minimal length guaranteeing a positive matrix of the C_{lk}'s. Therefore, the second term of the right hand side is a necessary condition to generate non-empty pragmatic variations for all the prototypes. The first term adds then the time needed to span such pragmatic variations in order to reduce the chances of misunderstandings and increase expected payoffs according to the η-efficiency.

Theorem 2 *Let $\eta > 0$, for any communication length $n \in [\widehat{n}, \infty)$,*

$$\Pi_{v_0} - \Pi_v^n < \eta$$

We would like to remark that although the corpus works quite efficiently in most of the cases, there may exist situations where communication is so short that some prototypes may be enable to generate a meaning. This create inefficiencies that could be easily avoided by a reassignment of the signals. More precisely, in the corpus construction, $\frac{n}{|\Omega|}$ out of the available n signals, those equal to 0, are used to distinguish each standard prototype from any other one. If there exists a state ω_j such that $C_{jk} < 0$ for any $k = 1, ..., |\Omega|$, $k \neq j$, then $Y_j = \varnothing$ and $p(Y_j | \sigma_j^S) = 0$. In this case, the Receiver's action $\widehat{a}_j$ will never be chosen and the $\frac{n}{|\Omega|}$ signals devoted to distinguish σ_j^S from the other prototypes are wasted. To avoid this inefficiency, the corpus is modified such that the 0' signals used in those prototypes sequences such that $p(Y_j | \sigma_j^S) = 0$ are reassigned to the other prototypes.

Example 4. Consider the incomplete information two-player game $\Gamma = \{\Gamma_1, \Gamma_2, \Gamma_3\}$, where nature chooses ω_j according to law $q = (q_1, q_2, q_3) = (0.1, 0.3, 0.6)$. Payoffs for the three states of nature are $M_1 = 3$, $M_2 = 20$ and $M_3 = 100$.

The matrix of parameters $C_{lk} - m$ is ,

$$\begin{pmatrix} * & 1.67 & 2.95 \\ -1.67 & * & 1.29 \\ -2.95 & -1.29 & * \end{pmatrix}$$

For $n = 3$, then $m = \frac{n}{|\Omega|} = 1$, and the sets of pragmatic variations Y_1 and Y_2 are empty, the corresponding actions $\widehat{a}_1$, $\widehat{a}_2$ will never be chosen and the communication game is equivalent to the silent game, where no-communication takes place. On the other hand, for $n \geq 9$, and then $m \geq 3$, all the pragmatic variations are non-empty and the Receiver's three actions will be played with ex-ante positive probability.

For the intermediate value of $n = 6$ $(m = 2)$, the matrix of vicinity bounds, C_{lk}, after the integer approximations, is:

$$\begin{pmatrix} * & 3 & 4 \\ 0 & * & 3 \\ -1 & 0 & * \end{pmatrix}$$

In this case the prototype $\sigma_1^S = 001111$ is enable to generate a meaning (the corresponding Receiver's action $\widehat{a}_1$) and the two 0-signals of the sequence devoted to distinguish it from the others are wasted. To avoid this inefficiency, players could act as if they were playing another (truncated) game, with only two states of nature ω_2 and ω_3, each of them taking place with probabilities $\widetilde{q}_2 = q_2 + q_1 \frac{q_2}{q_2+q_3}$ and $\widetilde{q}_3 = q_3 + q_1 \frac{q_3}{q_2+q_3}$, respectively. In this case the new corpus consists of the two standard prototypes $\widetilde{\sigma}_2^S = 000111$, $\widetilde{\sigma}_3^S = 111000$ and no signal is wasted now.

To formalize this idea, consider the game $\Gamma = \{\Gamma_1, ..., \Gamma_{|\Omega|}\}$, where Γ_j is chosen by nature with probability q_j, and its communication extension by adding n uses of the noisy channel v, denoted by Γ_v^n . Let as assume, without loss of generality, that

$p(Y_1|\sigma_1^S) \geq p(Y_2|\sigma_2^S) \geq ... \geq p(Y_{|\Omega|}|\ \sigma_{|\Omega|}^S)$ and let $j_0 = \min\{j = 1, ..., |\Omega|| p(Y_j|\sigma_j^S) > 0\}$.

Given Γ and Γ_v^n, define the auxiliary truncated game $\Gamma = \{\Gamma_1, ..., \Gamma_{j_0}\}$ where nature chooses state ω_j (and the the game Γ_j) with probability

$$\widetilde{q}_j = q_j + \frac{q_j}{\sum_{l=0}^{j_0} q_l} \sum_{k=j_0+1}^{|\Omega|} q_k \geq q_j$$

for $j = 1, \ldots, j_0$.

Let $\widetilde{\Gamma}_v^n$ the corresponding extended game and $\widetilde{\sigma}_j^S$, $j = 1, \ldots, j_0$, the standard prototypes of the new corpus. Since $p(Y_j|\widetilde{\sigma}_j^S) \geq p(Y_j|\sigma_j^S)$ for $j = 1, \ldots, j_0$ and $p(Y_j|\sigma_j^S) = 0$ for $j = j_0 + 1, \ldots, |\Omega|$, we have that

$$\widetilde{\Pi}_v^n = \sum_{j=1}^{j_0} \widetilde{q}_j M_j p(Y_j^n|\widetilde{\sigma}_j^S) \geq \sum_{j=1}^{j_0} q_j M_j p(Y_j^n|\sigma_j^S) = \sum_{j=1}^{|\Omega|} q_j M_j p(Y_j^n|\sigma_j^S) = \Pi_v^n$$

and no message is wasted trying to distinguish among actions that will never be chosen.

6 Concluding Remarks

We have shown that pragmatic Languages with a universal grammar are a powerful coordination device when there may exists communication misunderstandings. Reduced dictionaries[20], common knowledge simple grammars and standard prototypes help individuals to coordinate in spite of initial misunderstandings. This is accomplished by facilitating the inference of meaning and thus generating the pragmatic variations around each standard prototype. Our approach sheds light to the formation not only of target-oriented languages, but also to specific "organization" languages, professional languages, etc. .

When considering real life time-constraints, a language with structure based on different orderings of the enumerations turns out to be more useful for learning purposes rather than for meaning inference. Nevertheless, languages with universal grammars appear to have emerged because they ensure the successful transmission of languages themselves. The Chinese Language is an example of how the successful

[20] Nowak, Krakauer and Dress (1999) argue that, because of the noise, the fitness of a language cannot be increased arbitrarily by just adding more signals. On the contrary, the fitness can be increased by combining a small number of signal into words. This is called "phonemes" by linguists. Modern human languages have a limited number of phonemes: as reported by Nowak, Krakauer and Dress, all of 317 languages in the University of California Los Angeles Segment Inventory Database (UPSID) have between 11 and 141 phonemes, but 70% of these languages have between 20 and 37 phonemes.

transmission of information shapes some language characteristics. Spoken Chinese is distinguished by its high level of internal diversity (it is pragmatic and very local) though all spoken varieties of Chinese[21] are tonal and analytic; dictionaries are small with 6 vowels and 15 consonants and the grammar is compositional. On the contrary, written Chinese is highly complex: it comprises the written symbols used to represent spoken Chinese. Chinese characters do not constitute an alphabet or a compact syllabary; they are instead built up from simple parts representing objects or abstracts notions. There are around 47.035 ideograms or *hanzy*, but Chinese people do not manage more than 8000 of them.

One of the frequently asked questions in studies on language origins and evolution is how universal grammar structures in human languages could have emerged. One line of research assumes that such structures emerged from exploiting regularities found in protolanguages. Universal structures in language could have emerge when the learning examples do not cover the entire language (i.e., there was a bottleneck on the transmission of language). Other researchers have assumed that the ability to use syntax has evolved as a biological adaptation. In their seminal article which re-ignited much of recent burgeoning interest in language evolution, Pinker and Bloom (1990) argue persuasively that "a specialization for grammar evolved by a conventional neo-Darwinian process" (page 707), suggesting that human have evolved an innate, genetically specified module in the brain, which specifies a formal coding of the principles of Universal Grammar. These authors are firmly of the opinion that the selective advantage of the communicative function of language can explain the evolution of the language faculty itself. But, Chomsky (1988), perhaps somewhat surprisingly, given his introduction of the very idea of Universal Grammar, argues that the role of natural selection in language evolution is very limited. Much effort is in computer simulations of language evolution is been done to give more precise answers.

To conclude, we would like to call the attention about the way of precisely defining the notions of a language and a "common language", from an economic viewpoint. As stressed in Balinski and Laraki (2007,b) different models need different notions of both languages and common knowledge languages but some unifying rules are still lacking.

7 References

Azrieli and Lehrer, E. (2007): "Categorization"

Balinski, M, and R. Laraki (2007,a): "Election by Majority Judgement: Experimental Evidence", Cahier du Laboratoire d' Econometrie, École Polythecnique, n^o 2007-28.

[21] The standardized form of spoken Chinese is the Standard Mandarin.

Balinski, M, and R. Laraki (2007,b): "A theory of measuring, electing and ranking", *Proceedings of the National Academy of Sciences, U.S.A.,* vol. 104(2007), 8720-8725.

Batali, J. (1998): "Computational simulations of the emergence of grammar". In J.R. Hurford, M. Studdert-Kennedy, and C. Knight (editors), *Approaches to the evolution of language: Social and cognitive bases*, 405-426. Cambridge, Cambridge University Press.

Blume, A. (2000): "Coordination and Learning with a Partial Language", *Journal of Economic Theory* 95, 1-36.

Blume, A. (2005): "A Learning-Efficiency Explanation of Structure in Language", *Theory and Decision,* 57, 265-285.

Blume, A. and O. J. Board (2008): "Intentional Vagueness", mimeo.

Blume, A., O. J. Board and K. Kawamura (2007), "Noisy Talk", *Theoretical Economics*, Vol. 2, 395-440.

Chomsky, N. (1988): *Language and problems of knowledge*, Cambridge, MA. The MIT Press.

Crawford, V. and J. Sobel (1982): "Strategic Information Transmission", *Econometrica*, Vol. 50, 1431-1451.

Crémer, J., L. Garicano and A. Prat (2007): Language and the Theory of the Firm", *Quarterly Journal of Economics,* 122(1), 373-408.

Demichelis, S and J. Weibull (2008): "Language, meaning and games: A model of communication, coordination and evolution", *American Economic Review,* forthcoming.

Gärdenfors, P. (2000): *"Conceptual Spaces".* Cambridge, MA: The MIT Press.

Gärdenfors, P. (2002): "Cooperation and the evolution of symbolic communication", LUCS 91, ISSN 1101.

Grice, H.P. (1975): "Logic and Conversation". In P. Cole and J.L. Morgan (editors), *Syntax and semantics*, vol. 3, 41-58. New York. Academic Press.

Jäger, G. (2007):"The evolution of convex categories", *Linguist and Philos*, 30, 551-564. Published online: 26 March 2008.

Hernández,P., Urbano, A. and Vila J. (2009): "Nash Equilibrium and coding and decoding information transmission rules", mimeo.

Kirby, S. (2001): "Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity", *IEEE Journal of Evolutionary Computation*, 5(2), 102-110.

Kirby, S. (2002): "Learning, bottlenecks and the evolution of recursive syntax". In E. Bricoe, editor, *Linguistics evolution through language acquisition: Formal and Computational models*, 173-203, Cambridge, Cambridge University Press.

Kirby, S. (2007): "The evolution of the meaning-espace structure through iterated learning", In Lyon, C., Nehaniv, C., and Cangelosi, A., editors, *Emergence of Communication and Language*, 253-268, Spring-Verlang.

Koessler, F. (2000): "Common Knowledge and Interactive Behaviors: A Survey", *European Journal of Economic and Social Systems*, 14(3), 271-308.

Koessler, F. (2001): "Common Knowledge and Consensus with Nosy Communication", *Mathematical Social Sciences*, 42(2), 139-159.

Koessler, F. (2004): "Strategic Knowledge Sharing in Bayesian Games", *Games and Economic Behavior*, 48(2), 292-320.

Lipman, B., and D. Seppi (1995): "Robust Inference in Communication Games with Partial Peovability", *Journal of Economic Theory* 66, 370-405.

Nowak, M. A. and D. C. Krakauer (1999): "The evolution of language", *Proc. Natl. Acad. Sci. USA*, vol. 96, 8028-8033.

Nowak, M. A., D.C. Krakauer and A. Dress (1999): "An error limit for the evolution of language", *Proceedings of the Royal Society B: Biological Sciences*, 266, 2131-2136.

Pinker, S and P. Bloom (1990): "Natural language and natural selection", *Behavioral and Brain Sciences*, 13, 707-784.

Reiter, E. and S. Sripada (2002): "Human Variation and Lexical Choice", *Association for Computational Linguistics*, 28, 545-553.

Rubinstein, A. (1996): "Why are certain properties of binary relations relatively more common in natural language, *Econometrica*, Vol. 64, 343-355.

Rubinstein, A. (2000): *Economics and Language*, Cambridge University Press, Cambridge.

Selten, R.and M. Warglien (2007): "The emergence of simple languages in an experimental coordination game", *PNAS*, vol. 104 (18), 7361-7366.

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

Smith, A.D.M. (2003): *Evolving Communication through the inference of Meaning*, PhD thesis, University of Edinburg.

Vogt, P. (2000): *Lexicon grounding on mobile robots.* PhD Thesis, Vrije Universiteit, Brussels.

Vogt, P. (2005): "The emergence of compositional structures in perceptually grounded language games", *Artificial Intelligence* 167(1-2): 206-242.

8 Appendix A

Proof of Lemma 1: If $\varepsilon_0 + \varepsilon_1 < 1$, then $p(1|0) = \varepsilon_0 < 1 - \varepsilon_1 = p(1|1)$ and $p(0|1) = \varepsilon_1 < 1 - \varepsilon_0 = p(0|0)$. Clearly, the conditional probability of receiving a 0, when a 0 was sent is higher than the one of receiving a 0 when a 1 was sent, therefore, $\frac{p(0|0)}{p(0|1)}1$. And similarly for the conditional probability of receiving a 1, i.e., $\frac{p(1|1)}{p(1|0)} \neq 1$. Thus, information transmission is informative since $p(r|s) \neq p(r|\widehat{s})$ for any $r \in \{0,1\}$ and $s, \widehat{s} \in \{0,1\}$ and thus,

$$\frac{p(s|r)}{p(\widehat{s}|r)} = \frac{p(r|s)p(s)}{p(r|\widehat{s})p(\widehat{s})} \neq \frac{p(s)}{p(\widehat{s})}$$

Now, let $r = 0$ and $\widehat{r} = 1$, $s = 0$ and $\widehat{s} = 1$, then $p(r|s) = p(0|0) = 1 - \varepsilon_0$, $p(\widehat{r}|\widehat{s}) = p(1|1) = 1 - \varepsilon_1$, $p(r|\widehat{s}) = p(0|1) = \varepsilon_1$ and $p(\widehat{r}|s) = p(1|0) = \varepsilon_0$. Let us check that output signal 0 is more favorable than output signal 1, for input signal 0:

$$\frac{p(r|s)}{p(r|\widehat{s})} = \frac{p(0|0)}{p(0|1)} = \frac{(1-\varepsilon_0)}{\varepsilon_1} > \frac{\varepsilon_0}{(1-\varepsilon_1)} = \frac{p(1|0)}{p(1|1)} = \frac{p(\widehat{r}|s)}{p(\widehat{r}|\widehat{s})}$$

Similarly, letting now $r = 1$ and $\widehat{r} = 0$, $s = 1$ and $\widehat{s} = 0$, then $p(r|s) = p(1|1) = 1 - \varepsilon_1$, $p(\widehat{r}|\widehat{s}) = p(0|0) = 1 - \varepsilon_0, p(r|\widehat{s}) = p(1|0) = \varepsilon_0$ and $p(\widehat{r}|s) = p(0|1) = \varepsilon_1$. As above, since $(1 - \varepsilon_0)(1 - \varepsilon_1) > \varepsilon_1\varepsilon_0$, then $p(r|s)p(\widehat{r}|\widehat{s}) > p(r|\widehat{s})p(\widehat{r}|s)$ and output signal 1 is more favorable than output signal 0, for input signal 1.

If $\varepsilon_0 + \varepsilon_1 > 1$, then $p(1|0) = \varepsilon_0 > 1 - \varepsilon_1 = p(1|1)$ and $p(0|1) = \varepsilon_1 > 1 - \varepsilon_0 = p(0|0)$. In words, the conditional probability of receiving a 0, when a 0 was sent is lower than the one of receiving a 0 when a 1 was sent. And similarly for the conditional probability of receiving a 1. Information transmission is informative again since $p(r|s) \neq p(r|\widehat{s})$ for any $r \in \{0,1\}$ and $s, \widehat{s} \in \{0,1\}$, and then

$$\frac{p(s|r)}{p(\widehat{s}|r)} = \frac{p(r|s)p(s)}{p(r|\widehat{s})p(\widehat{s})} \neq \frac{p(s)}{p(\widehat{s})}$$

Moreover, let $r = 1$ and $\widehat{r} = 0$, $s = 0$ and $\widehat{s} = 1$ then output signal 1 is now more favorable than output signal 0, for input signal 0, since $p(r|s) = \varepsilon_0 > (1 - \varepsilon_1) = p(r|\widehat{s})$ and $p(\widehat{r}|\widehat{s}) = \varepsilon_1 > (1 - \varepsilon_0) = p(\widehat{r}|s)$ and then $\varepsilon_1\varepsilon_0 > (1 - \varepsilon_0)(1 - \varepsilon_1)$, or $p(r|s)p(\widehat{r}|\widehat{s}) \geq p(r|\widehat{s})p(\widehat{r}|s)$. For the same reason, taking now $r = 0$ and $\widehat{r} = 1$, $s = 1$ and $\widehat{s} = 0$, output signal 0 is now more favorable than output signal 1, for input signal 1.

Finally notice that when $\varepsilon_0 + \varepsilon_1 = 1$, then $p(1|0) = \varepsilon_0 = 1 - \varepsilon_1 = p(1|1)$ and $p(0|1) = \varepsilon_1 = 1 - \varepsilon_0 = p(0|0)$. Now input signals are not informative at all, since the conditional probability of receiving a 0, when a 0 was sent is equal to the one of receiving a 0 when a 1 was sent. And similarly for the conditional probability of receiving a 1. In other words, $p(r|s) = p(r|\widehat{s})$, for any $r \in \{0,1\}$ and s, $\widehat{s} \in \{0,1\}$ and then information transmission is useless since

$$\frac{p(s|r)}{p(\widehat{s}|r)} = \frac{p(r|s)p(s)}{p(r|\widehat{s})p(\widehat{s})} = \frac{p(s)}{p(\widehat{s})}$$

Finally, if $\varepsilon_1 = (1 - \varepsilon_0)$ and $\varepsilon_0 = (1 - \varepsilon_1)$, then for any $s \in \{0,1\}$ and any $r \in \{0,1\}$, $\varepsilon_1\varepsilon_0 = (1 - \varepsilon_0)(1 - \varepsilon_1)$, or $p(r|\widehat{s})p(\widehat{r}|s) = p(r|s)p(\widehat{r}|\widehat{s})$ and then output signal 0 (or 1) is as informative as output signal 1 (or 0), for input signal 0 (or 1), i.e., when output signal 0 (or 1) is observed, then it is equally likely that it comes from input signal 0 (or 1) than from input signal 1 (or 0).

Proof of Lemma 2: 1) Recall that $\widehat{\sigma}_k^S = \{x_k^j\}_{j\in\{1,\ldots,n\}}$ and $\widehat{\sigma}_{k'}^S = \{x_{k'}^j\}_{j\in\{1,\ldots,n\}}$ where $x_k^j = x_{k'}^j = 1$ but the blocks k and k' (i.e. j such that $(i-1)m-1 \leq j \leq im$ for $i \in \{k,k'\}$). Therefore, the Hamming distance in the block l is:

$$\begin{aligned} h_l(\widehat{\sigma}_k^S, y) &= \sum_{j=1}^{m} I_{y_{lm+j} \neq (\widehat{\sigma}_k^S)_{lm+j}} = \sum_{j=1}^{m} I_{y_{lm+j} \neq 1} \\ &= \sum_{j=1}^{m} I_{y_{lm+j} \neq (\widehat{\sigma}_{k'}^S)_{lm+j}} = h_l(\widehat{\sigma}_{k'}^S, y) \end{aligned}$$

2) Let us compute $h_k(\widehat{\sigma}_k^S, y) + h_k(\widehat{\sigma}_{k'}^S, y)$ if $k \neq k'$. Notice that $\{x_k^j\}_{j\in\{km+1,\ldots,(k+1)m\}} = 0$ and $\{x_{k'}^j\}_{j\in\{km+1,\ldots,(k+1)m\}} = 1$

$$\begin{aligned} h_k(\widehat{\sigma}_k^S, y) + h_k(\widehat{\sigma}_{k'}^S, y) &= h((0,\ldots,0),(y^{km+1},\ldots,y^{(k+1)m})) \\ &+ h((1,\ldots,1),(y^{km+1},\ldots,y^{(k+1)m})) \\ &= h((0,\ldots,0),(y^{km+1},\ldots,y^{(k+1)m})) \\ &+ m - h((0,\ldots,0),(y^{km+1},\ldots,y^{(k+1)m})) \\ &= m \end{aligned}$$

Proof that: Sets Y_l, $l = \{1,\ldots,|\Omega|\}$, are a true partition of Y^n, given the above description of Y_l.

1) Without loss of generality, suppose on the contrary that $Y_1 \cap Y_2 \neq \emptyset$.

a. Let $1 \in \tilde{\Omega}_2$ and $y \in Y_1 \cap Y_2$. Then,

$$\begin{aligned} h_1(\widehat{\sigma}_1^S, y) + h_2(\widehat{\sigma}_1^S, y) &\leq C_{12}, \text{ and} \\ h_2(\widehat{\sigma}_2^S, y) + h_1(\widehat{\sigma}_2^S, y) &\leq C_{21} \end{aligned}$$

and adding $h_1(\widehat{\sigma}_1^S, y) + h_1(\widehat{\sigma}_2^S, y) + h_2(\widehat{\sigma}_1^S, y) + h_2(\widehat{\sigma}_2^S, y) \leq C_{12} + C_{21}$, that by Lemma 1(2) is $m + m \leq 2m - 1$, a contradiction.

b. Let $1 \notin \tilde{\Omega}_2$ and $y \in Y_1 \cap Y_2$.

$$\begin{aligned} h_1(\widehat{\sigma}_1^S, y) + h_2(\widehat{\sigma}_1^S, y) &\leq C_{12}, \text{ and} \\ h_2(\widehat{\sigma}_2^S, y) + h_1(\widehat{\sigma}_2^S, y) &< C_{21} \end{aligned}$$

and adding $h_1(\widehat{\sigma}_1^S, y) + h_1(\widehat{\sigma}_2^S, y) + h_2(\widehat{\sigma}_1^S, y) + h_2(\widehat{\sigma}_2^S, y) < C_{12} + C_{21}$, that by Lemma 1(2) is $2m < 2m$, a contradiction again.

Proof of Proposition 1: By Bayes's Theorem,

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{\frac{p(y|\widehat{\sigma}_l^S)c(\widehat{\sigma}_l^S)}{p(y)}}{\frac{p(y|\widehat{\sigma}_k^S)p(\widehat{\sigma}_k^S)}{p(y)}} = \frac{q_l}{q_k}\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)}$$

The conditional probability of the channel to generate output y if message $\widehat{\sigma}_l^S$ is sent can be written as:

$$\begin{aligned}
p(y|\widehat{\sigma}_l^S) &= \prod_{t=1}^{n} p(y_t|(\widehat{\sigma}_l^S)_t) \\
&= \prod_{t=1}^{lm} p(y_t|1) \prod_{t=lm+1}^{(l+1)m} p(y_t|0) \prod_{t=(l+1)m+1}^{n} p(y_t|1) \\
&= \varepsilon_0^{h_l(\widehat{\sigma}_l^S,y)}(1-\varepsilon_0)^{m-h_l(\widehat{\sigma}_l^S,y))} \prod_{\substack{\alpha=1 \\ \alpha\neq l}}^{|\Omega|} \left[\varepsilon_1^{h_\alpha(\widehat{\sigma}_l^S,y)}(1-\varepsilon_1)^{m-h_\alpha(\widehat{\sigma}_l^S,y)}\right]
\end{aligned}$$

and the conditional probability to generate the same y if message $\widehat{\sigma}_k^S$ is sent instead is:

$$p(y|\widehat{\sigma}_k^S) = \varepsilon_0^{h_k(\widehat{\sigma}_k^S,y)}(1-\varepsilon_0)^{m-h_k(\widehat{\sigma}_k^S,y)} \prod_{\substack{\alpha=1 \\ \alpha\neq k}}^{|\Omega|} \left[\varepsilon_1^{h_\alpha(\widehat{\sigma}_k^S,y)}(1-\varepsilon_1)^{m-h_\alpha(\widehat{\sigma}_k^S,y)}\right]$$

Consider the likelihood ratio of messages $\widehat{\sigma}_l^S$ and $\widehat{\sigma}_k^S$, when y is realized, $\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)}$. It is not difficult to show by some cumbersome algebra that this ratio can be splitted in three separated terms.

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{q_l}{q_k} \times Ratio_1 \times Ratio_2 \times Ratio_3$$

where:

$Ratio_1$: Measures the probability of transformation of the 0's bits in block l with respect to the corresponding probability in block k. By lemma 2, part 2,

$$\begin{aligned}
Ratio_1 &= \frac{\varepsilon_0^{h_l(\widehat{\sigma}_l^S,y)}(1-\varepsilon_0)^{m-h_l(\widehat{\sigma}_l^S,y)}}{\varepsilon_0^{h_k(\widehat{\sigma}_k^S,y)}(1-\varepsilon_0)^{m-h_k(\widehat{\sigma}_k^S,y)}} \\
&= \frac{\varepsilon_0^{h_l(\widehat{\sigma}_l^S,y)}(1-\varepsilon_0)^{m-h_l(\widehat{\sigma}_l^S,y)}}{\varepsilon_0^{m-h_k(\widehat{\sigma}_l^S,y)}(1-\varepsilon_0)^{h_k(\widehat{\sigma}_l^S,y)}} \\
&= \left(\frac{\varepsilon_0}{1-\varepsilon_0}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m}
\end{aligned}$$

Ratio$_2$: Refers to the probability of transformation of the 1's bits in block k with respect to the corresponding probability in block l. Then, by lemma 2, part 2:

$$\begin{aligned}
Ratio_2 &= \frac{\varepsilon_1^{h_k(\widehat{\sigma}_l^S,y)}(1-\varepsilon_1)^{m-h_k(\widehat{\sigma}_l^S,y)}}{\varepsilon_1^{h_l(\widehat{\sigma}_k^S,y)}(1-\varepsilon_1)^{m-h_l(\widehat{\sigma}_k^S,y)}} \\
&= \frac{\varepsilon_1^{m-h_l(\widehat{\sigma}_l^S,y)}(1-\varepsilon_1)^{h_k(\widehat{\sigma}_l^S,y)}}{\varepsilon_1^{h_l(\widehat{\sigma}_k^S,y)}(1-\varepsilon_1)^{m-h_l(\widehat{\sigma}_k^S,y)}} \\
&= \left(\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m}
\end{aligned}$$

Ratio$_3$: Refers to the probability of transformation of the 1's bits in the remaining blocks (all the blocks but l and k), then, using lemma 2, part 1

$$\begin{aligned}
Ratio_3 &= \prod_{\substack{\alpha=1 \\ \alpha\neq l,k}}^{|\Omega|} \frac{\varepsilon_1^{h_\alpha(\widehat{\sigma}_l^S,y)}(1-\varepsilon_1)^{m-h_\alpha(\widehat{\sigma}_l^S,y)}}{\varepsilon_1^{h_\alpha(\widehat{\sigma}_k^S,y)}(1-\varepsilon_1)^{m-h_\alpha(\widehat{\sigma}_k^S,y)}} \\
&= 1
\end{aligned}$$

Putting these three ratios together, we have that

$$\begin{aligned}
\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} &= \frac{q_l}{q_k} \times Ratio_1 \times Ratio_2 \times Ratio_3 \\
&= \frac{q_l}{q_k}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m}\left(\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m} \\
&= \frac{q_l}{q_k}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S,y)+h_k(\widehat{\sigma}_l^S,y)-m}
\end{aligned}$$

and the proposition holds.

Proof of the Theorem 1: *The Receiver's condition*:

Given the Sender equilibrium strategy $\{\widehat{\sigma}_j^S\}_j$, and the Receiver's information set Y^n, consider the realization $y \in Y^n$. The Receiver's strategy is defined by

$$\widehat{\sigma}_y^R = \widehat{a}_l \Leftrightarrow y \in Y_l$$

where for each $l \in \{1, \ldots, |\Omega|\}$, and $\tilde{\Omega}_l = \{k \in \{1, \ldots, |\Omega|\}$ such that $q_l M_l \neq q_k M_k\}$,

$$\begin{aligned}
Y_l = \{y \in Y^n| \quad & h_l(\widehat{\sigma}_l^S, y)) + h_k(\widehat{\sigma}_l^S, y)) \leq C_{lk}, \text{ for all } k \in \tilde{\Omega}_l \\
& h_l(\widehat{\sigma}_l^S, y)) + h_{k'}(\widehat{\sigma}_l^S, y)) \leq C_{lk'}, \text{ for all } k' \epsilon \notin \tilde{\Omega}_l \ , \ k' < l \\
& h_l(\widehat{\sigma}_l^S, y)) + h_{k'}(\widehat{\sigma}_l^S, y)) < C_{lk'}, \text{ for all } k' \notin \tilde{\Omega}_l \ , \ k' > l\}
\end{aligned}$$

The parameters $\{C_{lk}\}_{l\neq k}$ are given by:

$$C_{lk} = \frac{Ln \frac{M_k q_k}{M_l q_l}}{Ln \frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + m$$

with associated payoff $\pi_y^R(\{\widehat{\sigma}_j^S\}_j, \widehat{a}_l) = \sum_{j=1}^{|\Omega|} p(\widehat{\sigma}_j^S|y)u(\widehat{a}_l, \omega_j) = p(\widehat{\sigma}_l^S)|y)M_l$.

Consider any other strategy $a_\alpha \neq \widehat{a}_l$ and suppose that its associated payoff is higher than $\widehat{a}_l$. Then,

$$\begin{aligned} \pi_y^R(\{\{\widehat{\sigma}_j^S\}_j, a_\alpha) &> \pi_y^R(\{\widehat{\sigma}_j^S\}_j, \widehat{a}_l) \\ \sum_{j=1}^{|\Omega|} p(\widehat{\sigma}_j^S)|y)u(\widehat{a}_\alpha, \omega_j) &> \sum_{j=1}^{|\Omega|} p(\widehat{\sigma}_j^S|y)u(\widehat{a}_l^R, \omega_j), \end{aligned}$$

which by the linearity of π_y^R in probabilities $p(\widehat{\sigma}_j^S)|y)$, is equal to,

$$p(\widehat{\sigma}_\alpha^S|y)M_\alpha > p(\widehat{\sigma}_l^S|y)M_l$$

or

$$\frac{M_\alpha}{M_l} > \frac{p(\widehat{\sigma}_l^S)|y)}{p(\widehat{\sigma}_\alpha^S|y)}$$

By Proposition 1, these inequalities can be written as

$$\frac{M_\alpha}{M_l} > \frac{q_l}{q_\alpha}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S, y)) + h_\alpha(\widehat{\sigma}_l^S, y)) - m}$$

or

$$\frac{q_\alpha M_\alpha}{q_l M_l} > \left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_l(\widehat{\sigma}_l^S, y)) + h_\alpha(\widehat{\sigma}_l^S, y)) - m}$$

We write this condition with the Logarithm operator:

$$\ln\left(\frac{q_\alpha M_\alpha}{q_l M_l}\right) > (h_l(\widehat{\sigma}_l^S, y)) + h_\alpha(\widehat{\sigma}_l^S, y)) - m) \ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)$$

Since $\varepsilon_0 + \varepsilon_1 < 1$, then $\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1} < 1$ and $\ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right) < 0$. Then, the above inequality is equivalent to:

$$\begin{aligned} h_l(\widehat{\sigma}_l^S, y)) + h_\alpha(\widehat{\sigma}_l^S, y) &> \frac{\ln\left(\frac{q_\alpha M_\alpha}{q_l M_l}\right)}{\ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)} + m, \text{ or} \\ h_l(\widehat{\sigma}_l^S, y)) + h_\alpha(\widehat{\sigma}_l^S, y) &> C_{l\alpha} \end{aligned}$$

by the definition of $C_{l\alpha}$. But this contradicts that $y \in Y_l$, since by definition $Y_l = \{y \in Y^n |\ h_l(\sigma_l^S, y)) + h_k(\sigma_l^S, y)) \leq C_{lk}$, for all $k \neq l\}$, in particular for $k = \alpha$. Therefore for each $y \in Y^n$ there is no profitable deviation from $\widehat{\sigma}_y^R$, and $\widehat{\sigma}_y^R$ is a best response to $\{\widehat{a}_j^S\}_j$.

The Sender's condition. Truth-telling: The Sender's strategy at state ω_j consists of sending a message and thus it suffices to show that there is no profitable deviation by sending another message different from $\widehat{\sigma}_j^S$, when R plays $\{\widehat{\sigma}_y^R\}_y$. The associated payoff of $\widehat{\sigma}_j^S$ at state ω_j when the Receiver plays his equilibrium strategy $\{\widehat{\sigma}_y^R\}_y$ is

$$\pi_j^S(\widehat{\sigma}_j^S, \{\widehat{\sigma}_y^R\}_y) = \sum_{y \in Y^n} p(y|\widehat{\sigma}_j^S) u(\widehat{\sigma}_y^R, \omega_j) = M_j \sum_{y \in Y_j} p(y|\widehat{\sigma}_j^S)$$

since $u(a_t, \omega_j) = 0$ for all $a_t \neq \widehat{a}_j^R$.

Consider the associated payoff of sending any other message $x \in X^n$,

$$\sum_{y \in Y^n} p(y|x) u(\widehat{\sigma}_y^R, \omega_j) = M_j \sum_{y \in Y_j} p(y|x)$$

Let $f = h(\widehat{\sigma}_j^S, x)$ be the Hamming distance between messages $\widehat{\sigma}_j^S$ and x. We can construct a sequence of messages $\{\theta_0, \theta_1, \ldots, \theta_f\}$ such that $\theta_i \in X^n, \theta_0 = \widehat{\sigma}_j^S, \theta_f = x$ satisfying that, for all $d = 0, \ldots, f-1$,

$$\begin{aligned} h(\widehat{\sigma}_j^S, \theta_{d+1}) &= h(\widehat{\sigma}_j^S, \theta_d) + 1 \\ h(\theta_d, \theta_{d+1}) &= 1 \end{aligned}$$

This sequence transforms message $\widehat{\sigma}_j^S$ into message x by only changing one element at each step. Let us show that, for all $d = 0, \ldots, f-1$,

$$\sum_{y \in Y_j} \frac{p(y|\theta_d)}{p(y|\theta_{d+1})} \geq 1$$

Let i_d be the location of the (unique) mismatch between θ_d and θ_{d+1}. Then,

$$\begin{aligned} \sum_{y \in Y_j} \frac{p(y|\theta_d)}{p(y|\theta_{d+1})} &= \sum_{y \in Y_j} \frac{\prod_{i=1,\ldots,n} p(y^i|\theta_d^i)}{\prod_{i=1,\ldots,n} p(y^i|\theta_{d+1}^i)} \\ &= \sum_{y \in Y_j} \frac{p(y^{i_d}|\theta_d^{i_d})}{p(y^{i_d}|\theta_{d+1}^{i_d})} \\ &= \sum_{y \in Y_j | y^{i_d}=0} \frac{p(0|\theta_d^{i_d})}{p(0|\theta_{d+1}^{i_d})} + \sum_{y \in Y_j | y^{i_d}=1} \frac{p(1|\theta_d^{i_d})}{p(1|\theta_{d+1}^{i_d})} \end{aligned}$$

Let us consider two different cases:

Case 1: $(j-1)m-1 \leq i_d \leq jm$. The mismatch occurs at block j and as $y \in Y_j$, then the element $\theta_d^{i_d}$ coincides with the element $y^{i_d} = 0$ and $\theta_{d+1}^{i_d} = 1$. The above expression is now given by:

$$\begin{aligned}\sum_{y\in Y_j} \frac{p(y|\theta_d)}{p(y|\theta_{d+1})} &= \sum_{y\in Y_j|y^{i_d}=0} \frac{p(0|0)}{p(0|1)} + \sum_{y\in Y_j|y^{i_d}=1} \frac{p(1|0)}{p(1|1)} \\ &= \sum_{y\in Y_j|y^{i_d}=0} \frac{1-\varepsilon_0}{\varepsilon_1} + \sum_{y\in Y_j|y^{i_d}=1} \frac{\varepsilon_0}{1-\varepsilon_1} \geq 1\end{aligned}$$

Notice that there exists at least an element $y \in Y_j$ with $y^{i_d} = 0$ and the ratio $\frac{1-\varepsilon_0}{\varepsilon_1} \geq 1$ because $\varepsilon_0 + \varepsilon_1 < 1$. Therefore $\sum_{y\in Y_j|y^{i_d}=0} \frac{1-\varepsilon_0}{\varepsilon_1} \geq 1$.

Case 2: $i_d < (j-1)m-1$ or $i_d > jm$. The mismatch occurs in a different block of j and as above $y \in Y_j$, then the element $\theta_d^{i_d}$ coincides with the element $y^{i_d} = 1$ and $\theta_{d+1}^{i_d} = 0$. The above expression is now given by:

$$\begin{aligned}\sum_{y\in Y_j} \frac{p(y|\theta_d)}{p(y|\theta_{d+1})} &= \sum_{y\in Y_j|y^{i_d}=0} \frac{p(0|1)}{p(0|0)} + \sum_{y\in Y_d|y^{i_d}=1} \frac{p(1|1)}{p(1|0)} \\ &= \sum_{y\in Y_j|y^{i_d}=0} \frac{\varepsilon_1}{1-\varepsilon_0} + \sum_{y\in Y_j|y^{i_d}=1} \frac{1-\varepsilon_1}{\varepsilon_0} \geq 1\end{aligned}$$

The set of elements in Y_j such that $y^{i_d} = 1$ has cardinality greater or equal than 1. Therefore $\sum_{y\in Y_j|y^{i_d}=1} \frac{1-\varepsilon_1}{\varepsilon_0} \geq 1$. From the above reasoning, the probability $p(y|\theta_d)$ decreases at each step of the deviation chain $\{\widehat{\sigma}_j^S, \theta_1, \ldots, x\}$. We conclude that $\sum_{y\in Y_j} p(y|\widehat{\sigma}_j^S)) \geq \sum_{y\in Y_j} p(y|x)$ and the associated payoffs for both messages $\widehat{\sigma}_j^S$ and x verify the condition $M_j \sum_{y\in Y_j} p(y|\widehat{\sigma}_j^S)) \geq M_j \sum_{y\in Y_j} p(y|x)$ that closes the proof. Hence, for each $\widehat{\sigma}_j^S$ is a best response to $\{\widehat{\sigma}_y^R\}_y$.

Since at each state ω_j, the Sender pure strategy $\widehat{\sigma}_j^S$ is a best response to $\{\widehat{\sigma}_y^R\}_y$ and for each $y \in Y^n$, the Receiver pure strategy $\widehat{\sigma}_y^R$ is a best response to $\{\widehat{\sigma}_j^S\}_j$, then the pair of tuples $(\{\widehat{\sigma}_j^S\}_j, \{\widehat{\sigma}_y^R\}_y)$ is a pure strategy Nash equilibrium of Γ_v^n.

Proof of proposition 2*:* Without loss of generality, let us assume that $j = 1$ and $C_{12} \leq C_{13} \leq ... \leq C_{1|\Omega|}$. Under these assumptions the theorem is proved provided that:

$$p(Y_1|\widehat{\sigma}_1^S) \geq 1 - \varepsilon_1^{c_{12}+1}\phi(\varepsilon_0, \varepsilon_1)$$

where $\phi(\varepsilon_0, \varepsilon_1)$ is a function on ε_0 and ε_1 such that $0 \leq \phi(\varepsilon_0, \varepsilon_1) \leq 1$. The probability $p(Y_1|\widehat{\sigma}_1^S)$ can be written as

$$p(Y_1|\widehat{\sigma}_1^S) = 1 - p(Y^n - Y_1|\widehat{\sigma}_1^S)$$

Let $J \subseteq \{2, 3, \ldots, |\Omega|\}$ and let us E_J denote the event

$$E_J = \{y \in Y^n - Y_1 | h_1(\widehat{\sigma}_1^S, y) + h_j(\widehat{\sigma}_1^S, y) > C_{1j} \text{ if } j \in J \text{ and } h_1(\widehat{\sigma}_1^S, y) + h_l(\widehat{\sigma}_1^S, y) \leq C_{1l} \text{ if } l \notin J\}$$

We have that:

$$\begin{aligned} E_{J_1} \cap E_{J_2} &= \emptyset \text{ if } J_1, J_2 \subseteq \{2, 3, ..|\Omega|\},\ J_1 \neq J_2 \\ Y^n - Y_1 &= \bigcup_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} E_J \end{aligned}$$

In other words, $\{E_J\}_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}}$ is a partition of $Y^n - Y_1$ formed by disjoint events. In such a case, the probability $p(y \in Y^n - Y_1 | \widehat{\sigma}_1^S)$ can be written as:

$$p(y \in Y^n - Y_1 | \widehat{\sigma}_1^S) = p(\bigcup_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} E_J | \widehat{\sigma}_1^S) = \sum_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} p(E_J | \widehat{\sigma}_1^S)$$

Similarly, for each $J \subseteq \{2, 3, ..|\Omega|\}$, define the event

$$F_J = \{y \in E_J | h_1(\widehat{\sigma}_1^S, y) + h_j(\widehat{\sigma}_1^S, y) = C_{12} + 1\}$$

Since

$$p(F_J | \widehat{\sigma}_1^S) = \varepsilon_0^{h_1(\widehat{\sigma}_1^S, y)} \varepsilon_1^{h_j(\widehat{\sigma}_1^S, y)} \leq \varepsilon_1^{h_1(\widehat{\sigma}_1^S, y)} \varepsilon_1^{h_j(\widehat{\sigma}_1^S, y)} = \varepsilon_1^{C_{12}+1}$$

We have then

$$\begin{aligned} p(y &\in Y^n - Y_1 | \widehat{\sigma}_1^S) = \sum_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} p(E_J | F_J, \widehat{\sigma}_1^S) p(F_J | \widehat{\sigma}_1^S) \\ &\leq \varepsilon_1^{C_{12}+1} \sum_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} p(E_J | F_J, \widehat{\sigma}_1^S) = \varepsilon_1^{C_{12}+1} p(\bigcup_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} E_J | F_J, \widehat{\sigma}_1^S) \end{aligned}$$

Let

$$\phi(\varepsilon_0, \varepsilon_1) = p(\bigcup_{\substack{J \subseteq \{2,3,..|\Omega|\} \\ J \neq \emptyset}} E_J | F_J, \widehat{\sigma}_1^S)$$

then

$$p(Y_1 | \widehat{\sigma}_1^S) \geq 1 - \varepsilon_1^{c_{12}+1} \phi(\varepsilon_0, \varepsilon_1)$$

and the theorem holds.

9 Appendix B

The Receiver's best response for any feasible corpus or encoding rule:

Notations:

To make easier the comparison between any observed output sequence y and any pair of standar prototypes $\widehat{\sigma}_i^S = (x_i^1, ..., x_i^n)$ and $\widehat{\sigma}_j^S = (x_j^1, ..., x_j^n)$, let us split the sequences $\widehat{\sigma}_i^S$, $\widehat{\sigma}_j^S$ in four different blocks or subsequences. Each block is formed by the elements of the sequences with subindexes in the following sets:

$$\begin{aligned}
T_{00}^{ij} &= \{l \mid 1 \leq l \leq n \text{ and } x_i^l = 0 \text{ and } x_j^l = 0\} \\
T_{01}^{ij} &= \{l \mid 1 \leq l \leq n \text{ and } x_i^l = 0 \text{ and } x_j^l = 1\} \\
T_{10}^{ij} &= \{l \mid 1 \leq l \leq n \text{ and } x_i^l = 1 \text{ and } x_j^l = 0\} \\
T_{11}^{ij} &= \{l \mid 1 \leq l \leq n \text{ and } x_i^l = 1 \text{ and } x_j^l = 1\}
\end{aligned}$$

For instance, $l \in T_{10}^{ij}$ means that the element of $\widehat{\sigma}_i^S$ placed in position l is a 1, while the same element in the l-position in $\widehat{\sigma}_j^S$ is a 0. Moreover, some of the sets $T_{\alpha\beta}^{ij}$ may be empty, for $\alpha, \beta = 0, 1$. Notice that this subdivision of any two sequences in four blocks may be differente for each pair of prototypes to be compared (the use of superindexes ij is then needed to distinguish among each pair of prototyopes).

To further proceed, we need to introduce some additional notation representing the cardinality of these subsets and the Hamming distance among the four subsequences of the output and prototypes to be compared. Formally, for $\alpha, \beta = 0, 1$, define:

$$n_{\alpha\beta}^{ij} = |T_{\alpha\beta}^{ij}|$$

$$h_{\alpha\beta}^{ij}(y, \widehat{\sigma}_i^S) = \sum_{l \in T_{\alpha\beta}^{ij}} I_{y^l \neq x_i^l} \text{ and } h_{\alpha\beta}^{ij}(y, \widehat{\sigma}_j^S) = \sum_{l \in T_{\alpha\beta}^{ij}} I_{y^l \neq x_j^l}$$

where I stands for the indicator function.

Example. For instance, T_{01}^{ij} is the subset of indexes $l = 1, ..., n$ corresponding to those position ocupied by zeros in $\widehat{\sigma}_i^S$ and by ones in $\widehat{\sigma}_j^S$ (i. e., $x_i^l = 0$ and $x_j^l = 1$). If $n = 6$ and $\widehat{\sigma}_i^S = (1, 1, 0, 0, 1, 0)$ and $\widehat{\sigma}_j^S = (1, 0, 0, 1, 0, 1)$ we have that $T_{01}^{ij} = \{4, 6\}$ and $n_{01}^{ij} = 2$. Additionally $h_{01}^{ij}(\widehat{\sigma}_i^S, y) = h((0, 0), (y^4, y^6))$ is the number of elements of y located in positions 4 and 6 that are not zero and $h_{01}^{ij}(\widehat{\sigma}_j^S, y) = h((1, 1), (y^4, y^6))$ is the number of elements of y located in positions 4 and 6 that are not one. Notice that each sequence y is splitted if four separate blocks whose elements are placed at positions $T_{00}^{ij} = \{3\}$, $T_{01}^{ij} = \{4, 6\}$, $T_{10}^{ij} = \{2, 5\}$ and $T_{11}^{ij} = \{1\}$.

It is straightforward to check that:

Lemma 3 *For all $i, j = 1, 2, \ldots, |\Omega|, i \neq j$ and for all $y \in \{0, 1\}^n$ we have that*

1. $n_{00}^{ij} + n_{01}^{ij} + n_{10}^{ij} + n_{11}^{ij} = n$

2. $n_{00}^{ij} = n_{00}^{ji}; n_{11}^{ij} = n_{11}^{ji}; n_{01}^{ij} = n_{10}^{ji}$

3. $h_{00}^{ij}(y, \widehat{\sigma}_i^S) = h_{00}^{ij}(y, \widehat{\sigma}_j^S)$

4. $h_{11}^{ij}(y, \widehat{\sigma}_i^S) = h_{11}^{ij}(y, \widehat{\sigma}_j^S)$

5. $h_{01}^{ij}(y, \widehat{\sigma}_i^S) = n_{01}^{ij} - h_{01}^{ij}(y, \widehat{\sigma}_j^S)$

6. $h_{10}^{ij}(y, \widehat{\sigma}_i^S) = n_{10}^{ij} - h_{10}^{ij}(y, \widehat{\sigma}_j^S)$

Proposition 3 *For all $l, k = 1, \ldots, |\Omega|$, $l \neq k$ and for all $y \in Y$,*

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{q_l}{q_k}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)+h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)^{n_{10}^{lk}}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)^{n_{01}^{lk}}$$

Proof: Given any $k \neq l$, all the elements of $\widehat{\sigma}_l^S$ in each block $T_{\alpha\beta}^{lk}$ $(\alpha, \beta = 0, 1)$ are constant and equal to α. Then, $h_{\alpha\beta}^{lk}(\widehat{\sigma}_l^S, y)$ is just the number of the $n_{\alpha\beta}^{lk}$ elements with value α in $\widehat{\sigma}_l^S$ (placed at positions $T_{\alpha\beta}^{lk}$) misstransmited by the channel, where misstransmision takes place with probability ε_α. Alternatively, $n_{\alpha\beta}^{lk} - h_{\alpha\beta}^{lk}(\widehat{\sigma}_l^S, y)$ is the number of elements α sent properly (each of them with probability $1 - \varepsilon_\alpha$). Since the channel transforms elements independently, the $n_{\alpha\beta}^{lk}$ elements of $\widehat{\sigma}_l^S$ in $T_{\alpha\beta}^{lk}$ become the corresponding $n_{\alpha\beta}^{lk}$ of y with probality

$$(1-\varepsilon_\alpha)^{n_{\alpha\beta}^{lk}-h_{\alpha\beta}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_\alpha^{h_{\alpha\beta}^{lk}(\widehat{\sigma}_l^S,y)}$$

Applying this reasoning to the four blocks $T_{\alpha\beta}^{lk}$, we can write the probability of the noisy channel generating output y when the prototype $\widehat{\sigma}_l^S$ was sent as:

$$\begin{aligned} p(y|\widehat{\sigma}_l^S) &= (1-\varepsilon_0)^{n_{00}^{lk}-h_{00}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{00}^{lk}(\widehat{\sigma}_l^S,y)} \times (1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)} \times \\ & \quad (1-\varepsilon_1)^{n_{11}^{lk}-h_{11}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{11}^{lk}(\widehat{\sigma}_l^S,y)} \times (1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)} \end{aligned}$$

Similarly:

$$\begin{aligned} p(y|\widehat{\sigma}_k^S) &= (1-\varepsilon_0)^{n_{00}^{lk}-h_{00}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_0^{h_{00}^{lk}(\widehat{\sigma}_k^S,y)} \times (1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_k^S,y)} \times \\ & \quad (1-\varepsilon_1)^{n_{11}^{lk}-h_{11}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_1^{h_{11}^{lk}(\widehat{\sigma}_k^S,y)} \times (1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_k^S,y)} \end{aligned}$$

and then:

$$\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)} = \frac{(1-\varepsilon_0)^{n_{00}^{lk}-h_{00}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{00}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_0)^{n_{00}^{lk}-h_{00}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_0^{h_{00}^{lk}(\widehat{\sigma}_k^S,y)}} \times \frac{(1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_k^S,y)}} \times \frac{(1-\varepsilon_1)^{n_{11}^{lk}-h_{11}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{11}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_1)^{n_{11}^{lk}-h_{11}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_1^{h_{11}^{lk}(\widehat{\sigma}_k^S,y)}} \times \frac{(1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_k^S,y)}}$$

By the above Lemma, $h_{00}^{lk}(\widehat{\sigma}_l^S,y) = h_{00}^{lk}(\widehat{\sigma}_k^S,y)$ and $h_{11}^{lk}(\widehat{\sigma}_l^S,y) = h_{11}^{lk}(\widehat{\sigma}_k^S,y)$, then the first and third ratio of the above expression are 1. Hence

$$\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)} = \frac{(1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_k^S,y)}} \frac{(1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_k^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_k^S,y)}}$$

In addition, the above lemma also expresses the Hamming distances to $\widehat{\sigma}_k^S$ in terms of those to $\widehat{\sigma}_l^S$ and then:

$$\begin{aligned}
\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)} &= \frac{(1-\varepsilon_0)^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_0)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_0^{n_{01}^{lk}-h_{01}^{lk}(\widehat{\sigma}_l^S,y)}} \frac{(1-\varepsilon_1)^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)}}{(1-\varepsilon_1)^{h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\varepsilon_1^{n_{10}^{lk}-h_{10}^{lk}(\widehat{\sigma}_l^S,y)}} \\
&= \left(\frac{\varepsilon_1}{1-\varepsilon_0}\right)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{01}^{lk}(\widehat{\sigma}_l^S,y)-n_{01}^{lk}} \left(\frac{\varepsilon_0}{1-\varepsilon_1}\right)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{01}^{lk}(\widehat{\sigma}_l^S,y)-n_{10}^{lk}} \\
&= \left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{01}^{lk}(\widehat{\sigma}_l^S,y)} \left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)^{n_{10}^{lk}} \left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)^{n_{01}^{lk}}
\end{aligned}$$

and since

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} = \frac{q_l}{q_k}\frac{p(y|\widehat{\sigma}_l^S)}{p(y|\widehat{\sigma}_k^S)}$$

the Proposition holds.

The next proposition characterizes the best response of R to any received message y in terms of the Hamming distances among subsequences of y and the standard prototypes of any corpus, the noisy parameters and the payoffs:

Proposition 4 *Given any corpus $\{\widehat{\sigma}_1^S,..,\widehat{\sigma}_{|\Omega|}^S\}$ and any output sequence $y \in \{0,1\}^n$, the action $\widehat{a}_l$ is a best response to y if and only if, the following system of $|\Omega|-1$ inequalities is satisfied:*

$$\{h_{01}^{lk}(\widehat{\sigma}_l^S,y) + h_{10}^{lk}(\widehat{\sigma}_l^S,y)\} \leq \frac{Ln\frac{q_k M_k}{q_l M_l}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} + \frac{n_{10}^{lk}Ln\frac{\varepsilon_0}{1-\varepsilon_1} + n_{01}^{lk}Ln\frac{\varepsilon_1}{1-\varepsilon_0}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}}, \; for \; all \; k \neq l\}$$

Proof: Given $y, \widehat{a}_l$ is best response to y if and only if

$$\frac{p(\widehat{\sigma}_l^S|y)}{p(\widehat{\sigma}_k^S|y)} \geq \frac{M_k}{M_l}$$

By Proposition 1, the best response condition can be written as

$$\frac{q_l}{q_k}\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)^{n_{10}^{lk}}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)^{n_{01}^{lk}} \geq \frac{M_k}{M_l}$$

$$\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)^{h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{10}^{lk}(\widehat{\sigma}_l^S,y)}\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)^{n_{10}^{lk}}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)^{n_{01}^{lk}} \geq \frac{q_k M_k}{q_l M_l}$$

$$(h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{10}^{lk}(\widehat{\sigma}_l^S,y))Ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)+n_{10}^{lk}Ln\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)+n_{01}^{lk}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right) \geq Ln\frac{q_k M_k}{q_l M_l}$$

$$(h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{10}^{lk}(\widehat{\sigma}_l^S,y))Ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right) \geq Ln\frac{q_k M_k}{q_l M_l}-n_{10}^{lk}Ln\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)-n_{01}^{lk}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)$$

$$(h_{01}^{lk}(\widehat{\sigma}_l^S,y)+h_{10}^{lk}(\widehat{\sigma}_l^S,y)) \leq \frac{Ln\frac{q_k M_k}{q_l M_l}-n_{10}^{lk}Ln\left(\frac{1-\varepsilon_1}{\varepsilon_0}\right)-n_{01}^{lk}\left(\frac{1-\varepsilon_0}{\varepsilon_1}\right)}{Ln\left(\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}\right)}$$

and the proposition holds.

The characterization of the proposition applies to any feasible corpus (i. e. encoding rule) used by the sender. In words, $\widehat{a}_l$ is chosen whenever the distances of y and prototype $\widehat{\sigma}_l^S$ in blocks T_{01}^{lk} and T_{10}^{lk}, for all $k \neq l$, $k = \{1, 2, ..., \Omega\}\backslash l$, are smaller than some bounds involving both game parameters and encoding parameters. Specifically the first part of the bound for the Hamming distance between y and $\widehat{\sigma}_l^S$:

$$\frac{Ln\frac{q_k M_k}{q_l M_l}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}}$$

refers only to the ratio of expected payoffs and the noise parameters, and is not related to any specific corpus. Meanwhile, the second part

$$\frac{n_{10}^{lk}Ln\frac{\varepsilon_0}{1-\varepsilon_1}+n_{01}^{lk}Ln\frac{\varepsilon_1}{1-\varepsilon_0}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}}$$

depends on the particular encoding rule, specifically on parameters n_{10}^{lk} and n_{01}^{lk}. Notice that if the encoding rule or corpus satisfies that $n_{10}^{lk} = n_{01}^{lk} = m$ for all $l, k = 1, ..., |\Omega|$, then

$$\frac{n_{10}^{lk} Ln\frac{\varepsilon_0}{1-\varepsilon_1} + n_{01}^{lk} Ln\frac{\varepsilon_1}{1-\varepsilon_0}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} = \frac{mLn\frac{\varepsilon_0}{1-\varepsilon_1} + mLn\frac{\varepsilon_1}{1-\varepsilon_0}}{Ln\frac{\varepsilon_0}{1-\varepsilon_0}\frac{\varepsilon_1}{1-\varepsilon_1}} = m$$

It is clear that the block coding rule satisfies this property.