The Language Game and Evolutionary Game Theory

Akky van de Kerkhof
5970539

A thesis presented for the degree of
Bachelor of Mathematics

Written under supervision of
Dr. M. Ruijgrok

Utrecht University

June 2019

Contents

Chapter 1

Introduction

1.1 Motivation

This section is based mostly on [2], [3] and [5].

For many years, information was primarily transferred genetically from one individual to another, from one generation to the next. About one million years ago, human language started to evolve. Information could then be transferred by language as well, which gave rise to a new form of evolution, namely, cultural evolution (M.A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*, [2]). However, humans from different language communities make use of different sounds to denote the same object, action or property. A main topic in modern linguistics therefore is the search for properties that are shared by all languages. A certain collection of concept-to-sign mappings can be seen as a proto-language. Evolutionary game theory - which provides a formal framework for studying both biological and cultural evolution of frequency-dependent phenomena - can be used to show how such a proto-language can evolve from a pre-linguistic environment (C. Pawlowitsch. *Finite populations choose an optimal language*, [5]). From a linguistic point a view, an interesting question in the evolution of concept-to-sign mappings is whether a simple replication mechanism will always lead to *bidirectionality*, that is, the property that whenever a specific signal is used to communicate a specific object, this signal will also evoke the image of this object. It turns out that in an infinitely large population, this is not necessarily the case, since the deterministic replicator dynamics can be blocked in a suboptimum state, where one object of communication is linked to two or more signals, or where one signal is used for two or more objects (C. Pawlowitsch. *Why evolution does not always lead to an optimal signaling game*, [6]). However, in the beginning of language, small population size presumably played a crucial role. Hence, in order to obtain a good model for the beginning of language, it is important to consider finite population sizes.

1.2 Synopsis

In this thesis we discuss the evolution of language in a finite population by considering a mathematical framework based on game theory and evolutionary dynamics. In particular, we are interested in the evolution of vocabulary, hence the association between signals and objects. Therefore we consider the Language Game. The Language Game is a sender-receiver game in which the sender wishes to transfer information about a certain object to a receiver and where we precisely examine associations between used signals and objects. We will explain the Language Game in more detail in chapter 4. Our aim in this thesis is to show that in a finite population, efficient proto-languages, where one object of communication is associated by a unique signal and vice versa, are the only strategies that are protected by selection. In order to show this, we discuss the evolution of a lexical matrix in a finite population under the frequency dependent Moran process in the style of Nowak et al. [4], as Pawlowitsch did in [5]. Eventually, the Moran process leads to fixation of a single strategy throughout the entire population. However, new variants can arise by mutations. A crucial aspect of the frequency-dependent Moran process in a finite population is that a single mutant strategy can take over the entire population through the effects of drift, even if this mutant strategy has a slight disadvantage in terms of relative fitness against the resident type. Hence, it is not straightforward that there exist evolutionary stable strategies.

In the first two chapters we introduce some background information on evolutionary game theory in a finite population, essential for this thesis. For this, we make use of [1], [2], [4], [5] and [7]. In chapter 2 we discuss the influence of population size on evolutionary dynamics and give a short recap of the frequency dependent Moran process, which is often used to describe evolutionary dynamics in a finite population. In chapter 3 we give a definition on evolutionary stability in a finite population as proposed by Nowak et al. [4]. This definition requires two conditions: selection opposes a mutant strategy invading the resident strategy, and, selection opposes a mutant strategy replacing the resident strategy. In section 3.1 we show that in case of a symmetric game, the first condition is equivalent with the resident strategy being a strict Nash strategy. In section 3.2 we show that in a symmetric game under weak selection, the second condition is equivalent with saying that no mutant strategy has a chance of reaching fixation greater than the neutral threshold, which is equal to the inverse population size.

In chapter 4 we explain and analyse the Language Game, based on [3], [5], [6] and [8]. In section 4.1 we begin with introducing the strategy sets, where we suppose that an individual can either be in the role of the sender or in the role of the receiver. From this, we derive in section 4.2 the potential of communication between a sender and a receiver, which is similar to the payoff of the game. Then we determine the set of best responses for both the sender and the receiver in section 4.3. Furthermore we give a characterisation of best-response properties which are needed later on. The obtained symmetric payoff function together with the strategy sets of the sender and receiver then constitute a two-player asymmetric game with a symmetric payoff function. In section 4.4 we symmetrize the asymmetric game by assuming that an individual adopts both the role of a sender and a receiver with equal probabilities depending on its relative position in different situations. From this, we derive the payoff in the symmetric game in section 4.5 and we give an upper bound for this payoff in section 4.6. Lastly, we determine (strict) Nash equilibria of the game in section 4.7. Here we will also see that in a Nash equilibrium, the strategies of an individual played in both the role of the sender and the receiver have to be best responses to each other, that is, equivalent to the best responses in the asymmetric game.

In chapter 5 we view the definition of evolutionary stability in finite populations, as given in chapter 3, in context of the Language Game. Our main focus lies on the second condition of this definition. Here, we show that the only evolutionary stable strategies of the Language Game are indeed efficient proto-languages, where there is a bijective map from objects of communication to used signals and vice versa.

In chapter 6 we give a few examples of competing languages to clarify the notion of evolutionary stability and the outcome that efficient proto-languages are the only evolutionary stable strategies.

We end this thesis by combining all results in a recap in chapter 7.

This thesis is based mostly on the article "Finite Populations Choose an Optimal Language" by Christina Pawlowitsch [5]. In all chapters we clarify the sources that are used mostly.

Chapter 2

Evolutionary Game Theory

This chapter is based on [2], [5] and [7].

Evolutionary game theory is based on populations of players of a certain type and focuses on the evolutionary dynamics, which can be both genetic or cultural. Hence, the distribution of the population can evolve over time. This can be due to fitness of a certain type or efficiency of strategies. Since evolutionary game theory is based on the assumption that fitness is frequency dependent, it follows that the evolutionary dynamics are influenced by the frequency in occurrence of the competing strategies inside the population. Furthermore, new strategies may arise over time. The main question in evolutionary game theory is therefore which strategies become extinct and which can survive over time. Generally, evolutionary dynamics in games is studied through replicator equations, which show the growth rate of the proportion of the population that plays a certain strategy. However, these equations are not representative for the real situation in a finite population. In the next section, we therefore explain the influence of the population size on evolutionary dynamics.

2.1 Infinite and finite population sizes

This section is based mostly on [7].

We draw a distinction between an infinite and a finite population size. If the model is based on an infinite population size, the frequency changes of different types are deterministic. This is due to dynamic rules of the reproduction process, such as the replicator dynamics. Here, the frequency of a certain type changes deterministically in proportion to the difference of its own fitness and the average population fitness. If the population size is sufficiently large, the replicator dynamics gives a good approximation (P.D. Taylor and L. Jonker *Evolutionary stable strategies and game dynamics*, [7]). However, the deterministic replicator dynamics does not take any effects of drift into consideration. Whereas in a finite population, drift is automatically present since differences in relative fitness only translates into expected and not realised offspring. Hence, random genetic drift is considered to play an important role in the evolution of finite populations. Therefore, if the model is based on a finite population size, the frequencies will fluctuate by chance. The smaller the population is, the greater the frequency fluctuations are. A model that captures such stochastic effects and that is often used to describe finite populations is the Moran process, as we will elaborate on in the next section.

In this thesis we assume well-mixed populations, which mathematically corresponds to populations under random mating.

2.2 The Moran process

This section is based mostly on [2].

The Moran process is a simple stochastic process that describes the evolutionary dynamics in a finite population of constant size N with overlapping generations. In order to ensure that the total population size remains constant, there is always assumed one birth and one death event in each time step. Hence, one individual is chosen at random for reproduction and one individual is chosen at random for removal. The identical offspring of the first individual replaces the second individual in the population. If the individual chosen to reproduce is under uniform random sampling among all individuals inside the population, the process is under neutral evolution. If this sampling depends on the fitness of a certain type, then selection acts on the population dynamics. As a consequence, a type that has a higher fitness is more likely to be chosen for reproduction. Note that it is possible that an individual gets replaced by its own offspring. This is the case when one and the same individual from the population is chosen to reproduce and die at the same time.

The only stochastic variable in the Moran process is the number of individuals inside the population that are of a certain type. Hence, in a population consisting of only two types of individuals, I_1 and I_2, we can let the number of individuals of type I_1 equal k, consequently the number of individuals of type I_2 equals $N-k$. It follows that the Moran process is defined on the state space $k = 0, 1, ..., N$. We denote the state of the population with the vector $X = (k, N-k)$. As this process can be seen as a Brownian motion, the Moran process eventually leads to fixation of a single type throughout the entire population, where new variation can only be brought in by mutations.

Chapter 3

Evolutionary Stability

This chapter is based on [4] and [5].

As mentioned in section 2.2, the frequency-dependent Moran process leads to fixation of a single strategy throughout the entire population, where the only way new variation can be brought in is by mutations. Assume that the population has reached a state where all individuals use the same strategy. We refer to this strategy as the resident strategy. In order to evaluate the stability of this resident strategy in an evolutionary sense, we need the notion of evolutionary stability. For finite populations, there is no formal definition for a strategy to be evolutionary stable. However, we consider the following definition as proposed by Nowak et al. [4].

Definition 3.1 (Nowak et al. [4])**.** In a finite population of size N, a strategy is evolutionary stable if it satisfies the following two conditions

(1) selection opposes all mutant strategies invading the resident strategy,

(2) selection opposes all mutant strategies replacing the resident strategy.

Here, condition (1) means that a single mutant must have a lower fitness than the rest of the population, who use the resident strategy. And condition (2) means that the probability that a single mutant will invade and take over the entire population, is smaller than $1/N$, the neutral threshold. Hence, for an evolutionary stable language, it holds true that once it is fixed in a population, natural selection alone is sufficient to prevent mutant strategies from invading successfully.

In the sequel we assume that there are only two competing strategies: the resident strategy, indicated by R, and an arbitrary mutant strategy, indicated by M. Restricting attention to these two strategies, the payoff matrix becomes

$$\begin{array}{cc} & \begin{array}{cc} R & M \end{array} \\ \begin{array}{c} R \\ M \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{array}$$

Where a, b, c and d are real numbers.

From this payoff matrix, we can derive the fitness of R and M. Assume that the state of the population is given by $X = (X_1, X_2)$, where X_1 is the number of individuals who use the resident strategy R and X_2 is the number of individuals who use the mutant strategy M. Since R and M are the only two competing strategies in the population, we have $X_1 + X_2 = N$. The fitness of R, depending linearly on frequencies, can be derived as follows. Consider an individual who uses the resident strategy R. This individual can interact with the other $N-1$ individuals in the population. From this $N-1$ individuals there are $X_1 - 1$ individuals left who use R and X_2 individuals who use M. Hence, the chance of meeting another individual who uses R is $\frac{X_1-1}{N-1}$ and the chance of meeting an individual who uses M is $\frac{X_2}{N-1}$. The payoff from interaction with another individual who uses R equals a, whereas the payoff from interaction with an individual who uses M equals b. Therefore the expected payoff for an individual who uses the resident strategy R equals $\frac{X_1-1}{N-1} \cdot a + \frac{X_2}{N-1} \cdot b$. In the same way we can determine the expected payoff for an individual who uses a mutant strategy M. We obtain that the fitness of R, and, respectively M, given the state of the population, is equal to

$$F(R|X_1, X_2) = \frac{X_1 - 1}{N - 1} \cdot a + \frac{X_2}{N - 1} \cdot b. \tag{3.1}$$

$$F(M|X_1, X_2) = \frac{X_1}{N - 1} \cdot c + \frac{X_2 - 1}{N - 1} \cdot d. \tag{3.2}$$

In the next two sections we shall discuss conditions (1) and (2) from Definition 3.1 in more detail.

3.1 Selection opposes all mutant strategies invading the resident strategy

This section is based mostly on [4].

The first condition that needs to be satisfied in order for the resident strategy to be evolutionary stable, is that selection must oppose all mutant strategies invading the resident strategy. This is the case when a single mutant has a lower fitness than the rest of the population, who use the resident strategy. We look further into when this condition is satisfied. Specifically, we prove the following theorem.

Theorem 3.2. *In case of a symmetric game, selection opposes all mutant strategies invading the resident strategy if and only if the resident strategy is a strict Nash strategy.*

This theorem states that in case of a symmetric game, condition (1) of Definition 3.1 is equivalent to the condition that the resident strategy should be a strict Nash strategy. We prove the above theorem.

Proof. First we show the direct implication. Given a symmetric game. Let M be an arbitrary mutant strategy and let R be the resident strategy. In this case, after the mutation has appeared, there is one individual who plays strategy M in a population consisting otherwise of individuals playing R. Therefore, we have $X_1 = N - 1$ and $X_2 = 1$. Assume that selection opposes M invading R. That is, the single M mutant has a lower fitness than the rest of the population, who play R. Hence

$$F(M \,|\, N-1, 1) < F(R \,|\, N-1, 1). \tag{3.3}$$

By filling in the corresponding fitness-functions (3.1) and (3.2), equation (3.3) is equivalent to

$$\frac{N-1}{N-1} \cdot c + \frac{1-1}{N-1} \cdot d < \frac{N-2}{N-1} \cdot a + \frac{1}{N-1} \cdot b,$$

hence

$$c < \frac{N-2}{N-1} \cdot a + \frac{1}{N-1} \cdot b. \tag{3.4}$$

By using symmetry of the game, that is $b = c$, we obtain that

$$b < \frac{N-2}{N-1} \cdot a + \frac{1}{N-1} \cdot b,$$

hence

$$\frac{N-2}{N-1} \cdot b < \frac{N-2}{N-1} \cdot a,$$

and thus

$$b < a.$$

Therefore equation (3.3) is equivalent to

$$a > b. \tag{3.5}$$

This means that the resident strategy R is a unique best response to itself. Since M was an arbitrary mutant strategy, it follows that R is a unique best response to itself against all mutant strategies M'. Hence by definition, the resident strategy R is a strict Nash strategy in the complete strategy space. This proves the direct implication. For the reverse implication, one can simply go through the steps of the proof given, but in reverse order. This concludes the proof of theorem 3.2. □

This interpretation of the first condition of Definition 3.1 seems to make sense, since the entire concept of a Nash equilibrium is precisely based on the idea of a single deviation. Usually this would mean that the strategy choice of a single individual has a fading effect on the population's average strategy. However, in a finite population this is not necessarily true. Looking at condition (3.4), where the mutant's payoff is compared to the payoff of another, non-mutant, individual inside the population, the effect of the deviation is brought into the equation. If the weight of the deviant fades, then we do find ourselves in the case where the strategy choice of a single individual has a fading effect on the population's average. However, the same is true if the term that reflects the deviation cancels out for some other reason, such as symmetry of the payoff function.

3.2 Selection opposes all mutant strategies replacing the resident strategy

This section is based mostly on [1] and [4].

The second condition that needs to be satisfied in order for the resident strategy to be evolutionary stable, is that selection must oppose all mutant strategies replacing the resident strategy. Even though a single mutant strategy has a lower fitness than the regular type, the frequency dependent Moran process can still favor the fixation of this mutant strategy due to drift. In order to avoid this, we must have that the probability that a single mutant will invade and take over the entire population is smaller than the inverse population size. We look further into the derivation of this condition and when this condition is satisfied.

First of all, the frequency-dependent Moran process allows us to introduce a parameter $\omega \in [0, 1]$, which determines the contribution of the game's payoff to fitness. Therefore, this parameter measures the intensity of selection. We replace the fitness functions (3.1) and (3.2) by the modified fitness functions

$$F_\omega(R \mid X_1, X_2) = 1 - \omega + \omega \cdot F(R \mid X_1, X_2). \tag{3.6}$$

$$F_\omega(M \mid X_1, X_2) = 1 - \omega + \omega \cdot F(M \mid X_1, X_2). \tag{3.7}$$

If $\omega = 0$, then the game's payoff does not contribute to fitness at all and it follows that $F_0(R \mid X_1, X_2) = F_0(M \mid X_1, X_2) = 1$. This is the case of neutral evolution. There is no selection, but random drift can replace R by M. If $\omega = 1$, then equations (3.6) and (3.7) are equivalent to the fitness functions (3.1) and (3.2). Therefore, the game's payoff completely determines fitness and selection is considered strong.

We can determine the probability ρ that a single individual who uses an arbitrary mutant strategy M can invade and take over a population consisting otherwise of individuals playing the resident strategy R. This probability is also referred to as the fixation probability of M. We can consider this as the probability that a stochastic process starting from the state where $X_2 = 1$ reaches the absorbing state $X_2 = N$ rather than $X_2 = 0$, which is given by (S. Karlin and H. M. Taylor [1])

$$\rho = 1 \Bigg/ \left(1 + \sum_{k=1}^{N-1} \prod_{X_2=1}^{k} \frac{F_\omega(R \mid X_1, X_2)}{F_\omega(M \mid X_1, X_2)} \right). \tag{3.8}$$

Note that in case of neutral evolution, hence if $\omega = 0$, it follows from equations (3.6) and (3.7) that $F_0(R \mid X_1, X_2) = F_0(M \mid X_1, X_2) = 1$. Therefore, equation (3.8) reduces to $\rho = 1/N$ and we obtain a neutral threshold equal to $1/N$.

Now, the idea of Nowak et al. [4] for evaluating the stability of the resident strategy in an evolutionary sense is to compare the fixation probability ρ of a single M mutant under the frequency-dependent Moran process to this neutral threshold. If $\rho < 1/N$, we say that selection opposes a mutant strategy M replacing the resident strategy R. If $\rho > 1/N$, we say that selection favors a mutant strategy M replacing the resident strategy R.

In practice, it can require much work to calculate fixation probabilities. However, under weak selection ($\omega << 1$) we show that the following theorem holds.

Theorem 3.3 (Nowak et al. [4]). *For a symmetric game, given a finite population of size $N \geq 3$ and sufficiently weak selection ($\omega << 1$), it holds true that selection favors a mutant strategy replacing the resident strategy, that is $\rho > \frac{1}{N}$, if and only if*

$$d + c > 2a.$$

Remark We notice that weak selection is indeed the natural case. In reality, the overall fitness of an individual does not depend exclusively on its communicative strategy, but on other cultural and biological traits as well. Therefore, the payoff of the game is only a small part that is added to the background fitness of an individual, hence $\omega << 1$.

For the proof of Theorem 3.3, we make use of the following lemma.

Lemma 3.4 (Nowak et al. [4]). *For a given finite population of size N and sufficiently weak selection ($\omega << 1$) it holds true that $\rho > \frac{1}{N}$ if and only if*

$$d(N-2) + c(2N-1) > b(N+1) + a(2N-4).$$

We shall not prove Lemma 3.4. However, for more details one can view Nowak et al.[4]. Below we prove Theorem 3.3.

Proof. First, we show the direct implication. Let M be an arbitrary mutant strategy and let R be the resident strategy. Assume that selection favors the mutant strategy M replacing the resident strategy R. Hence $\rho > \frac{1}{N}$. By Lemma 3.4 it follows that

$$d(N-2) + c(2N-1) > b(N+1) + a(2N-4). \tag{3.9}$$

Since we assume a symmetric payoff function, that is $b = c$, equation (3.9) reduces to

$$d(N-2) + c(2N-1) > c(N+1) + a(2N-4),$$

that is

$$d(N-2) + c(N-2) > a(2N-4). \tag{3.10}$$

For $N \geq 3$ we can divide equation (3.10) by $(N-2)$ to obtain

$$d + c > 2a. \tag{3.11}$$

This proves the direct implication. For the reverse implication, one can simply go through the steps of the proof given, but in reverse order. This concludes the proof of theorem 3.3. □

Note, that we can rearrange the terms in equation (3.11) to obtain that

$$d - a > a - c.$$

Hence,

$$\rho > \frac{1}{N} \quad \Leftrightarrow \quad d - a > a - c. \tag{3.12}$$

We can interpret this as follows. A single mutant strategy M that appears in a population consisting otherwise of the resident strategy R, can reach fixation with a probability greater than $1/N$ if and only if its disadvantage in relative fitness against the resident type, that is $c - a < 0$, is outweighed by a payoff advantage that the mutant strategy has against itself relative to the payoff that the resident strategy has against itself, that is $d - a$. Hence, a mutant strategy that appears in a population consisting otherwise of the resident type, can reach fixation with a probability greater than the inverse population size, even though it has a strict disadvantage in relative fitness against the resident type.

Chapter 4

The Language Game

This chapter is based on [3], [5], [6] and [8].

The Language Game is a sender-receiver game in which we examine associations between signals and objects. Consider two players, one in the role of the sender (speaker) and one in the role of the receiver (listener). Both players are able to produce a number of signals (sounds). The sender wants to transfer information about a certain object. An object can be anything that can be referred to, such as items, events or other individuals. In order to do this, the sender uses one of the signals it possesses. If the chosen signal evokes the image of the particular object in the receiver, then communication has been successful and both players receive a relative high positive payoff. If communication has not been successful, hence if the chosen signal does not evoke the image of the particular object in the receiver, then both players receive a relative low positive payoff (could be 0 as well). In the next section, we formulate the corresponding model based on [3]. After that, we discuss various charasteristics of the game.

4.1 The Model

This section is based mostly on [5].

Consider a group of individuals. Suppose there are n objects that can potentially become the subject of communication. Furthermore, suppose there are m available signals to describe these n objects. A strategy in role of the sender can be represented by a $n \times m$ probability matrix P, also referred to as the active matrix, in which rows represent objects and columns represent signals. The entries p_{ij} denote the probability that a sender uses signal j in order to refer to object i. Since a sender will always link a signal to an object, it follows that the rows of P add up to 1. In a similar way we represent a strategy in the role of the receiver by a $m \times n$ probability matrix Q, also referred to as the passive matrix, in which rows represent signals and columns represent objects. The entries q_{ji} denote the probability that a receiver infers object i when observing signal j. Since a receiver will always link an object to a signal, it follows that the rows of Q add up to 1 as well. Hence we obtain the strategies

$$P = \begin{pmatrix} p_{11} & \dots & p_{1j} & \dots & p_{1m} \\ \vdots & & \vdots & & \vdots \\ p_{i1} & \dots & p_{ij} & \dots & p_{im} \\ \vdots & & \vdots & & \vdots \\ p_{n1} & \dots & p_{nj} & \dots & p_{nm} \end{pmatrix} \in \mathcal{P}_{n\times m}, \quad Q = \begin{pmatrix} q_{11} & \dots & q_{1i} & \dots & q_{1n} \\ \vdots & & \vdots & & \vdots \\ q_{j1} & \dots & q_{ji} & \dots & q_{jn} \\ \vdots & & \vdots & & \vdots \\ q_{m1} & \dots & q_{mi} & \dots & q_{mn} \end{pmatrix} \in \mathcal{Q}_{m\times n},$$

where

$$\mathcal{P}_{n\times m} = \left\{ P \in \mathbb{R}_+^{n\times m} \;\middle|\; \sum_{j=1}^{m} p_{ij} = 1 \;\forall i \right\}, \tag{4.1}$$

$$\mathcal{Q}_{m\times n} = \left\{ Q \in \mathbb{R}_+^{m\times n} \;\middle|\; \sum_{i=1}^{n} q_{ji} = 1 \;\forall j \right\}. \tag{4.2}$$

4.2 The potential of communication

This section is based mostly on [3] and [5].

When a sender and a receiver wish to interact with each other, communication depends on the degree to which sender and receiver understand each other. Hence, when the sender uses certain signals to transfer information about a particular object i, it is crucial that the receiver evokes the image of this object when observing these signals from the sender. Therefore, the probability that a sender with strategy P and a receiver with strategy Q communicate correctly about object i is given by

$$\sum_{j=1}^{m} p_{ij} q_{ji}.$$

If we take the sum over all possible n objects of communication, we obtain to what extent a sender with strategy P and a receiver with strategy Q are able to communicate correctly about all n objects. We call this the *potential of communication* of the pair (P, Q), given by

$$\pi(P, Q) := \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} q_{ji}.$$

Note that the potential of communication between the strategies P and Q equals the trace of the $n \times n$ matrix PQ, that is, the sum of the elements on the main diagonal of the square matrix PQ, since

$$\mathrm{tr}(PQ) = \sum_{i=1}^{n} (PQ)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} q_{ji}.$$

Assuming that correct communication is mutually beneficial, we assign the potential of communication to the payoff that both the sender and the receiver gain from interacting with each other. Hence

$$\pi_P(P, Q) = \pi(P, Q) = \pi_Q(P, Q). \tag{4.3}$$

Therefore, the strategy sets for the sender and receiver, as denoted in (4.1) and (4.2) respectively, together with the payoff function in (4.3) form a two-player asymmetric game with a symmetric payoff function. That is, the two players have different strategy sets but receive the same payoff out of their interaction.

4.3 Best responses

This section is based mostly on [5] and [6].

The strategy that yields the highest payoff for a player, considering the other player's strategy to be left unchanged, is called a best response to the other player's strategy. In the asymmetric game, for a receiver, the set of best responses to a strategy $P \in \mathcal{P}_{n\times m}$ of the sender, is given by

$$B(P) = \left\{ Q \in \mathcal{Q}_{m\times n} \,|\, \mathrm{tr}(PQ) \geq \mathrm{tr}(PQ') \,\text{for all}\, Q' \in \mathcal{Q}_{m\times n} \right\}.$$

And for a sender, the set of best responses to a strategy $Q \in \mathcal{Q}_{m\times n}$ of the receiver, is given by

$$B(Q) = \left\{ P \in \mathcal{P}_{n\times m} \,|\, \mathrm{tr}(PQ) \geq \mathrm{tr}(P'Q) \,\text{for all}\, P' \in \mathcal{P}_{n\times m} \right\}.$$

We give the following two propositions without proof as a characterisation of best-response properties. For a proof of Propositions 4.1 and 4.2 we refer to Pawlowitsch [6]. These propositions are needed for the proof of Proposition 5.1 later on.

Proposition 4.1 (Pawlowitsch [6], Lemma 2)**.** *Let* $P \in \mathcal{P}_{n\times m}$ *and* $Q \in \mathcal{Q}_{m\times n}$*.*

(a) If $Q \in B(P)$*, then*

$$\sum_{i \in argmax_i(p_{ij})} q_{ji} = 1 \quad \text{and} \quad q_{ji} = 0 \quad \text{for all } i \notin argmax_i(p_{ij}).$$

(b) If $P \in B(Q)$*, then*

$$\sum_{j \in argmax_j(q_{ji})} p_{ij} = 1 \quad \text{and} \quad p_{ij} = 0 \quad \text{for all } j \notin argmax_j(q_{ji}).$$

If p_{ij} *is the unique maximal element in the* j*-th column of* P*, then, for any* $Q \in B(P)$ *it follows that* $q_{ji} = 1$*. By interchanging the roles of* P *and* Q *we obtain the same result vice versa.*

Proposition 4.2 (Pawlowitsch [6], Lemma 4)**.** *Let* $P \in \mathcal{P}_{n\times m}$ *and* $Q \in \mathcal{Q}_{m\times n}$*.*

(a) If $Q \in B(P)$*, then*

$$q_{ji} \neq 0 \quad \Rightarrow \quad p_{ij} = \max_i(p_{ij}) \quad \Rightarrow \quad p_{ij} \neq 0 \quad or \quad p_{i'j} = 0 \quad \forall i'.$$

(b) If $P \in B(Q)$*, then*

$$p_{ij} \neq 0 \quad \Rightarrow \quad q_{ji} = \max_j(q_{ji}) \quad \Rightarrow \quad q_{ji} \neq 0 \quad or \quad q_{j'i} = 0 \quad \forall j'.$$

4.4 The symmetric game

This section is based mostly on [5].

When we research language as a social phenomena, we do not find individuals that are either only senders or only receivers. An individual shall adopt the role of sender or receiver depending on its relative position in different situations of interaction with other individuals. Therefore, we assume that interaction is pairwise and that individuals adopt both social roles with equal probabilities. Formally, this corresponds to symmetrizing the asymmetric game. A strategy of the symmetric game, then, is a pair of an active and a passive matrix

$$L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n},$$

also called a *language* in the sequal.

4.5 Communication payoff

This section is based mostly on [3] and [5].

Consider two individuals I_1 and I_2 who use languages $L_1 = (P_1, Q_1)$ and $L_2 = (P_2, Q_2)$ respectively. For individual I_1, the entries $p_{ij}^{(1)}$ denote the probability of making sound j when seeing object i, whereas the entries $q_{ji}^{(1)}$ denote the probability of thinking of object i when hearing sound j. For individual I_2 these probabilities are given by the entries $p_{ij}^{(2)}$ and $q_{ji}^{(2)}$. Hence, the probability that individual I_1 successfully conveys information about object i to individual I_2 is given by $\sum_{j=1}^{m} p_{ij}^{(1)} q_{ji}^{(2)}$. By summing this probability over all n objects, we obtain to what extent I_1 is able to transfer information to I_2, that is $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij}^{(1)} q_{ji}^{(2)}$. The total payoff for interaction between I_1 and I_2 is then constructed by taking the average sum of I_1's ability to convey information to I_2, and I_2's ability to convey information to I_1. Therefore the communication payoff is given by

$$f(L_1, L_2) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(p_{ij}^{(1)} q_{ji}^{(2)} + p_{ij}^{(2)} q_{ji}^{(1)} \right) = \frac{1}{2}\mathrm{tr}(P_1 Q_2) + \frac{1}{2}\mathrm{tr}(P_2 Q_1). \tag{4.4}$$

Since both individuals are considered once in the role of speaker and once in the role of listener together with symmetry of the payoff function in the asymmetric game (4.3), it follows that there is symmetry in equation (4.4), that is $f(L_1, L_2) = f(L_2, L_1)$.

It follows that the Language Game is a doubly symmetric game, that is, a symmetric game with a symmetric payoff function.

4.6 Maximum communication payoff

This section is based mostly on [5] and [8].

Logically, the payoff function is maximal when two individuals speak the same language $L = (P, Q)$. It then follows from equation (4.4) that

$$f(L, L) = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} q_{ji}. \tag{4.5}$$

Since the rows of both P and Q add up to 1 by definition, we know that

$$f(L, L) \leq \min\{m, n\}.$$

The following theorem shows that if $m = n$, then the maximum payoff is reached if and only if $L = (P, Q)$ is an efficient proto-language, that is, P is a permutation matrix and $Q = P^T$.

Theorem 4.3 (P.E. Trapa and M.A. Nowak [8], Lemma 3.3)**.** *Suppose* $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$. *Then* $f(L, L) = n$ *if and only if* L *is an efficient proto-language.*

Proof. First we prove the direct implication. Assume that $f(L, L) = n$. By equation (4.5) we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}q_{ji} = n.$$

Then it must hold true that

$$\sum_{j=1}^{n} p_{ij}q_{ji} = 1 \quad \text{for all} \quad i = 1, 2, ..., n.$$

Fix $i^* \in \{1, 2, ..., n\}$. Since the rows of P add up to 1 by definition, it holds true that $\sum_{j=1}^{n} p_{i^*j} = 1$. Hence we must have $q_{ji^*} = 1$ for all j such that $p_{i^*j} \neq 0$. Fix $j^* \in \{1, 2, ..., n\}$ such that $q_{j^*i^*} = 1$. Since the rows of Q add up to 1 by definition as well, it holds true that $\sum_{i=1}^{n} q_{j^*i} = 1$. But $q_{j^*i^*} = 1$, thus $q_{j^*i} = 0$ for all $i \neq i^*$. Therefore $p_{ij^*} = 0$ for all $i \neq i^*$. Hence, in P, it follows that there is only one non-zero entry per column, while each row adds up to 1. Since $m = n$, this can only be the case if P is a permutation matrix. Furthermore, since for each i^* it holds true that $q_{ji^*} = 1$ for all j such that $p_{i^*j} \neq 0$ and since $q_{j^*i} = 0$ for all $i \neq i^*$, it follows that $Q = P^T$. Hence, $L = (P, P^T)$ with P a permutation matrix. This proves the direct implication.

For the reverse implication. Let $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ be an efficient proto-language, that is, P is a permutation matrix and $Q = P^T$. We determine $f(L, L)$. By equation (4.5) we know that

$$f(L, L) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}q_{ji}.$$

Fix $i^* \in \{1, 2, ..., n\}$. By definition we have that the rows of P add up to 1, hence $\sum_{j=1}^{n} p_{i^*j} = 1$. Since L is an efficient proto-language it follows that $q_{ji^*} = 1$ if and only if $p_{i^*j} = 1$. Hence, $q_{ji^*} = 1$ for every j such that $p_{i^*j} \neq 0$. Therefore $\sum_{j=1}^{n} p_{i^*j}q_{ji^*} = 1$. Since i^* was arbitrary, we obtain that

$$f(L, L) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}q_{ji} = \sum_{i=1}^{n} 1 = n.$$

This proves the reverse implication. □

4.7 Nash Equilibria

This section is based mostly on [5], [6] and [8].

In Game Theory, a Nash equilibrium is a set of strategies of the game in which no player could benefit from changing its own strategy, considering the strategies of the other players to be left unchanged. Hence, all strategies involved have to be best responses to each other. Since the Language Game is a symmetric game, a Nash equilibrium is obtained when a strategy is a best response to itself. In the sequel, we call such strategies Nash strategies. That is, L is a Nash strategy if

$$f(L, L) \geq f(L, L') \quad \text{for all} \quad L' \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}.$$

It is called a strict Nash strategy if it is a unique best response to itself. Hence, if

$$f(L, L) > f(L, L') \quad \text{for all} \quad L' \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}.$$

Since the Language Game is a symmetric game, it holds true that in a Nash strategy $L = (P, Q)$, the strategies played in both roles have to be best responses to each other. That is, $P \in B(Q)$ and $Q \in B(P)$. In a strict Nash strategy P has to be a unique best response to Q and the other way around. We shall give a formal characterisation of the conditions that P and Q have to satisfy in order for $L = (P, Q)$ to be a (strict) Nash strategy.

Theorem 4.4 (P.E. Trapa and M.A. Nowak [8], Theorem 5.1)**.** *A language $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$, where neither P nor Q has a zero-column, is a Nash strategy if and only if there exist real numbers $p_1, ..., p_m \in (0, 1)$ and $q_1, ..., q_n \in (0, 1)$ such that*

(i) for each $j = 1, 2, ..., m$, the j-th column of P has its entries drawn from $\{0, p_j\}$,

(ii) for each $i = 1, 2, ..., n$, the i-th column of Q has its entries drawn from $\{0, q_i\}$,

(iii) for all $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, the entry p_{ij} in P is strict positive if and only if the entry q_{ji} in Q is strict positive.

For the proof of Theorem 4.4 we refer to P.E. Trapa and M.A. Nowak [8]. From theorem 4.4 we deduce that in a Nash strategy, both synonymy as homonymy is allowed. That is, multiple signals may be used to describe one object and multiple objects may be associated with the same signal. However, syno-homonymy, where two or more signals denote exactly the same multiple objects, is only allowed when this happens in equal proportions.

Theorem 4.5 (P.E. Trapa and M.A. Nowak [8], Theorem 3.1)**.** *A language* $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ *is a strict Nash strategy in the complete strategy space if and only if*

(i) the number of signals is equal to the number of objects, that is $n = m$*,*

(ii) the active matrix P *is a permutation matrix, that is a square binary matrix with exactly one entry equal to 1 in every row and every column and all other entries equal to 0,*

(iii) the passive matrix Q *is the transpose of the active matrix* P*, that is* $Q = P^T$*.*

This theorem states that the only strict Nash strategies of the Language Game are efficient proto-languages, where one object can be referred to by exactly one signal and where each signal refers to exactly one object. Note from Theorem 4.3 that in this case, the maximum payoff is reached. Therefore, any player would be worse off by playing a different strategy, as desired. Since a permutation matrix has exactly one 1 entry in each row and each column, while all other entries are 0, it follows that for a given number n of objects and signals, there are $n!$ permutations and hence $n!$ such matrices. Since $Q = P^T$ is fixed, we obtain that there are $n!$ strict Nash strategies in a Language Game where we consider n objects and signals. We end this section by providing a proof of theorem 4.5.

Proof. First we prove by contradiction that if a language $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ satisfies conditions $(i), (ii)$ and (iii), then L is a strict Nash strategy. By conditions (i) and (ii), we have that P is a $n \times n$-permutation matrix. By condition (iii), we have $Q = P^T$. Since we can rearrange the order of the signals in our language, we are allowed to assume without loss of generality that $P = P^T = Q = \mathbb{I}_n$, the $n \times n$-identity matrix. Now, assume that L is not a strict Nash strategy. Hence, there is a language $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ with $L' \neq L$ such that

$$f(L, L) \leq f(L, L'). \tag{4.6}$$

Since L satisfies conditions $(i), (ii)$ and (iii), it holds true that L is an efficient proto-language. Hence by Theorem 4.3 we know that $f(L, L) = n$. Therefore equation (4.6) is equivalent to

$$n \leq f(L, L').$$

From this it follows that

$$\begin{aligned} n &\leq \frac{1}{2}\big(\operatorname{tr}(PQ') + \operatorname{tr}(P'Q)\big) \\ &\leq \frac{1}{2}\big(\operatorname{tr}(\mathbb{I}_n Q') + \operatorname{tr}(P'\mathbb{I}_n)\big) && \text{(since } P = Q = \mathbb{I}_n\text{)} \\ &\leq \frac{1}{2}\big(\operatorname{tr}(Q') + \operatorname{tr}(P')\big). \end{aligned}$$

Hence

$$\frac{1}{2}\big(\operatorname{tr}(Q') + \operatorname{tr}(P')\big) \geq n. \tag{4.7}$$

Since the rows of both P' and Q' must add up to 1 by definition and since $n = m$, it follows that $\operatorname{tr}(P') \leq n$ and $\operatorname{tr}(Q') \leq n$. Therefore

$$\frac{1}{2}\big(\operatorname{tr}(Q') + \operatorname{tr}(P')\big) \leq n. \tag{4.8}$$

By equations (4.7) and (4.8) we obtain that

$$n \leq \frac{1}{2}\big(\operatorname{tr}(Q') + \operatorname{tr}(P')\big) \leq n,$$

hence

$$\frac{1}{2}\big(\operatorname{tr}(Q') + \operatorname{tr}(P')\big) = n.$$

Since $\operatorname{tr}(P') \leq n$ and $\operatorname{tr}(Q') \leq n$, we must have that $\operatorname{tr}(Q') = \operatorname{tr}(P') = n$. Hence, all entries on the main diagonal of both P' as Q' must be 1, while each row adds up to 1. It follows from this that all entries of P' and Q' that are not on the main diagonal must equal 0. Therefore $P' = Q' = \mathbb{I}_n$ and we obtain that $L' = L$. This is a contradiction. Hence, the assumption that L is not a strict Nash strategy was false. We conclude that L is indeed a strict Nash strategy. This proves the first half of Theorem 4.5.

Next we would like to show that if $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then L satisfies conditions $(i), (ii)$ and (iii). For this, we start by proving some lemma's.

Lemma 4.6 (P.E. Trapa and M.A. Nowak [8], Lemma 3.5)**.** *If $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then P is binary.*

Proof. We use a proof by contradiction. Assume that $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy and that P is not binary. Since all rows of P add up to 1 by definition, while P is not binary, it follows that there is at least one row with more than one non-zero entry. Assume this is the case for the i-th row in P and consider the i-th column in Q. Let j be an index such that q_{ji} is a maximal entry in the i-th column of Q. We define a new matrix $P' \neq P$ as follows.

$$p'_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j \\ 0 & \text{if } k = i \text{ and } l \neq j \\ p_{kl} & \text{else} \end{cases}$$

Hence, P' is similar to P, except for the i-th row where P has more than one non-zero entry, whereas P' only has one non-zero entry p_{ij} equal to 1. Note that the rows of P' add up to 1 as well. By construction, we have $\text{tr}(P'Q) \geq \text{tr}(PQ)$. On the other hand, since $L = (P, Q)$ is a strict Nash strategy, we have that $\text{tr}(PQ) > \text{tr}(P'Q)$ by definition. Hence, we have reached a contradiction. Our assumption that P is not binary was therefore false. We conclude that if $L = (P, Q)$ is a strict Nash strategy, then P is binary. □

Lemma 4.7 (P.E. Trapa and M.A. Nowak [8], Lemma 3.5)**.** *If $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then $Q = P^T$.*

Proof. Assume that $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy. Then by Lemma 4.6 we know that P is binary. Since the rows of P add up to 1 by definition, it follows that there is exactly one 1 entry in each row of P, while all other entries equal 0. Fix a row i of P and let j be the unique index such that $p_{ij} = 1$. From our analysis in the proof of Lemma 4.6 we know that from this, it follows that q_{ji} is a maximal entry in the i-th column of Q. Suppose that this is not a unique maximal entry in the i-th column of Q. We show that this leads to a contradiction. Since q_{ji} is not a unique maximal entry in the j-th column of Q, there is a $h \neq j$ such that $q_{ji} = q_{hi}$. We define a new matrix $P' \neq P$ as follows.

$$p'_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = h \\ 0 & \text{if } k = i \text{ and } l \neq h \\ p_{kl} & \text{else} \end{cases}$$

Hence, P' is similar to P, except for the i-th row where P has a 1 at entry p_{ij} (and 0 elsewhere), whereas P' has a 1 at entry p'_{ih} (and 0 elsewhere). By construction, we have $\text{tr}(P'Q) = \text{tr}(PQ)$. On the other hand, by definition we have $\text{tr}(PQ) > \text{tr}(P'Q)$, since $L = (P, Q)$ is a strict Nash strategy. Hence, we have reached a contradiction. Our assumption that q_{ji} is not a unique maximal entry in the i-th column of Q was therefore false. We conclude that if p_{ij} is the unique maximal entry equal to 1 in the i-th row of P, then q_{ji} is the unique maximal entry in the i-th column of Q. Since i was arbitrary and since all rows of Q must add up to 1 by definition, it follows that the unique maximal entry in each column of Q is equal to 1 as well. Thus we obtain that $Q = P^T$. We conclude that if $L = (P, Q)$ is a strict Nash strategy, then $Q = P^T$. □

Lemma 4.8 (P.E. Trapa and M.A. Nowak [8], Lemma 3.6)**.** *If $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then P is a binary matrix with no two non-zero entries in the same column.*

Proof. We use a proof by contradiction. Assume that $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy. Then by Lemma 4.6 it follows that P is a binary matrix. Suppose to the contrary that there is a column in P that has two non-zero entries. Let j be this column in P and assume that the entries p_{ij} and p_{hj} are non-zero. Since P is binary, it follows that $p_{ij} = p_{hj} = 1$. Consider the entries q_{ji} and q_{jh} in Q. By Lemma 4.7 we know that $Q = P^T$, therefore $q_{ji} = q_{jh} = 1$ as well. We define a new matrix $Q' \neq Q$ as follows

$$q'_{lk} = \begin{cases} 1+\epsilon & \text{if } l = j \text{ and } k = i \\ 1-\epsilon & \text{if } l = j \text{ and } k = h \\ q_{lk} & \text{else} \end{cases}$$

with $\epsilon > 0$ small. Hence, Q' is similar to Q, except for the j-th row where in Q a sufficiently small positive number ϵ is subtracted from q_{jh} and added to q_{ji} to obtain Q'. By construction, we have $\mathrm{tr}(PQ) = \mathrm{tr}(PQ')$. On the other hand, since $L = (P, Q)$ is a strict Nash strategy, we have $\mathrm{tr}(PQ) > \mathrm{tr}(PQ')$ by definition. Hence, we have reached a contradiction. Our assumption that there is a column in P that has two non-zero entries was therefore false. We conclude that if $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then P is a binary matrix with no two non-zero entries in the same column. □

Now, let $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ be a strict Nash strategy. Then it follows from Lemma 4.6 that all entries of P are equal to 0 or 1. Since the rows of P must add up to 1 by definition, we obtain that there is exactly one 1 entry in each row of P, while all other entries equal 0. From Lemma 4.8 it follows that there can also be only one 1 entry in each column of P, while all other entries equal 0. Hence, there is exactly one entry in each row and each column equal to 1. That is, each object is associated by a unique signal and vice versa. Therefore, the number of objects and signals must be identical and we obtain that $n = m$. From this it follows that P is a square binary matrix, hence P is a permutation matrix. By Lemma 4.7 we also have that $Q = P^T$. Therefore, we conclude that if $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is a strict Nash strategy, then $n = m$ and P is a permutation matrix and $Q = P^T$. This proves the second half of Theorem 4.5, which concludes the proof of Theorem 4.5. □

Chapter 5

Evolutionary stable strategies in the Language Game

This chapter is based on [5] and [6].

We would like to determine whether there are evolutionary stable strategies in the Language Game. Hence, if there are strategies $L \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ that satisfy conditions (1) and (2) from Definition 3.1. In section 3.1 we derived that condition (1) is equivalent with being a strict Nash strategy. From Theorem 4.5 we know that the only strict Nash strategies of the Language Game in the complete strategy space are efficient proto-languages. Hence, strategies $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ where P is a permutation matrix and $Q = P^T$. Therefore, the only strategies in the Language Game that might fit the definition for evolutionary stability in a finite population, are efficient proto-languages. We are left with determining if efficient proto-languages satisfy condition (2) of definition 3.1 as well. That is, the probability that a single mutant will invade and take over the entire population, is smaller than the neutral threshold. This is indeed the case. More precisely, the following implications can be shown.

Proposition 5.1 (Pawlowitsch [5], Proposition 1)**.** *Let $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ be a strategy. Under the frequency-dependent Moran process with weak selection ($\omega << 1$), for $N \geq 3$, the following implications hold true*

1. *if L is an efficient proto-language, hence if L is a strict Nash strategy in the complete strategy space $\mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$, then there is no $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$, with $L' \neq L$, such that $\rho' \geq \frac{1}{N}$.*
2. *if L is not an efficient proto-language, then there is some $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$, with $L' \neq L$, such that $\rho' > \frac{1}{N}$.*

Since we know by Theorem 4.5 that all efficient proto-languages are strict Nash strategies and vice versa, Proposition 5.1 tells us the following. The first implication states that if $L \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ is an efficient proto-language, then condition (2) of Definition 3.1 is satisfied, since all mutant strategies $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ have a fixation probability lower than the neutral threshold, that is $\rho' < \frac{1}{N}$. From this it directly follows that all efficient proto-languages are evolutionary stable strategies of the game. The second implication states that if $L \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ is not an efficient proto-language, then there exists a mutant strategy $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ which selection can favor replacing L. Hence, in this case L does not satisfy condition (2) of Definition 3.1 on being an evolutionary stable strategy.

In order to prove Proposition 5.1, we need the notion of neutral stability.

Definition 5.2 (Pawlowitsch [5], Definition 1)**.** A strategy $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ is neutrally stable if and only if

(i) it is a Nash strategy, and if

(ii) whenever $f(L, L) = f(L', L)$ for some $L' = (P', Q') \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$, then

$$f(L, L') \geq f(L', L').$$

We conclude this chapter by proving the implications of proposition 5.1.

Proof. First we prove implication (1). We use a proof by contradiction. Let $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ be an efficient proto-language and assume that there is a strategy $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ such that $\rho' \geq \frac{1}{N}$. Since $\rho' \geq \frac{1}{N}$, it immediately follows from Theorem 3.3 that

$$f(L', L') + f(L', L) \geq 2f(L, L). \tag{5.1}$$

We show why this leads to a contradiction. Since L is an efficient proto-language, that is, P is a permutation matrix and $Q = P^T$, it follows from Theorem 4.3 that L exploits the maximum payoff. Hence,

$$f\left(L, L\right) = \operatorname{tr}\left(P, P^T\right) = n \geq \operatorname{tr}\left(\tilde{P}, \tilde{Q}\right) = f\left(\tilde{L}, \tilde{L}\right),$$

for all $\tilde{L} \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$. In particular it follows that

$$f(L, L) \geq f(L', L'). \tag{5.2}$$

Also, since L is a strict Nash strategy, it follows that L is a unique best reply to itself. Therefore, $f(L,L) > f(L,\tilde{L})$ for all $\tilde{L} \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$. From which it follows that

$$f(L,L) > f(L,L'). \tag{5.3}$$

Hence, we obtain that

$$\begin{aligned} f(L',L') + f(L',L) &= f(L',L') + f(L,L') && \text{(by symmetry)} \\ &\leq f(L,L) + f(L,L') && \text{(by (5.2))} \\ &< f(L,L) + f(L,L) && \text{(by (5.3))} \\ &= 2f(L,L). \end{aligned}$$

Thus

$$f(L',L') + f(L',L) < 2f(L,L),$$

which is a contradiction with equation (5.1). Therefore, the assumption that there exists a strategy $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ such that $\rho' \geq \frac{1}{N}$ was false. We conclude that if $L \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ is an efficient proto-language, then there is no $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$, with $L' \neq L$, such that $\rho' \geq \frac{1}{N}$. This proves (1).

Next, we prove implication (2). We use a proof by exhaustion. Let $L \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ where L is not an efficient proto-language. Consider the following three cases

1. L is not a Nash strategy,
2. L is a Nash strategy, but not a neutrally stable strategy,
3. L is a neutrally stable strategy.

We show for all cases that there is some $L' \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$, with $L' \neq L$, such that $\rho' > \frac{1}{N}$.

1. Assume $L = (P,Q)$ is not a Nash strategy. Then either Q is not a best response to P, or P is not a best response to Q, or both. Without loss of generality, assume $Q \notin B(P)$. Then by definition there exists a $Q' \in \mathcal{Q}_{n\times n}$ with $Q' \neq Q$ such that

$$\mathrm{tr}(PQ) < \mathrm{tr}(PQ').$$

Consider the strategy $L' = (P, Q')$. Then $L' \neq L$ and $\mathrm{tr}(PQ') > \mathrm{tr}(PQ)$. Therefore $\frac{3}{2} \cdot \mathrm{tr}(PQ') > \frac{3}{2} \cdot \mathrm{tr}(PQ)$, from which we obtain that

$$\mathrm{tr}\left(PQ'\right) + \left(\frac{1}{2}\mathrm{tr}\left(PQ'\right) + \frac{1}{2}\mathrm{tr}(PQ)\right) > 2 \cdot \mathrm{tr}(PQ).$$

Hence,

$$f(L', L') + f(L', L) > 2f(L, L).$$

By Theorem 3.3, this implies that $\rho' > \frac{1}{N}$.

2. Assume $L = (P, Q)$ is a Nash strategy, but not a neutrally stable strategy. Since L is a Nash strategy, but not a neutrally stable strategy, it follows by Definition 5.2 on being a neutrally stable strategy that there is some $L' = (P', Q') \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ such that

$$f(L, L) = f(L', L), \tag{5.4}$$

and

$$f(L, L') < f(L', L'). \tag{5.5}$$

From this it follows that $L' \neq L$ and

$$\begin{aligned} f(L', L') + f(L', L) &> f(L, L') + f(L', L) && \text{(by (5.5))} \\ &= f(L', L) + f(L', L) && \text{(by symmetry)} \\ &= 2f(L', L) \\ &= 2f(L, L). && \text{(by (5.4))} \end{aligned}$$

Hence

$$f(L', L') + f(L', L) > 2f(L, L),$$

which by Theorem 3.3 implies that $\rho' > \frac{1}{N}$.

3. The last case requires some more steps. We start with the following lemmas.

Lemma 5.3 (Pawlowitsch [6], Theorem 1)**.** *Let $L = (P, Q) \in \mathcal{P}_{n\times m} \times \mathcal{Q}_{m\times n}$ be a Nash strategy. Then L is a neutrally stable strategy if and only if*

(i) at least one of the matrices P or Q has no zero-column, and

(ii) neither P nor Q has a column with multiple maximal elements that are strictly between 0 and 1.

For a proof of Lemma 5.3, we refer to Pawlowitsch [6].

Lemma 5.4. *If $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ is a neutrally stable strategy, but not an efficient proto-language, then at least one of the matrices P or Q has at least one row that has at least two entries strictly between* 0 *and* 1*, such that they are unique maximal entries in their respective columns.*

Proof of Lemma 5.4. Assume $L = (P, Q) \in \mathcal{P}_{n\times n} \times \mathcal{Q}_{n\times n}$ is a neutrally stable strategy, but not an efficient proto-language. Then it follows from Lemma 5.3 condition (i) that at least one of the matrices P or Q has no zero-column. Without loss of generality we can assume that P has no zero-column. By condition (ii) of Lemma 5.3 we also know that P does not have a column with multiple maximal elements that are strictly between 0 and 1. Together these conditions imply that for each column of P, its maximum must satisfy one of the following cases, that is, the maximum is

1. unique and equal to 1,
2. unique, but not equal to 1,
3. equal to 1, but not unique.

In case all columns of P have a unique maximal element which equals 1, it follows that P is a permutation matrix since all rows of P must add up to 1 by definition. Furthermore, since $L = (P, Q)$ is a neutrally stable strategy and hence a Nash strategy, it follows that Q is a best response to P. Thus, by Proposition 4.1 it follows that $Q = P^T$. However, this would mean that $L = (P, Q)$ is precisely an efficient proto-language, which is a contradiction with the assumption that L is not an efficient proto-language. Therefore, there must be at least one column in P that has a unique maximal element unequal to 1, or, that has multiple maximal elements equal to 1. We evaluate both cases.
Suppose P has a column with a unique maximal element unequal to 1. Let j' be this column and let $p_{ij'} \in (0, 1)$ be the unique maximal element. Notice that $p_{ij'} \neq 0$ otherwise j' would be a zero-column which contradicts the assumption that P has no zero-column. Since $L = (P, Q)$ is a neutrally stable strategy and therefore a Nash strategy, it follows that Q is a best response to P. Then, by Proposition 4.1 we obtain that $q_{j'i} = 1$ and $q_{j'i}$ is a maximal element in column i of Q, hence $j' \in \text{argmax}_j(q_{ji})$. Furthermore, since P is a best response to Q as well, we know from Proposition 4.1 that

$$\sum_{j \in \text{argmax}_j(q_{ji})} p_{ij} = 1,$$

while $p_{ij'} \in (0, 1) \neq 1$. Therefore, there exists at least one $j^* \in \text{argmax}_j(q_{ji})$ with $j^* \neq j'$ such that $p_{ij^*} \neq 0$ and we obtain that $q_{j'i}$ is not a unique maximal element in column i of Q. Notice that from this it follows that row i in P has at least two entries $p_{ij'}$

and p_{ij^*} that are strictly between 0 and 1. We show that all such entries are unique maximal entries in their respective columns. Since $q_{ji} \neq 0$ for all $j \in \text{argmax}_j(q_{ji})$ and since Q is a best response to P, it follows from Proposition 4.2 that p_{ij} is a maximal element in the corresponding column j in P. Again by Proposition 4.2 this implies for all $j \in \text{argmax}_j(q_{ji})$ that either $p_{ij} \neq 0$, or, $p_{i'j} = 0$ for all i'. Since we assumed that P has no zero-column, we must have that $p_{ij} \neq 0$ for all $j \in \text{argmax}_j(q_{ji})$. Again, since

$$\sum_{j \in \text{argmax}_j(q_{ji})} p_{ij} = 1,$$

while $p_{ij'} \in (0, 1) \neq 1$, it follows for all $j \in \text{argmax}_j(q_{ji})$ that $p_{ij} \neq 1$ as well. Therefore $0 < p_{ij} < 1$ for all $j \in \text{argmax}_j(q_{ji})$. Now, we obtain for all $j \in \text{argmax}_j(q_{ji})$ that $p_{ij} \in (0, 1)$ while p_{ij} is a maximal element in column j in P. However, since L is a neutrally stable strategy, Lemma 5.3 tells us that P has no column with multiple maximal elements that are strictly between 0 and 1. Thus it follows that p_{ij} is the unique maximal element in column j for all $j \in \text{argmax}_j(q_{ji})$, as desired. Therefore, the matrix P has at least one row that has at least two entries strictly between 0 and 1, such that they are unique maximal entries in their respective columns. This proves the lemma in this case.

Suppose P has a column with multiple maximal elements equal to 1. Then there must be two rows in P that both have a entry equal to 1 in the same column position. Since all rows of P must add up to 1 by definition, it follows that all other entries in both rows must equal 0. However $P \in \mathcal{P}_{n \times n}$, that is, P is a square matrix. Since by assumption P has no zero-column, it follows that there must be at least two columns that have a unique maximal element strictly between 0 and 1. Hence, we find ourselves in the same case as above, for which we have already proven the statement of the lemma to be true. Of course, the roles of P and Q can be interchanged. This concludes the proof of Lemma 5.4. □

Now, let $L = (P, Q) \in \mathcal{P}_{n \times n} \times \mathcal{Q}_{n \times n}$ be a neutrally stable strategy. By the initial assumption in condition (2) of Proposition 5.1, we have that L is not an efficient proto-language. Therefore, it follows by Lemma 5.5 that at least one of the matrices P or Q has at least one row that has at least two entries strictly between 0 and 1, such that they are unique maximal entries in their respective columns. Assume without loss of generality that P is this matrix. Let i_1 be the row that has at least two entries strictly between 0 and 1, such that they are unique maximal entries in their respective columns in P. Since $L = (P, Q)$ is a neutrally stable strategy and therefore a Nash strategy, we know that Q is a best response to P and vice versa. Therefore Proposition 4.1 tells us that $q_{ji_1} = 1$ whenever $p_{i_1 j} \neq 0$. Since there are at least two entries strictly between 0 and 1 in the i_1-th row of P, it follows that there are also at least two entries equal to 1 in the i_1-th column of Q. As a consequence, since all rows of Q must add up to 1 by definition, there are now two cases

1. the matrix Q has at least one zero-column, or,
2. the matrix Q has at least two columns with unique maximal elements strictly between 0 and 1.

For both cases, we can construct a potential mutant language $L' = (P', Q')$ for which it holds true that its fixation probability ρ' is greater than the neutral threshold. In case Q has a zero column, say the i_0-th column, we construct $P' \neq P$ from P as follows. Replace the i_0-th row with a vector that has a 1 at some position j such that $j \in \text{argmax}_j(q_{ji_1})$ and 0 elsewhere, and, replace the i_1-th row with a vector that has a 1 at some position $j' \in \text{argmax}_j(q_{ji_1})$, with $j' \neq j$, and 0 elsewhere. Furthermore, we construct $Q' \neq Q$ from Q by swapping the entries in the j-th row, such that $q'_{ji_0} = 1$ and $q'_{ji} = 0$ for all $i \neq i_0$, leaving the other rows unchanged. For clarification of this construction, we give the following example.

$$\text{if} \qquad P = \begin{array}{cc} & \begin{array}{ccc} j' & j & \end{array} \\ \begin{array}{c} i_1 \\ i_0 \\ \end{array} & \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array}, \qquad Q = \begin{array}{cc} & \begin{array}{ccc} i_1 & i_0 & \end{array} \\ \begin{array}{c} j' \\ j \\ \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array},$$

$$\text{then} \qquad P' = \begin{array}{cc} & \begin{array}{ccc} j' & j & \end{array} \\ \begin{array}{c} i_1 \\ i_0 \\ \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}, \qquad Q' = \begin{array}{cc} & \begin{array}{ccc} i_1 & i_0 & \end{array} \\ \begin{array}{c} j' \\ j \\ \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}.$$

In case Q has two columns with unique maximal elements strictly between 0 and 1, say the i_2-th and i_3-th column, it follows that these unique maximal entries must appear in one and the same row j^*. This is because all rows must add up to 1 by definition and the entries in the i_1-th column are already 1 for $j \in \text{argmax}_j(q_{ji_1})$. We can construct $P' \neq P$ from P in the same way as above by considering $i_2 \equiv i_0$. Similar, we can construct $Q' \neq Q$ from Q as above considering $i_2 \equiv i_0$, but repeating this step for the j^*-th row, that is, $q'_{j^*i_3} = 1$ and $q'_{j^*i} = 0$ for all $i \neq i_3$. Again, we give an illustration for clarification.

$$\text{if} \qquad P = \begin{array}{c} \\ i_1 \\ i_2 \\ i_3 \end{array} \overset{\begin{array}{ccc} j' & j & j^* \end{array}}{\begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}}, \qquad Q = \begin{array}{c} \\ j' \\ j \\ j^* \end{array} \overset{\begin{array}{ccc} i_1 & i_2 & i_3 \end{array}}{\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \beta & 1-\beta \end{pmatrix}},$$

$$\text{then} \qquad P' = \begin{array}{c} \\ i_1 \\ i_2 \\ i_3 \end{array} \overset{\begin{array}{ccc} j' & j & j^* \end{array}}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}, \qquad Q' = \begin{array}{c} \\ j' \\ j \\ j^* \end{array} \overset{\begin{array}{ccc} i_1 & i_2 & i_3 \end{array}}{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}.$$

In both cases, it can easily be seen by a comparison in payoff, that the payoff advantage that the constructed mutant language $L' = (P', Q')$ has against itself relative to the payoff that $L = (P, Q)$ has against itself, is higher than the payoff disadvantage that the constructed mutant language L' has against L relative to the payoff that L has against itself. For more details on the payoff comparison in general, one can view Pawlowitsch [5]. From this it follows by Theorem 3.3 and equation (3.12) that $\rho' > 1/N$, as desired.

Hence, in all three cases we have seen that if L is not an efficient proto-language, then there exists some $L' \neq L$ such that $\rho' > 1/N$. The proof of the second part of proposition 5.1 is therefore finished. This concludes the proof of Proposition 5.1. □

Chapter 6

A few examples

This chapter is based on [5].

In this chapter we give a few examples of competing languages in a finite population of size N, discussing their stability in an evolutionary sense. In all cases, we assume that there are only two languages competing with each other. We look at a number of situations.

6.1 Two neutrally stable strategies competing.

First we consider the case where there are two neutrally stable strategies in the complete strategy space competing with each other. Assume that both strategies have the same synonymy and homonymy in their active and passive matrices, but with different probabilities. Let these strategies be given by

$$L_1 = (P_1, Q_1) = \left[\begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-\beta & \beta \end{pmatrix} \right],$$

$$L_2 = (P_2, Q_2) = \left[\begin{pmatrix} 1-\gamma & \gamma & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-\delta & \delta \end{pmatrix} \right],$$

where $\alpha, \beta, \gamma, \delta \in (0,1)$, with $\alpha \neq \gamma$ and $\beta \neq \delta$. Notice that both L_1 and L_2 satisfy conditions $(i), (ii)$ and (iii) of Theorem 4.4 for being a Nash strategy. With some simple calculations, we easily see that $f(L_1, L_1) = f(L_1, L_2) = f(L_2, L_1) = f(L_2, L_2) = 2$.

Hence, we obtain the corresponding payoff matrix

$$\begin{array}{c} \\ L_1 \\ L_2 \end{array} \begin{array}{c} \begin{array}{cc} L_1 & L_2 \end{array} \\ \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \end{array}$$

By observing the above payoff matrix, we notice the following. First of all, there are no strict Nash strategies in this game, since no strategy is a unique best response to itself. Hence, by Theorem 3.2, both L_1 and L_2 do not satisfy condition (1) of Definition 3.1 on evolutionary stability. Furthermore, it holds true that

$$f(L_1, L_1) + f(L_1, L_2) = 2 + 2 = 4 = 2 \cdot 2 = 2 \cdot f(L_2, L_2).$$

and

$$f(L_2, L_2) + f(L_2, L_1) = 2 + 2 = 4 = 2 \cdot 2 = 2 \cdot f(L_1, L_1),$$

Hence, we are in the case of neutral evolution. The fixation probabilities of both L_1 and L_2 are equal to the neutral threshold. That is, $\rho_1 = 1/N$ and $\rho_2 = 1/N$. Therefore, drift is the only evolutionary force at work.

6.2 An efficient proto-language competing with a simple Nash strategy.

Next we consider the case where there is an efficient proto-language competing with a simple Nash strategy. Let these strategies be given by

$$L_1 = (P_1, Q_1) = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right],$$
$$L_2 = (P_2, Q_2) = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right].$$

Notice that L_1 is indeed an efficient proto-language. With some simple calculations, we easily see that $f(L_1, L_1) = 3$ and $f(L_1, L_2) = f(L_2, L_1) = f(L_2, L_2) = 2$. Hence, we obtain the corresponding payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} L_1 & L_2 \end{array} \\ \begin{array}{c} L_1 \\ L_2 \end{array} & \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \end{array}$$

By observing the above payoff matrix, we can say the following. First of all,

$$f(L_2, L_2) = 2 = f(L_2, L_1),$$

hence L_2 is a best response to itself, but not a unique best response. The strategy L_1 is a best response to L_2 as well. Therefore, a single L_1 mutant that appears in a population consisting otherwise of L_2 has the same fitness as the resident language L_2. On the other hand, since

$$f(L_1, L_1) = 3 > 2 = f(L_1, L_2),$$

it follows that L_1 is a unique best response to itself and thus a strict Nash strategy of this game. Therefore, the fitness of L_1 will already be greater than the fitness of L_2 once there appears a second L_1 mutant in the population, as can be seen when calculating the fixation probabilities of L_1 and L_2. Since

$$f(L_1, L_1) + f(L_1, L_2) = 3 + 2 = 5 > 4 = 2 \cdot 2 = 2 \cdot f(L_2, L_2),$$

it follows by Theorem 3.2 that the fixation probability of L_1 is greater than the neutral threshold, that is $\rho_1 > 1/N$. On the other hand

$$f(L_2, L_2) + f(L_2, L_1) = 2 + 2 = 4 < 6 = 2 \cdot 3 = 2 \cdot f(L_1, L_1),$$

hence it follows by Theorem 3.2 that the fixation probability of L_2 is not greater than the neutral threshold, that is $\rho_2 \leq 1/N$. We conclude that in a population where the strategy L_2 is set as the resident language, selection favors the mutant strategy L_1 replacing L_2. Whereas in a population where the strategy L_1 is set, selection opposes the mutant strategy L_2 replacing L_1. Since L_1 is a strict Nash strategy as well, it follows that L_1 satisfies both conditions (1) and (2) of Definition 3.1 on evolutionary stability. Therefore L_1 is an evolutionary stable strategy. This general result for efficient proto-languages of course has already been proven in Theorem 4.5 together with Proposition 5.1.

6.3 An efficient proto-language competing with a neutrally stable strategy.

In this case there is an efficient proto-language competing with a neutrally stable strategy. Let these strategies be given by

$$L_1 = (P_1, Q_1) = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right],$$

$$L_2 = (P_2, Q_2) = \left[\begin{pmatrix} \alpha & 1-\alpha & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1-\beta & \beta \end{pmatrix} \right].$$

Where $\alpha, \beta \in (0,1)$. Notice that L_1 is indeed an efficient proto-language and that L_2 satisfies conditions $(i), (ii)$ and (iii) of Theorem 4.4 for being a Nash strategy. With some simple calculations, we easily see that $f(L_1, L_1) = 3$ and $f(L_2, L_2) = 2$ and $f(L_1, L_2) = f(L_2, L_1) = 1 + \frac{\alpha}{2} + \frac{\beta}{2}$. Hence, we obtain the corresponding payoff matrix

$$\begin{array}{c} \\ L_1 \\ L_2 \end{array} \begin{array}{c} \begin{array}{cc} L_1 & \qquad L_2 \end{array} \\ \begin{pmatrix} 3 & 1 + \frac{\alpha}{2} + \frac{\beta}{2} \\ 1 + \frac{\alpha}{2} + \frac{\beta}{2} & 2 \end{pmatrix} \end{array}$$

By observing the above payoff matrix, we can say the following. First of all, since $\alpha, \beta < 1$ it follows that

$$f(L_2, L_2) = 2 = 1 + \frac{1}{2} + \frac{1}{2} > 1 + \frac{\alpha}{2} + \frac{\beta}{2} = f(L_2, L_1),$$

hence L_2 is a unique best response to itself. Also

$$f(L_1, L_1) = 3 > 2 = 1 + \frac{1}{2} + \frac{1}{2} > 1 + \frac{\alpha}{2} + \frac{\beta}{2} = f(L_1, L_2),$$

hence L_1 is a unique best response to itself as well. Therefore, both L_1 and L_2 are strict Nash strategies of this game. Hence, a single L_1 mutant that appears in a population consisting otherwise of L_2 has a strictly lower fitness than the resident language L_2, and vice versa. Therefore, we cannot distinguish between L_1 and L_2 based on the fitness of a single mutant. We determine the fixation probabilities of L_1 and L_2. Since $\alpha, \beta > 0$, it follows that

$$f(L_1, L_1) + f(L_1, L_2) = 3 + 1 + \frac{\alpha}{2} + \frac{\beta}{2} > 4 = 2 \cdot 2 = 2 \cdot f(L_2, L_2).$$

Hence it follows from Theorem 3.2 that the fixation probability of L_1 is greater than the neutral threshold, that is $\rho_1 > 1/N$. Also, since $\alpha, \beta < 1$, we have

$$f(L_2, L_2) + f(L_2, L_1) = 2 + 1 + \frac{\alpha}{2} + \frac{\beta}{2} < 3 + \frac{1}{2} + \frac{1}{2} = 4 < 6 = 2 \cdot 3 = 2 \cdot f(L_1, L_1),$$

hence it follows by Theorem 3.2 that the fixation probability of L_2 is not greater than the neutral threshold, that is $\rho_2 \leq 1/N$. We conclude that in a population where the strategy L_2 is set as the resident language, selection favors the mutant strategy L_1 replacing L_2. However, in a population where the strategy L_1 is set as the resident language, selection opposes the mutant strategy L_2 replacing L_1. Hence, we can distinguish L_1 from L_2 by comparing their fixation probabilities. It follows that only L_1 satisfies conditions (1) and (2) of Definition 3.1 on evolutionary stability. Therefore L_1 is an evolutionary stable strategy. Again, this general result for efficient proto-languages has already been proven in Theorem 4.5 together with Proposition 5.1.

6.4 Two efficient proto-languages competing.

At last, we consider two efficient proto-languages competing with each other. Let these strategies be given by

$$L_1 = (P_1, Q_1) = \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right],$$

$$L_2 = (P_2, Q_2) = \left[\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right].$$

Notice that L_1 and L_2 are indeed efficient proto-languages. With some simple calculations, we easily see that $f(L_1, L_1) = 3$ and $f(L_1, L_2) = f(L_2, L_1) = 0$ and $f(L_2, L_2) = 3$. Hence, we obtain the corresponding payoff matrix

$$\begin{array}{c} \\ L_1 \\ L_2 \end{array} \begin{array}{c} \begin{array}{cc} L_1 & L_2 \end{array} \\ \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \end{array}$$

By observing the above payoff matrix, we notice the following. First of all,

$$f(L_1, L_1) = 3 > 0 = f(L_1, L_2),$$

hence L_1 is a unique best reply to itself. Also

$$f(L_2, L_2) = 3 > 0 = f(L_2, L_1),$$

hence L_2 is a unique best reply to itself as well. Therefore, both L_1 as L_2 are strict Nash strategies of this game. Of course, this property of efficient proto-languages was already proven in Theorem 4.5. Hence, a single L_1 mutant that appears in a population consisting otherwise of L_2 has a strictly lower fitness than the resident language L_2, and vice versa. Hence, we cannot distinguish L_1 from L_2 based on the fitness of a single mutant. We determine the fixation probabilities of L_1 and L_2. Since

$$f(L_1, L_1) + f(L_1, L_2) = 3 + 0 = 3 < 6 = 2 \cdot 3 = 2 \cdot f(L_2, L_2),$$

it follows by Theorem 3.2 that the fixation probability of L_1 is not greater than the neutral threshold, that is $\rho_1 \leq 1/N$. Also, since

$$f(L_2, L_2) + f(L_2, L_1) = 3 + 0 = 3 < 6 = 2 \cdot 3 = 2 \cdot f(L_1, L_1),$$

it follows by Theorem 3.2 that the fixation probability of L_2 is not greater than the neutral threshold either, that is $\rho_2 \leq 1/N$. We conclude that in a population where the strategy L_1 is set as the resident language, selection opposes the mutant strategy L_2 replacing L_1, and vice versa. Therefore we cannot distinguish L_1 and L_2 based on the comparison in fixation probabilities either. It follows that both L_1 and L_2 satisfy conditions (1) and (2) of Definition 3.1 on evolutionary stability. Therefore L_1 and L_2 are both evolutionary stable strategies. Again, this general result for efficient proto-languages has already been proven in Theorem 4.5 together with Proposition 5.1.

Chapter 7

Conclusions

We conclude this thesis with an interpretation of the results from the previous chapters. First of all we mention that in the modeling framework used here, what we call 'the payoff of the game' is just short hand for adding a fitness component to a birth-and-death process that introduces an element of frequency-dependent selection in addition to drift (C. Pawlowitsch, *Finite populations choose an optimal language*, [5]). In chapter 3 we introduced the definition of evolutionary stability in a finite population as proposed by Nowak et al. [4]. Definition 3.1 stated that a resident strategy is evolutionary stable if selection opposes all mutant strategies invading the resident strategy, and, if selection opposes all mutant strategies replacing the resident strategy. The first condition is satisfied when a single mutant that appears in a population consisting otherwise of the resident type, has a lower fitness relative to the rest of the population. In section 3.1 we proved with theorem 3.2 that in case of a symmetric game, like the Language game, this condition is equivalent with the resident strategy being a strict Nash strategy of the game. In section 4.7 we proved with Theorem 4.5 that the only strict Nash strategies of the Language Game are strategies for which $m = n$, the active matrix P is a binary matrix and the passive matrix Q is the transpose of P, hence $Q = P^T$. That is to say that the only strict Nash strategies of the Language Game are efficient proto-languages. Therefore, only efficient proto-languages seem to be eligible as an evolutionary stable language. The second condition of Definition 3.1 is satisfied if the fixation probability of all mutant strategies is smaller than the neutral threshold, which equals the inverse population size. Of course, in a finite population it follows that all mutant strategies have some positive probability to reach fixation due to the chances of random drift. In section 3.2 we proved with Theorem 3.3 that in case of a symmetric game, like the Language Game, and under the assumption of weak selection, which is the natural case, a mutant strategy has a fixation probability greater than the neutral threshold if and only if its disadvantage in relative fitness against the resident type is outweighed by a payoff advantage that the mutant strategy has against itself relative to the payoff that the resident language has against itself. Therefore, a mutant strategy that has a strict fitness disadvantage against the resident type can still reach fixation. With this, we proved in chapter 5 with Proposition 5.1 that under the frequency-dependent Moran process with weak selection, for a finite population size $N \geq 3$ and $m = n$, if the resident strategy is an efficient proto-language, then all mutant strategies have a fixation probability smaller than the neutral threshold. Hence, selection opposes all mutant strategies replacing the resident strategy in this case. Furthermore, Proposition 5.1 tells us that if the resident language is not an efficient proto-language, then there exists a mutant language with a fixation probability greater than the neutral threshold. Therefore, selection favors this mutant strategy replacing the resident strategy in case the resident strategy is not an efficient proto-language. Hence, efficient proto-languages are both protected and favored by selection. From Theorem 3.2 together with Theorem 4.5 and Proposition 5.1 it therefore follows that under a frequency-dependent Moran process with weak selection, the only strategies of the Language Game that satisfy both conditions of Definition 3.1 on evolutionary stability, are efficient proto-languages.

Bibliography

[1] S. Karlin and H. M. Taylor. A first course in stochastic processes. *Elsevier*, 1975.

[2] M.A.Nowak. Evolutionary dynamics: Exploring the equations of life. *The Belknap Press of Harvard University Press*, 2006.

[3] M.A. Nowak, J.B. Plotkin, and D.C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162, 1999.

[4] M.A. Nowak, A. Sasaaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, 2004.

[5] C. Pawlowitsch. Finite populations choose an optimal language. *Journal of Theoretical Biology*, 249:606–616, august 2007.

[6] C. Pawlowitsch. Why evolution does not always lead to an optimal signaling game. *Games and Economic Behavior*, 63:203–226, 2008.

[7] P.D. Taylor and L. Jonker. Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, 40:145–156, 1978.

[8] P.E. Trapa and M.A. Nowak. Nash equilibria for an evolutionary language game. *Journal of Mathematical Biology*, 41:172–178, september 2000.