INVESTIGATION OF EMERGENCE OF DIVERSITY IN LANGUAGE: PARENT ORIENTED TEACHER SELECTION

by

İbrahim Çimentepe

B.S., Computer Engineering, Boğaziçi University, 2012

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2016

ACKNOWLEDGEMENTS

This study would not have been possible without the support of many people. Many thanks to my thesis supervisor, Assoc. Prof. Dr. Haluk Bingöl, who encouraged me and read my numerous revisions and helped me to make some sense of the confusion.

I would also like to thank Dr. Suzan Üsküdarlı and Assoc. Prof. Dr. Şule Gündüz Öğüdücü for their participation in my thesis jury and their precious feedbacks.

Thanks to Boğaziçi University for providing me with the opportunity to complete this study and thanks to many friends I met here Suat Akbulut, Yekta Said Can, Mehmet Akif Ersoy and Uzay Çetin who helped with their encouragement and contributions.

My deepest gratitude is to my family. I want to thank my father, Mustafa Çimentepe who always provided me with his wisdom in many difficulties my all life. I also would like to thank my brothers and sisters; Enes, Tuba, Sümeyye and Esra who always stayed by my side. Finally, I would like to dedicate this study to my mother Ummahan and my grandfather Ibrahim whom I wish to be together in this milestone. Their love and hope for me will always be a responsibility on my shoulders.

ABSTRACT

INVESTIGATION OF EMERGENCE OF DIVERSITY IN LANGUAGE: PARENT ORIENTED TEACHER SELECTION

In this work, we investigated the emergence of diversity in language. Fundamental game theoretical model for emergence of language has been established and used in several studies. In these studies most basic assumption is that understanding everyone in the population gives evolutionary advantage to an agent. However, being in interaction with everyone is not achievable in reality where various limited resources such as memory and physical availability bind agents. Thus population should be organized in a way that each agent can interact only some percentage of population. In our model, agents select their teachers, who are responsible for the transmission of language, from neighbors of their parents, which forms a social network in population. Our findings include that emerged number of different languages in population is related to the percentage size of neighborhood and is independent of population size. Furthermore, we observe that seeking language-wise similarity in teacher selection makes no evolutionary difference in emergence of language, instead of seeking physical closeness. As a result, we see this study as an important contribution towards understanding the emergence of diversity in languages.

ÖZET

DİLDEKİ FARKLILIKLARIN OLUŞUMUNUN İNCELENMESİ: EBEVEYN MERKEZLİ ÖĞRETMEN SEÇİMİ

Bu çalışmamızda dilin oluşumundaki farklılıkları araştırdık. Dilin ortaya çıkışı ile ilgili temel, oyun teorisi modelleri bir çok çalışmada ortaya koyulmuş ve kullanılmıştır. Bu çalışmalarda en sık karşılaşılan varsayım toplumdaki herkes ile anlaşabilmenin evrimsel üstünlük sağlayacağı yönündedir. Fakat, toplumdaki herkesle iletişim içinde olmak gerçekte çok mümkün görünmemektedir, çünkü hafıza ve yerleşim gibi sınırlamalar bireyleri kısıtlamaktadır. Dolayısıyla toplum, her bireyin toplumun sadece belli bir kısmı ile iletişim halinde olacağı şekilde organize edilmelidir. Bizim modelimizde, bireyler dilin aktarımından sorumlu olan öğretmenleri ebeveynlerinin komşularından seçmektedir ve bu komşuluk ilişkileri toplumda bir sosyal ağ oluşmasına neden olmaktadır. Bulgularımızdan bazıları, komşu çevresinin büyüklüğü ile toplam oluşan farklı dil sayısının, tüm toplumun büyüklüğünden bağımsız olarak, ilişkili olduğu yönündedir. Üstelik, komşu seçiminde dil yakınlığı gözetmenin, fiziksel yakınlık gözetmeye göre herhangi bir evrimsel avantaj sağlamadığını gözlemledik. Sonuç olarak, biz bu çalışmayı dildeki farklılıkların oluşumunu anlamak adına önemli bir katkı olarak görüyoruz.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

$\mathbf{A}$	$M \times S$ association matrix
$a_{\mu x}$	Entry of association matrix
$\mathbf{D}$	$S \times M$ association matrix
$d_{x\mu}$	Entry of decoding matrix
$\mathbf{E}$	$M \times S$ encoding matrix
$e_{\mu x}$	Entry of encoding matrix
F_i	Fitness in Base Model
$F(i \to j)$	Comprehension from agent i to j
$F(i \leftrightarrow j)$	Mutual comprehension between agents i and j
$I(\mathcal{B} \to \mathcal{C})$	Comprehension from community $\mathcal{B}$ to $\mathcal{C}$
$I(\mathcal{B} \leftrightarrow \mathcal{C})$	Inter community comprehension between communities $\mathcal{B}$ and $\mathcal{C}$
$I(\mathbb{P}_K)$	Average inter community comprehension of partition $\mathbb{P}_K$
I^*	Average inter community comprehension in the best partition for given population $\mathcal{P}$
K	Cluster size
K^*	Community count in the best partition for given $\mathcal{P}$
$\mathcal{M}$	Set of meanings
M	Number of meanings
N	Number of agents
$\mathcal{P}$	Set of all agents
$\mathbb{P}_K$	Partition of population $\mathcal{P}$ with K clusters
$\overline{\mathbb{P}_K}$	Best partition for given K and $\mathcal{P}$
Q	Sampling size
R	Size of imitation set
r	Ratio of imitation set size to population size.
$\mathcal{S}$	Set of signals
S	Number of signals
$\mathcal{V}_i$	Neighborhood and imitation set of agent i

$W(\mathcal{C})$	Within community comprehension inside community $\mathcal{C}$
$W(i \to \mathcal{C})$	Language-wise closeness of agent i to community $\mathcal{C}$
$W_{\mathrm{r}}(\mathcal{C})$	Within community comprehension inside community $\mathcal{C}$ with random language
$W(\mathcal{P})$	Overall comprehension in population $\mathcal{P}$
$W(\mathbb{P}_K)$	Average within community comprehension of partition $\mathbb{P}_K$
W^*	Average within community comprehension in the best partition for given $\mathcal{P}$
$\pi(i,t)$	Probability of selecting one agent t within $\mathcal{V}_i$

LIST OF ACRONYMS/ABBREVIATIONS

1D	One Dimensional
NP	Non-deterministic Polynomial

1. INTRODUCTION

The common properties of animal and human communication have been understood partially. Certain signaling formations, which form a *language*, are used to transfer thoughts between individuals. It is now thought that the signals used to communicate can change through learning [1].

Language learning process may take certain forms between individuals. Trial and error is one possible form of learning, where individuals try to gain best benefit from communication, compensating their mistakes. On the other hand, learning can be in the form of *inheritance*, that is certain members such as parents, caregivers and teachers are responsible for the transmission of language. Imitation is central to language in all forms: individuals may imitate each other or prefer to imitate parent or role model members in population [2]. This raises the question of how the language is inherited or learned. Who should be selected as role models in community for the next generation? And which imitation strategies may lead to the emergence of language that is shared locally or across a population?

In this study, we examine the imitation mechanisms for the emergence of language where agents do not seek a descriptive relationship between meanings and signals as in complex human languages. Our imitation model is a simple one, called proto-imitation. In this model, agents are expected to imitate the agents, whom they select as teachers, unconditionally.

1.1. Motivation

The evolution of language has been studied as an ethological subject in numerous researches [3–8]. The general understanding is that humans have much more sophisticated conversational, auditory and conceptual abilities than animals [9]. However, how these traits are developed is still a question of research. As Reference [3–8] and many other ethological studies suggest, for gaining a deeper understanding of the origin of

language, Darwinian Frameworks are needed. On the other hand, studies addressing the origin of language with mathematical models are rare [10]. In Chapter 2, we will address the established frameworks in detail.

In this study, we make use of and extend the mathematical framework established in Reference [10]. Specifically, we investigate an alternate aspect of the origin of language: the emergence of diversity in language. This subject is very popular indeed; it is even addressed in the well-known story of the Tower of Babel. According to this story, the people who spoke the same language once scattered all around the world so that they could no longer understand each other, as their languages changed. This is a quite straight-forward example to the origin of diversity and is in part explained in Krakauer's study [11]. Understandably, if there are strict limitations that keep the individuals apart, the population would behave as several sub-populations. Thus, each of these sub-structures would possibly result in emergence of different languages, since language evolves and changes [1].

However, this type of organization in the population, where there is a strict separation between groups, is not always the case. Using a model that there are less strict social restrictions would be more suitable to understand the components of the emergence of diversity in languages.

In previous studies [10,12,13], a game theoretical approach has been proposed. In the games modeled, mutual comprehension is translated into biological fitness. Thus, individuals, who are successful in communication, produced a greater number of offspring. Their findings included that cooperation between individuals might be crucial for the evolution of language [13].

Nevertheless, there is more to evolution than just doing what is best for the population. Within populations, certain social structures are likely to have an important effect on how language evolves. *Social structures* can emerge as formations that are born out of individuals' predefined or asymmetrical priorities regarding whom to favor within the population. Basic reasons to these priorities might be kinship and

geographical closeness between individuals.

In populations, there are three different transmission types between agents [14]. Firstly, there can be a transmission between agents within a certain generation. This type of transmission is named *horizontal transmission*. Secondly, it is assumed that there is a certain transmission from parent to offspring, which is called *vertical transmission*. This transmission can be genetic or simply cultural, where parent teaches her experience to offspring. A parent's role in this transmission, which we refer as teaching, may be executed by someone close to a parent as well, as long as the transmission is not random. This last transmission type, where it is across generation lines but from non-parents, is called *oblique transmission*. These teacher members may be selected one of close kin. Thus, genetic similarity provides one mechanism for teacher selection process. Alternatively, candidates who may serve this purpose might be found in the neighborhood.

When individuals interact with others whom they find beneficial, they may decide to maintain contact leading to further learning. Moreover, if there is going to be a non-biological form of transmission to offspring, these individuals would more likely be selected to serve this purpose. This selection procedure is in fact Moran process of population genetics, where agents of a certain generation are replaced with the next generation of agents who are placed in the region close to their parents. As a result, neighbors in the next generation are expected to be similar to each other [15].

To sum up, we contribute towards understanding the emergence of diversity in language extending the mathematical models established in Reference [10]. The question of how diverse languages can be and underlying factors causing diversity have been investigated analytically in several studies [14, 16]. In this study we revisit these factors and propose a model to test the hypothesis that selection of teachers, who are neighbor to parent in terms of various criteria, causes the diversity in language. That is, we consider oblique and vertical transmissions. As a result, we find a significant correlation between the size of neighborhood and the number of different languages emerged.

In Chapter 2, we explain the related literature in detail. In Chapter 3, we provide the details of necessary models that we use as a background in this study. In Chapter 3, we first introduce the language model and the language-wise similarity measures between agents and communities. We explain the population dynamics, after then we introduce the community detection algorithm that is used in detection of sub-languages. We propose various models for teacher selection mechanisms, extending already established selection mechanism in Reference [13]. In Chapter 4, we introduce these models and accompanied simulations to test our hypothesis. In Chapter 5, there is a comparative discussion regarding the simulation results. Then in Chapter 6, we introduce the examples of possible extensions to this study. We simulated the models for wide range of parameters and results of additional simulations are shared in appendix sections Appendix A, Appendix B, Appendix C, Appendix D and Appendix E.

2. LITERATURE

Evolution by natural selection can be summed up in three Darwinian principles: (i) there is inheritance from parent to offspring, which causes correlation between parent and offspring, (ii) individuals are different from each other and (iii) some individuals leave more offspring as a result of having certain evolutionary advantage [14]. Many researchers [1, 3–7] eagerly apply these Darwinian frameworks to their investigation of the origin of language, assuming that ability to communicate is an evolutionary advantage. Some other researchers [17,18] are a little reluctant applying these predominantly used frameworks. For example, Reference [17] argues that language is not evolved for communication, but evolved for complex thoughts. It claims that if language evolved for communication, it would not suffer from ambiguity.

In Reference [9], these two approaches are investigated in detail. These two different views questioning the origin of language have one thing in common: learning a human language requires learning a very complex set of rules and signals. It requires highly developed organs for memorizing, auditory analysis and talking. What they disagree on is how these difficulties are overcome. Do humans need to be intelligent far superior to animals to speak language or are there other evolutionary factors affecting the process? In this study, we investigate the origin of language, making use of the Darwinian approaches extensively.

Language is composed of certain signals that have various meanings depending on the context, which they are used. In Reference [4] it is claimed that human language is a major transition in human evolution. We are as humans pretty much able to communicate using highly complex structures [16]. Yet to understand the origin of language, much more simplistic version of human language, which animals use, has been studied.

2.1. Ethological Studies

Ethology is the scientific study of animal behavior, viewing behavior as an evolutionary adaptive trait. Certain difference between human and animal languages is that animals have very simple signal systems, whereas humans express their ideas explicitly using complex structures [5]. In Reference [3] the signal system that is being made used of in animal communication is explained in detail. According to it, the language of animals in nature appears to use signals that (i) have certain, fixed references, (ii) are not ambiguous and (iii) lack syntax. These signals can be produced with various intentions such as food sharing and group cohesion in monkeys [3], alarm call in frogs or mating signal in crickets [6]. Another example is that velvet monkeys give different alarm calls to different classes of predators [5], that is when different stimuli presented, different signal is present. Squirrel monkeys, rhesus monkeys, gorillas, Japanese macaques, frogs, toads, canaries, zebra finches, bats, the sage grouse, mantis shrimp, the domestic chicken, tanagers, European blackbirds, honeybees and many other species are investigated for their communication systems [7].

In Reference [3], it is assumed that the instinct to produce certain signals is hardwired to animals and is activated whenever the appropriate stimulus occurs. That is, in nature, it usually is seen that animals broadcast certain signal whenever an appropriate circumstances for that particular signal arise [6]. Furthermore, different from human conversations, there usually is no certain recipient for these signals. And sender, does not usually seek necessity or effectiveness in this behavior. Nevertheless, signal is usually correctly understood by the recipients [3]. There must be an inherent knowledge regarding the meaning that is held by both the recipient and the sender. This raises the question how this inherent behavior evolved?

Imitation plays an important role in evolution of language in animals. In Reference [8] the presence of imitation has been investigated in chimpanzees. It is concluded that there is an imitation mechanism but it is significantly different from human children regarding their language learning ability. Chimpanzees imitated their partners mechanically, which cannot be characterized as *conversational competence*, it was more

like a mechanical, rote imitation. *Rote imitation* is the type of imitation mechanism that stems from the intention to copy another. This type of imitation is vital to learning process. Animals are exposed to certain situation and receive an accompanied signal. First receivers adjust subsequent actions on the basis of received signals, imitating the sender of the signal. After then, consequences of these changes in receiver's actions are displayed to the population, so that animals in the population learn to repeat the process in case the same set of events occurs [3].

2.2. Development of Language in Young Children

In young children's speech, repetition mechanism similar to rote imitation is seen. However, different from animals, imitation of children displays a presence of conversational competence [19]. Furthermore, even though apes could learn to combine two or more symbols in nonrandom ways [8], this type of syntax cannot be compared with highly sophisticated grammar of human language.

In Reference [20], the language development of young children is studied as a research project. In this project, the conversational competence of children is recorded, both semantic and grammatical aspects of their language development is studied, hopefully to understand early stages of grammatical constructions and the meanings that they convey. As a result, it is found that in very early stages children could begin to combine words to make sentences with semantic relations such as nomination, recurrence, disappearance, attribution, possession, agency, and a few others. These alone prove that human's ability to communicate is far more superior to animals.

Children are born with organs that allow them to communicate their basic needs, and they are able to further communicate with others through vocalization, eye gaze and gesture, even in infancy. Although initially they are unaware of the underlying meaning of these behaviors, the responses of caregivers during early years exchanges highlight and teach the communicative nature of language [21].

Through interactions with parents, friends, and caregivers, children learn how to use language appropriately in society. Because language is used for many purposes, lots of skills are involved in conversational competence. People need to learn to ask questions, make requests, give orders, express agreement or disagreement, apologize, refuse, joke, praise, and tell stories.

Additionally, language learning occurs under *maturational constraint*, that is language acquisition is successfully happens only during a maturationally bounded period [22]. Given similar input, after this period learners do not achieve the same outcome. This finding suggests that, people who are going to teach conversational skills to children is probably close to the parents, who presumably stay with children throughout this period.

2.3. Diversity of Culture

The mathematical premises of assessing the diversity in language are given in Reference [16]. An objective, quantitative measure to the diversity is developed, along with its relation to cultural elements such as political, economic, geographic and historic factors. In this study we use similar quantitative measures to calculate linguistic diversity between agents and communities.

The general conclusion in Reference [16] is that communication evolves rather than language itself. To elaborate, language is composed of certain signaling structures, which are used for communication. The signals do not change throughout the evolutionary progress, whereas how these signals are used changes in order to communicate more efficiently. Of course, language plays an important role in conveying certain meanings. However it seems that human languages are generally on the same level regarding efficiency. Reference [16] says that, traditional theories that take into account correlation between complexity of language and evolution of diversity fails to perceive the big picture. Understanding morphemes and how certain structures in language came to be is not sufficient to understand the diversity in languages. It seems that there is a correlation between communication and evolution of different cultures.

Our study is also related to Reference [14]. In Reference [14] diversity of culture is investigated in detail. In describing the evolution by natural selection, it is stated that there is a correlation in phenotype between parents and offspring. On the other hand, there is no reason why the correlation between parents and offspring should be explicitly genetic. Any phenomenon that causes children to be phenotypically nonrandom with respect to their parents will do. Thus children may simply imitate their parents by learning, or they may watch peers of their parents: that is, inheritance may be non-biological [14]. Reference [14] explores the consequences of supposing characteristics to be transmitted by non-genetic routes. In this book there is a discussion about what constitutes culture. Yet it is more interested in the question of how the distribution of these cultural traits would be.

The method is simple and direct. All possible mating are tabulated phenotypically, and the distribution of phenotypes of the offspring is specified. These may then be modified by encounters with possible teachers according to some frequency rule, and then weighted by Darwinian fitnesses. Using the approach described above, how much diverse languages in a population can be found analytically. In this study, we want to see how the diversity in languages emerges. Therefore we use these fundamentals to model a selection model, where agents are expected to choose teachers with respect to their closeness to parent figure.

2.4. Alternative Approaches

There are many researches that use the fundamental mathematical models established in Reference [10]. Our work is related to Reference [23–26] that investigates the effects of various structural organizations in population. For example in Reference [24], the population is divided into multiple groups, and the change in languages, as a result of interactions between these groups, is investigated. In Reference [25] and Reference [26], on the other hand, population is modeled in such a way that for each agent in the population there is a kinship and interaction structures. Kinship and interaction structures are used to investigate the effect of family and friends on the emergence of language respectively.

There are other studies that takes different approaches into consideration regarding the emergence of language. For example in Reference [27, 28], it is investigated through philosophical perspective, particularly the emergence of meaning is discussed.

As we mentioned in Section 1.1, there are three different transmission types (learning) between individuals. In this study, we investigate the model where vertical and oblique transmissions are used. In Reference [29] for example, emergence of language is investigated through a model that learning happens in one particular generation, which is a model of horizontal transmission.

If we consider horizontal transmission, the applicability of these structures expands. Learning in social networks has become a very popular subject lately with the spread of social media. In Reference [30], sequential and interaction based model is used to understand the behavior of financial markets. Similarly, there are numerous economics papers, such as Reference [31–35], that investigate the learning in social networks, that is considered as a complex system that consists of a large number of interacting units, which forms the economy.

3. BACKGROUND

In this study we examine the question of how language may emerge as a result of interactions between generations as the social structure evolves. We expect to find the underlying conditions of the evolutionary process of language in non-linguistic populations. Before we start, we need to introduce the language model.

3.1. Proto-language Model

We introduce a simple language communication model, called proto-language. *Proto-language* is a sign system where there are only simple associations between meanings and signals in population. Let $\mathcal{P}$ be the set of N agents. We have M meanings and S signals. The meaning space $\mathcal{M}$ is the set of all meanings, which in general may be an object or a status, which agents are required to describe. The signal space $\mathcal{S}$ is the set of all possible outputs that the agents are capable of producing to describe these meanings.

An agent $i \in \mathcal{P}$ selects certain meaning $\mu \in \mathcal{M}$ and wants to pass it to agent $j \in \mathcal{P}$. We assume that she does not have means to pass a meaning in her mind directly to the mind of j, therefore she has to use signals. She selects a signal $x \in \mathcal{S}$, which she thinks as a representation of μ, and passes the signal to j. In this context, the agent i who wants to pass a meaning is named *sender agent*, whereas the agent j who receives the signal is named *receiver agent*. We assume that there is no noisy channel between the sender and the receiver. So the receiver picks up exactly what is sent by the sender. Receiving x, j tries to interpret x in his own way. Hopefully j will interpret it as μ.

Clearly, mappings from μ to x and from x back to μ are very important for a successful communication. We need to specify how association of meaning and signal in sending and receiving ends are done. Suppose every agent has a statistic $a_{\mu x}$ of how frequently she uses signal x to mean meaning μ. Then we have an $M \times S$ *association matrix* $\mathbf{A} = [a_{\mu x}]$ from which we can derive encoding and decoding methods. The

following matrix is an example of such association matrix.

$$A_{M\times S} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1x} & \dots & a_{1S} \\ a_{21} & a_{22} & \dots & a_{2x} & \dots & a_{2S} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{\mu 1} & a_{\mu 2} & \dots & a_{\mu x} & \dots & a_{\mu S} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{M1} & a_{M2} & \dots & a_{Mx} & \dots & a_{MS} \end{bmatrix}$$

Encoding matrix, $\mathbf{E} = [e_{\mu x}]$, is an $M \times S$ matrix where $e_{\mu x}$ is the probability of using signal x for meaning μ. *Decoding matrix*, $\mathbf{D} = [d_{x\mu}]$, on the other hand, is an $S \times M$ matrix where $d_{x\mu}$ is the probability of understanding meaning μ for given signal x. Note that encoding and decoding matrices are on reciprocal sides of communication, which means encoding matrix **E** is used to select signal to describe given meaning, while decoding matrix, **D** is used to find out underlying meaning of given signal. The chance that signal x will be chosen to describe meaning μ is $e_{\mu x}$. Similarly the chance that μ will be understood from a given signal x is $d_{x\mu}$. Examples of such encoding and decoding matrices are given below.

$$\mathbf{E}_{M\times S} = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1x} & \dots & e_{1S} \\ e_{21} & e_{22} & \dots & e_{2x} & \dots & e_{2S} \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{\mu 1} & e_{\mu 2} & \dots & e_{\mu x} & \dots & e_{\mu S} \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{M1} & e_{M2} & \dots & e_{Mx} & \dots & e_{MS} \end{bmatrix}$$

$$\mathbf{D}_{S\times M} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1\mu} & \dots & d_{1M} \\ d_{21} & d_{22} & \dots & d_{2\mu} & \dots & d_{2M} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{x1} & d_{x2} & \dots & d_{x\mu} & \dots & d_{xM} \\ \vdots & \vdots & & \vdots & & \vdots \\ d_{S1} & d_{S2} & \dots & d_{S\mu} & \dots & d_{SM} \end{bmatrix}$$

These matrices are obtained from given association matrix as follows:

$$e_{\mu x} = \frac{a_{\mu x}}{\sum_{x'=1}^{S} a_{\mu x'}} \quad \text{and} \quad d_{x\mu} = \frac{a_{\mu x}}{\sum_{\mu'=1}^{M} a_{\mu' x}}$$

We will focus on $\mathbf{A}$ for language learning since $\mathbf{E}$ and $\mathbf{D}$ can be derived from $\mathbf{A}$. Note that, for each agent i, there is a dedicated association matrix $\mathbf{A}^{(i)}$, encoding matrix $\mathbf{E}^{(i)}$ and decoding matrix $\mathbf{D}^{(i)}$. Also note that, in both $\mathbf{E}$ and $\mathbf{D}$ matrix, each rows sum to 1:

$$\sum_{x'=1}^{S} e_{\mu x'} = 1 \quad \text{for all } \mu \in \{1, \cdots, M\}$$

and

$$\sum_{\mu'=1}^{M} d_{x\mu'} = 1 \quad \text{for all } x \in \{1, \cdots, S\}$$

3.2. Comprehension

Suppose agent i wants to pass meaning μ to agent j. In Figure 3.1, a diagram of such transmission is shown. Basically, first i encodes meaning μ using her encoding matrix. The output of this encoding process is a signal, let's say it is x, and is sent to

j. Now, j needs to decode this signal. To do so, she uses her own decoding matrix, and the output of the decoding process is a meaning. If j decodes μ from x, it means that comprehension of μ from i to j is successful.

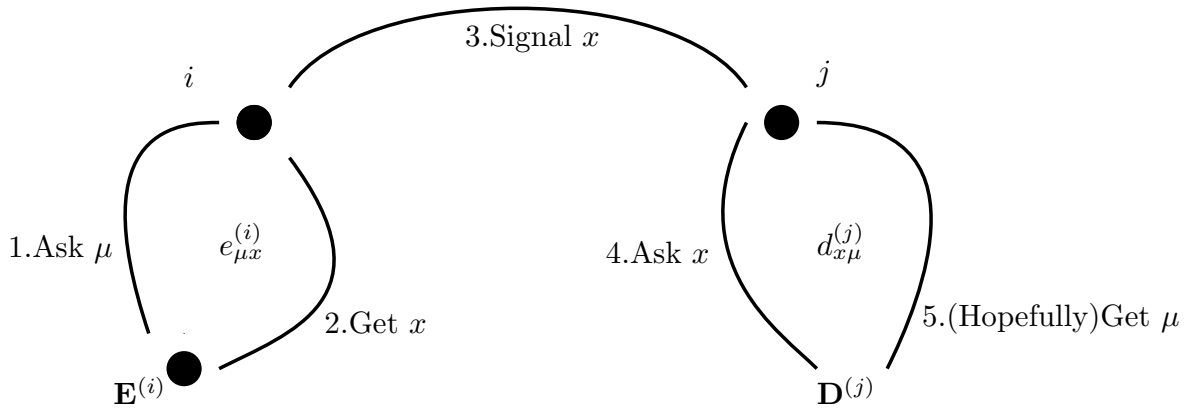


Figure 3.1. Simple transmission from agent i to agent j.

If we sum for all signals that can be used in between, probability of successfully communicating μ is

$$\sum_{x\in\mathcal{S}} e^{(i)}_{\mu x} d^{(j)}_{x\mu}$$

where $e^{(i)}_{\mu x}$ and $d^{(j)}_{x\mu}$ are from encoding matrix of i and decoding matrix of j, respectively. When we average that over all meanings, we obtain *comprehension* $F(i \to j)$ from i to j, that is

$$F(i \to j) = \frac{1}{M} \sum_{\mu\in\mathcal{M}} \sum_{x\in\mathcal{S}} e^{(i)}_{\mu x} d^{(j)}_{x\mu}$$

If we want them to communicate both ways, we consider *mutual comprehension*

$$F(i \leftrightarrow j) = \frac{F(i \to j) \ + \ F(j \to i)}{2}$$

We can define comprehension from one community to another community in a similar fashion. Let $\mathcal{B}$ and $\mathcal{C}$ be subsets of $\mathcal{P}$ and $\mathcal{B} \cap \mathcal{C} = \emptyset$. Then *comprehension* from $\mathcal{B}$ to $\mathcal{C}$ is defined as

$$I(\mathcal{B} \to \mathcal{C}) = \frac{1}{|\mathcal{B}||\mathcal{C}|} \sum_{i \in \mathcal{B}} \sum_{j \in \mathcal{C}} F(i \to j)$$

and *inter community comprehension* as

$$I(\mathcal{B} \leftrightarrow \mathcal{C}) = \frac{I(\mathcal{B} \to \mathcal{C}) \ + \ I(\mathcal{C} \to \mathcal{B})}{2}$$

We can also define comprehension within a community $\mathcal{C}$. *Within community comprehension* $W(\mathcal{C})$ is defined as the average comprehension in a community $\mathcal{C}$. Thus,

$$W(\mathcal{C}) = \frac{1}{|\mathcal{C}|(|\mathcal{C}| - 1)} \sum_{i \in \mathcal{C}} \sum_{\substack{j \in \mathcal{C} \\ j \neq i}} F(i \leftrightarrow j)$$

Note that, $W(\mathcal{C})$ is comprehension value between different members of the community, therefore we did not include $F(i \leftrightarrow i)$ to the calculation. Within community comprehension of the entire population, i.e., $W(\mathcal{P})$, is called *overall comprehension*.

We can also calculate language-wise closeness of agent i to community $\mathcal{C}$. *Closeness to community* $W(i \to \mathcal{C})$ is defined as the average comprehension between the agent i and the members of community $\mathcal{C}$. Thus,

$$W(i \to \mathcal{C}) = \frac{1}{|\mathcal{C} \setminus \{i\}|} \sum_{j \in \mathcal{C} \setminus \{i\}} F(i \leftrightarrow j)$$

Note that, in case $i \in \mathcal{C}$, we calculate the closeness between agent i and the rest of the community $\mathcal{C}$.

Let's consider the best mutual comprehension that can be possible between two agents who are speaking the same language. $F(i \leftrightarrow j) = 1$ case is only possible when each meaning μ can be passed successfully. *Ambiguity* is a phenomenon where some signal represents more than one meaning. If certain language has ambiguity, communication may not be successful. To eliminate ambiguity, first of all decoding matrix $\mathbf{D}$ must be binary matrix, thus for each signal x, there is at most 1 meaning to be understood.

Although there are some cases such as $M > S$ that ambiguity is inevitable. In this case, there is not enough signals to represent each meaning one by one. To maximize $F(i \leftrightarrow j)$, decoding matrix still should be binary matrix which $d_{x\mu} = 1$ if $e_{\mu x}$ is the largest entry in a column of $\mathbf{E}$; all other entries of $\mathbf{D}$ are 0. On the other hand, the maximum comprehension for encoding matrix $\mathbf{E}$ is obtained when $\mathbf{E}$ has exactly one 1 in every row ($M \leq S$) or in every column ($M > S$) [10].

As ambiguity increases, communication becomes less successful until the point where encoding and decoding processes are fully random. *Random communication* happens when all meaning signal associations are equally likely, that is, $e_{\mu x} = 1/S$ and $d_{x\mu} = 1/M$ for all possible (μ, x) pairs. In this case, the mutual comprehension between two agents i and j becomes

$$F(i \leftrightarrow j) = \frac{1}{M} \tag{3.1}$$

At this point, consider random comprehension within a community $\mathcal{C}$. *Random community comprehension* $W_{\mathrm{r}}(\mathcal{C})$ is defined as the average comprehension in a community where all communication is random. We find,

$$W_{\mathrm{r}}(\mathcal{C}) = \frac{1}{M} \tag{3.2}$$

from Equation 3.1.

3.3. Moran Model

The Moran process is used to study selection in finite populations. Let's say there is a population of fixed size N, in Moran process, population size always remains constant N in time. In each generation, one agent is randomly selected for reproduction, whereas another one is selected to be replaced by the child of the first agent. Note that, these two agents may not be different; that is, the first agent may be replaced by its own child [36]. At the end of this process, since there is only one birth and one death event, population size remains N.

In the context of evolution of language, this process is used to model transfer of language from one generation to another. In each generation, some agents are selected as teachers of the next generation. At each time step, one teacher and one parent is selected. Parent agent is replaced by its offspring where the offspring inherits the language of the teacher. At certain time steps later, all agents are replaced by new set of agents in the next generation keeping the total number of agents in the population constant.

3.4. Learning Model

The evolution of language can happen in two different ways. Language evolves either through agents interacting with each other within a generation, or it is transferred from one generation to the next by means of learning. In Reference [10], fundamentals of the latter form of interaction have been established.

First of all, the language of each agents in the first generation are initially random. After then, at each generation, population is replaced with new set of N agents. Agents of new generation have no meaning signal associations initially. That is, the association matrices of agents are empty. For language to be transferred from the generation of parents to the generation of children, some agents are assumed to be chosen as *teachers*.

In Reference [10], teacher selection is a result of fitness gains. The *fitness* of agent i is given as

$$F_i = \sum_{j \in \mathcal{P}} F(i \leftrightarrow j)$$

For the next generation, offspring are produced proportional to the fitness of an agent: the chance that a particular agent arises from agent i is proportional to

$$\frac{F_i}{\sum_{j \in \mathcal{P}} F_j}$$

That is, each child agent selects her teacher according to this probability distribution. Thus, agents who have better fitness are picked more. In Reference [10], it is stated that more than one teacher could be assigned for each child agent. This case is examined as a form of cultural learning, where some elite group of agents is responsible for transition of language. It is reported that since the selection mechanism remains the same, total number of teachers assigned only effects how fast the language emerges in such populations [10].

After teachers of the next generation are assigned, language is transferred from teacher to child. The learning process is no different than a naming game [37]. Child learns the language of her teacher by sampling their responses to specific meanings. The response is simply an encoding process where a teacher agent chooses a signal to call given meaning. For each meaning, the teacher provides Q responses and the child uses these to populate her association matrix, where Q is called *sampling size*.

After then child agent updates its association matrix $\mathbf{A}$ with this meaning signal pairs. Update process is as follows: Let i and t be the child and teacher agents respectively. For certain meaning $\mu \in \mathcal{M}$, let $x \in \mathcal{S}$ be the response signal of t. In the very basic case, child agent uses these parameters to update the entry $a_{\mu x}$ of its association matrix incrementing it by 1.

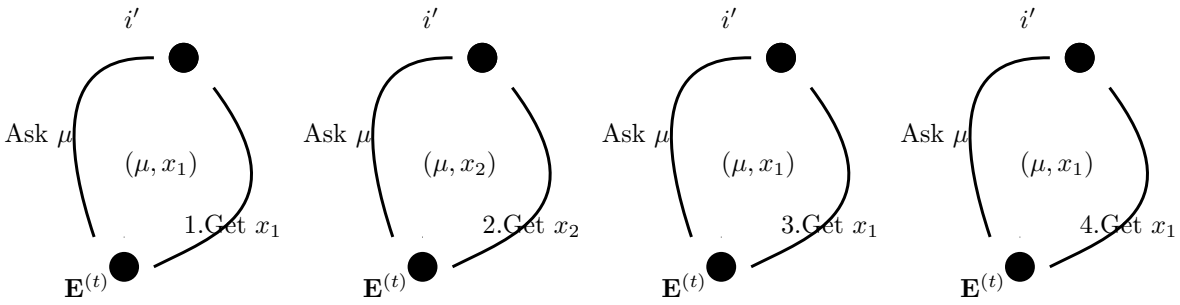


Figure 3.2. Teaching process of single meaning μ ($Q = 4$).

Example of a learning process of single meaning μ for $Q = 4$ can be found in Figure 3.2. In this example, μ is signaled with x_1 three times whereas it is only signaled with x_2 one time. This means that the probability of describing μ as x_1 and x_2 is 3/4 and 1/4 respectively. This corresponds to $a_{\mu x_1} = 3$ and $a_{\mu x_2} = 1$ in association matrix. When they are done editing association matrices, encoding and decoding matrices $\mathbf{E}$ and $\mathbf{D}$ are generated.

3.5. k-means Nearest Neighbors Algorithm

In this section, we will explain a method to detect sub-language groups. In order to do that we adapt k-means clustering algorithm to the context of language.

For a given cluster count K and a distance metric defined on set of observations, k-means clustering algorithm tries to find a partition of observations with K clusters in such a way that average within cluster distance is optimized [38]. It is a heuristic algorithm. One can find the best value of parameter K by trial and error. Pseudo-code of such algorithm can be found in Figure 3.3.

In this algorithm, observations are assigned to random clusters initially. At each iteration, each observation is assigned to the cluster that is the closest. To do that, first all cluster centroids are calculated. Then, for each observation, distances of this observation to each cluster centroids are calculated. After then each observation is assigned to a cluster that the distance between its centroid and the observation is

minimal. If the observation is already assigned to the selected cluster, no change is made. Some iterations later we expect the system to get stuck at some local maxima, that is observations are no longer assigned to different clusters.

To detect sub-languages, we use k-means algorithm. In this algorithm, k-means provides K communities in such a way that agents in the same community understand each other better. So the distance metric is mutual comprehension, and objective is maximization of within community comprehension in communities. Our approach has two steps: first we find the best partition for a given K, then we find the best K for our purpose.

Let $\mathbb{P}_K = \{\mathcal{C}_1, \mathcal{C}_2 \cdots, \mathcal{C}_K\}$ be a partition of set of agents $\mathcal{P}$ with K clusters. We consider clusters as *language communities*. The *average within community comprehension* is defined as

$$W(\mathbb{P}_K) = \frac{1}{K} \sum_{i=1}^{K} W(\mathcal{C}_i)$$

To find such partition, we use the modified version of k-means algorithm where the objective is to find the best partition for given K. The pseudo-code of this algorithm can be found in Figure 3.4. First of all we have agents instead of observations. In this algorithm, our goal is the maximization of the comprehension between agents and clusters. instead of distance minimization. Therefore at each step, different from standard k-means algorithm, agents are assigned to the clusters that they are closest to in terms of mutual comprehension.

In this version of k-means algorithm, we have chosen to use only average within community comprehension as objective function. To test if resulted communities are different from each other, we may also need average inter community comprehension values. The *average inter community comprehension* is defined as

$$I(\mathbb{P}_K) = \frac{1}{K(K-1)} \sum_{\mathcal{B} \in \mathbb{P}_K} \sum_{\substack{\mathcal{C} \in \mathbb{P}_K \\ \mathcal{C} \neq \mathcal{B}}} I(\mathcal{B} \leftrightarrow \mathcal{C})$$

Input:

$\mathcal{O} = \{o_1, o_2, \cdots, o_N\}$ (set of observations to be clustered)

K (number of clusters)

$MaxIters$ (limit of iterations)

Output:

$\mathbb{P} = \{C_1, C_2, \cdots, C_K\}$ (set of clusters)

$\mathcal{L} = \{l(o_i) | i = 1, 2, \cdots, N\}$ (set of distances between o_i and assigned cluster for each $o_i \in \mathcal{O}$)

```
1  begin
2      foreach C_i ∈ ℙ do
3          C_i ← o_j ∈ 𝒪 (e.g. random selection)
4      end
5      changed ← false,    iter ← 0;
6      repeat
7          foreach o_i ∈ 𝒪 do
8              minDist ← argMinDistance(o_i, ℙ); (find the distance between
                o_i and a centroid that the distance is minimal)
9              if minDist ≠ l(o_i) then
10                 l(o_i) ← minDist;,    changed ← true;
11             end
12         end
13         foreach C_i ∈ ℙ do
14             UpdateCluster(C_i); (assign each observation to a cluster that it
                is closest to)
15         end
16         iter = iter + 1;
17     until changed = true and iter ≤ MaxIters;
18 end
```

Figure 3.3. Standard k-means Algorithm.

Input:

$\mathcal{P} = \{p_1, p_2, \cdots, p_N\}$ (set of agents to be clustered)

K (number of clusters)

$MaxIters$ (limit of iterations)

Output:

$\mathbb{P} = \{C_1, C_2, \cdots, C_K\}$ (set of clusters)

$\mathcal{L} = \{l(p_i) | i = 1, 2, \cdots, N\}$ (set of closeness between p_i and assigned cluster for each $p_i \in \mathcal{P}$)

```
1  begin
2      foreach C_i ∈ ℙ do
3          C_i ← p_j ∈ 𝒫 (e.g. random selection)
4      end
5      changed ← false,    iter ← 0;
6      repeat
7          foreach p_i ∈ 𝒫 do
8              maxSim ← argMaxComprehension(p_i, ℙ); (find the closeness
                W(p_i → 𝒞_j) between p_i and 𝒞_j that the comprehension is
                maximal)
9              if maxSim ≥ l(p_i) then
10                 l(p_i) ← maxSim,    changed ← true;
11             end
12         end
13         foreach C_i ∈ ℙ do
14             UpdateCluster(C_i); (assign each agent to a cluster that it is
                closest to)
15         end
16         iter = iter + 1;
17     until changed = true and iter ≤ MaxIters;
18 end
```

Figure 3.4. Modified k-means algorithm.

There are many partitions of $\mathcal{P}$ with K communities. We select the partition $\overline{\mathbb{P}_K}$ with the maximum average within community comprehension for given K. That is,

$$\overline{\mathbb{P}_K} = \arg\max_{\mathbb{P}_K} \quad W(\mathbb{P}_K)$$

After then, we expect to find a suitable clustering for given system. Unfortunately, there is no algorithm to find the optimal community count. Therefore, we run the algorithm for $K \in \{K_{\min}, \ldots, K_{\max}\}$ and select the one with highest comprehension. Thus,

$$K^* = \arg\max_{K} \quad W(\overline{\mathbb{P}_K})$$

is the *optimum community count*. The corresponding partition $\overline{\mathbb{P}_{K^*}}$ is the *optimum partition* with the *optimum within community comprehension* value and *optimum inter community comprehension* value of

$$W^* = W(\overline{\mathbb{P}_{K^*}}) \quad I^* = I(\overline{\mathbb{P}_{K^*}})$$

respectively.

4. MODEL

We propose an evolutionary model where every generation has N agents. Every agent i makes exactly one child i'. Each child learns her language from a single agent called *teacher*, donated by t. The teacher provides Q samples for each meaning and the child fills her association matrix based on these samples. After learning process is completed, i is replaced by i'.

In this study, various teacher selection methods and their effects to emergence of diversity in language are investigated. First of all, parent may not be the teacher but she effects the selection of it. Selection of teacher is done in two steps. In the first step, each parent is assigned to R agents. The teacher t of child i' is selected from the candidates in $\mathcal{V}_i$, which is called the *imitation set* of i.

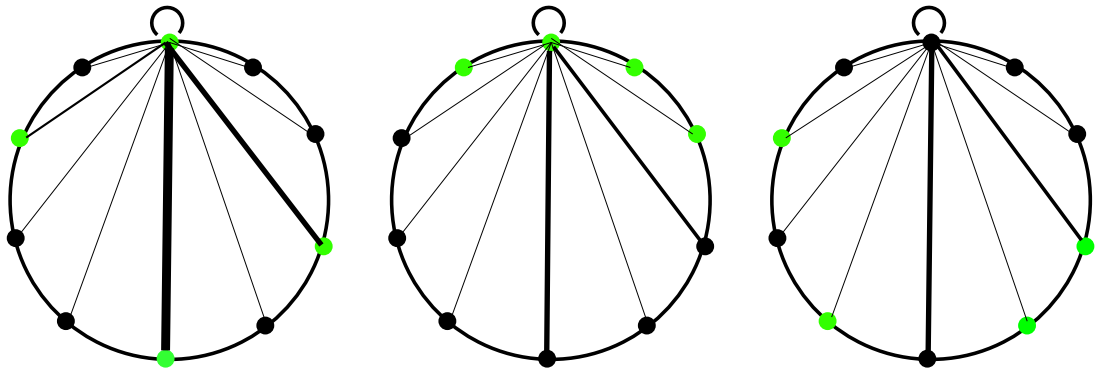
We consider three different ways to select R agents of imitation set.

(i) *Model-A.* Here, we are trying to construct a social structure that is similar to lifetime encounters. The most basic assumption is that agents make friends with whom they comprehend better. Therefore we select R agents that are closest to the parent language-wise.
Specifically, imitation set $\mathcal{V}_i$ of particular agent i is selected as follows. At each generation, agent i calculates $F(i \leftrightarrow j)$ for all $j \in \mathcal{P}$, picks R agents who have the highest comprehension with i. Thus, first R agents in population that the parent comprehends best forms the imitation set of the offspring.

(ii) *Model-B.* In accordance with the previous model, here we will also report the results of the case where the population is spatially organized. In particular, we assume that N agents are placed in equidistant sites on a 1D ring lattice. Then we select R agents that are physically closest to the parent. Note that child replaces parent in the lattice.

(iii) *Model-C.* We also examined the case where free associations are taken into account. This is a form of selection where agents' behavior is not limited with

structural restrictions. In this case, each agent samples random R members at each generation.



(a) Model-A. (b) Model-B. (c) Model-C.

Figure 4.1. Example of teacher selections for $N = 10$ and $R = 4$.

In Figure 4.1, we shared the selection results for a simple example where $N = 10$ and $R = 4$. Figure 4.1a, Figure 4.1b, Figure 4.1c shows the selection results for Model-A Model-B and Model-C respectively. In this example, each vertex represents a certain agent, and each edge represents a mutual comprehension between agents at two ends. The thickness of each edge is given in correlation with the mutual comprehension it represents, and the order from highest to lowest is as follows: $F(1 \leftrightarrow 6) > F(1 \leftrightarrow 1) = F(1 \leftrightarrow 3) = F(1 \leftrightarrow 8) > F(1 \leftrightarrow 2) = F(1 \leftrightarrow 4) = F(1 \leftrightarrow 5) = F(1 \leftrightarrow 7) = F(1 \leftrightarrow 9) = F(1 \leftrightarrow 10)$. The objective is to select the agents of $\mathcal{V}_1$, therefore we only showed the mutual comprehension between agent 1 and others. Note that agent 1 may not be in the imitation set but certainly effects the selection.

In Model-A agent 1 is supposed to select agents that are closest to her language-wise, and we see in Figure 4.1a that agents 6, 1, 3 and 8 have higher mutual comprehension values. Thus, $\mathcal{V}_1 = \{1, 3, 6, 8\}$ for Model-A. In Model-B we simply need to pick physically closest agents. In Figure 4.1b we see that closest agents are agents 2, 10, 9 and herself. Thus, $\mathcal{V}_1 = \{1, 2, 9, 10\}$ for Model-B. In Model-C on the other hand, there is no restriction therefore the agents in $\mathcal{V}_1 = \{3, 5, 7, 8\}$ are selected randomly, as it can be seen in Figure 4.1c.

In the second step, an agent, who has better mutual comprehension with the parent, has better chances to teach her language to the child. That is, teacher t is selected within the imitation set $\mathcal{V}_i$ proportional to

$$\pi(i,t) = \frac{F(i \leftrightarrow t)}{\sum_{j \in \mathcal{V}_i} F(i \leftrightarrow j)}$$

Note that, for each child i' exactly one teacher is chosen from the imitation set of i. Note also that an agent can be chosen as teacher by several children or none of them.

Now let's go back to example in Figure 4.1c. In this example, we know that $\mathcal{V}_1 = \{3, 5, 7, 8\}$. When we look at the mutual comprehension values, we see that $F(1 \leftrightarrow 3) = F(1 \leftrightarrow 8) > F(1 \leftrightarrow 5) = F(1 \leftrightarrow 7)$. Thus, at second step, agents 3 and 8 has better chances than agents 5 and 7, to be selected as teacher by agent 1.

4.1. Selection of Simulation Parameters

We investigate the effect of imitation set size R. There are $N = 100$ agents using $S = 8$ signals to communicate $M = 8$ meanings. Each data point is an average result of 100 realizations. We used the parameter $r = R/N$ instead of R in domain, thus we are able to compare the results of the simulations with other population sizes other then 100, too. The simulation results for population sizes $N = 50, 100, 150, 200$ can be found in Appendix B.

We run each realization for 500 generations and the language of each agents in the first generation are initially random as it is described in Section 3.2. 500 generations is sufficient since as it can be seen in Appendix A that most of the simulations rapidly converged even in 100 generations to a state where there is no longer a change in $W(\mathcal{P})$, which indicates that the simulation has reached to steady state. There is an exception with Model-B which took 400 generations to stabilize. We also run the simulations for $1,000$ generations to check if there is a noticeable change in the simulations, which we could find none.

In Reference [10], emergence of language is investigated for sampling sizes $Q = 1, 4, 7, 10$. We set the sampling size $Q = 4$ in our simulations. As we mentioned earlier in the Section 3.4, sampling size is the parameter that is used in learning process. For very small Q, only a summary of teacher's language could be passed on. As Q gets higher, the passed language becomes more similar to what teacher had, in terms of encoding and decoding matrices. The simulation results for different sampling sizes can be found in Appendix C. In summary, we found that the quality of language is low for $Q = 1$, whereas it gets higher and stays approximately the same for $Q = 4, 7$ and 10. Therefore $Q = 4$ seems to be the reasonable selection in our simulation.

Regarding selection of meaning and signal counts in the simulations, we repeated the same principles that can be found in Reference [10]. To understand the components of language evolution, we are using rather simple language model that has $M = 8$ meanings and $S = 8$ signals. In the best case scenario, different signals would be assigned for each meaning. However, because some signals might get lost during learning process, we expect some ambiguity.

We tested for different meaning and signal counts and reported the results in Appendix D. There are three cases: (i) $M > S$; if meaning count exceeds signal counts, there is bound to be ambiguity, since there is not enough signals to represent each meaning uniquely, thus comprehension values are lower. (ii) $M < S$; when there are more signals than meanings, we observed better comprehension values; since even if some signals get lost, there is still sufficient amount of signals left to represent each meanings. (iii) $M = S$; in this case, we detected relatively lower comprehension values than we had in the case where $M < S$. The reason is that it is possible some signals to get lost during learning. Thus, it causes ambiguity to happen inevitably as in the case where $M > S$.

5. RESULTS AND DISCUSSION

5.1. Detection of Global Language

In Figure 5.1, we reported the overall comprehension $W(\mathcal{P})$ for different r values. We experimented the effects of different selection strategies to global language. Base Model is represented as straight line since imitation sets are not defined in this model. As we can see, Model-C resulted with the best overall comprehension, compared to Models A and B. They fail to develop a global language that provides successful communication in between all members of population unless r value is higher than 0.5, which is the case where at least half of the population is in candidates. This raises the question, why selection strategies that take into account either language-wise or spatial closeness fails to provide a medium for emergence of global language. One possible explanation could be that rather than single language, that is used by the entire population, many languages, that are used by small communities, emerged.

5.2. Detection of Sub-Languages

Testing the hypothesis above, we used k-means algorithm to see if such different communities emerge. As we mentioned before in Section 3.5, to find the optimal community count K^*, we need to apply k-means algorithm to the population with different cluster counts, and compare the results obtained by objective functions. Specifically, we applied k-means for $K \in \{1, 2, \ldots, 10\}$, and reported K^* and corresponding W^* values in Figure 5.2 and Figure 5.3 respectively.

In these results, K^* alone does not tell us much. We have to check if the corresponding comprehension W^* is high enough. In Figure 5.3, low W^* value indicates that we could not find any suitable communities, whereas when W^* is high, there is such a partition that agents of the same community comprehend each other quite well.

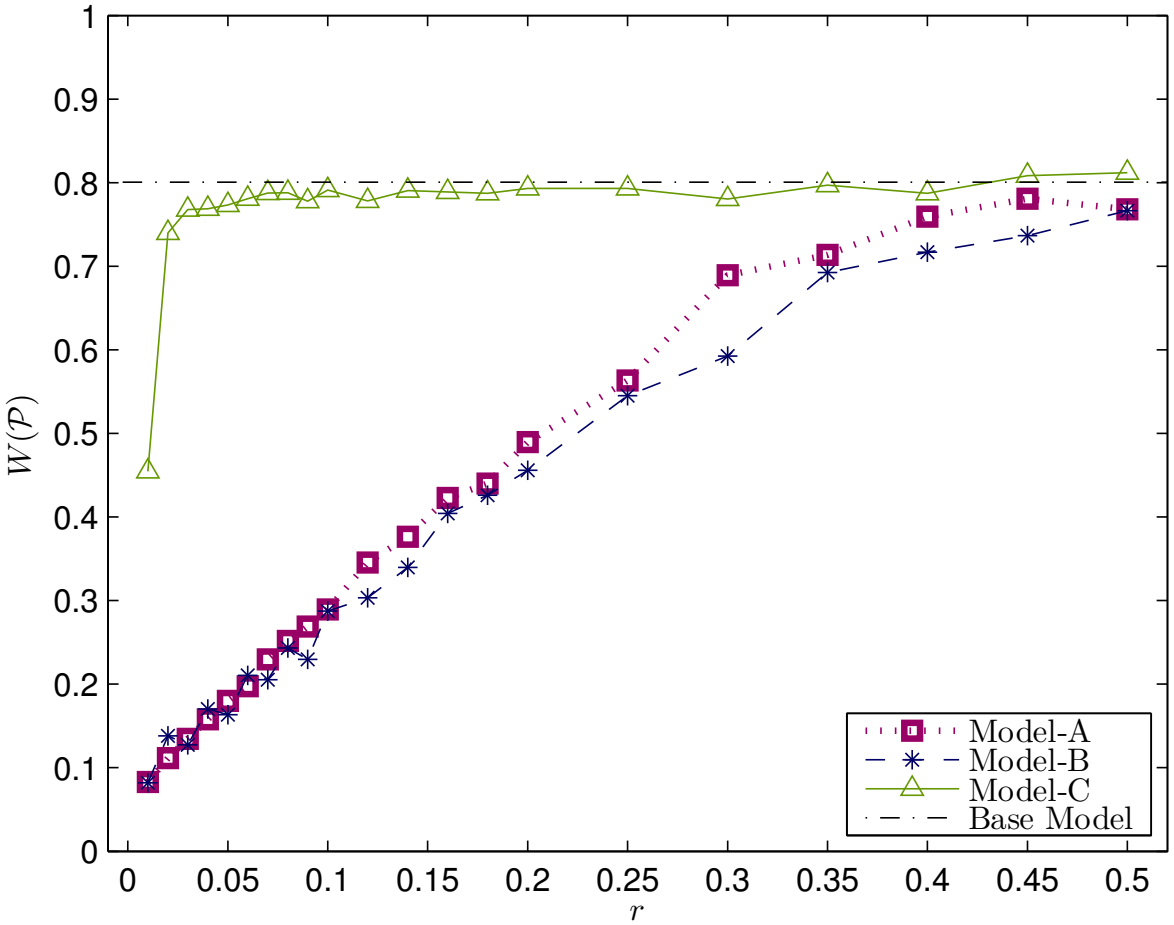


Figure 5.1. Overall comprehension of selection strategies for simulations of $(N, M, S, Q) = (100, 8, 8, 4)$.

5.3. Discussion

We assume that, as long as language conveys some information about true signals, which means that mutual comprehension is not random, it can be considered suitable. Thus,

$$W^* \geq W_{\mathrm{r}}(\mathcal{P}) = \frac{1}{M}$$

must hold in population. In our experiments, we tested for $M = 8$ meanings. With Equation 3.2 we can calculate $W_{\mathrm{r}}(\mathcal{P}) = 0.125$. As a result, as we can see in Figure 5.3, global or sub-languages exist in Model-A and Model-B for $R > 1$.

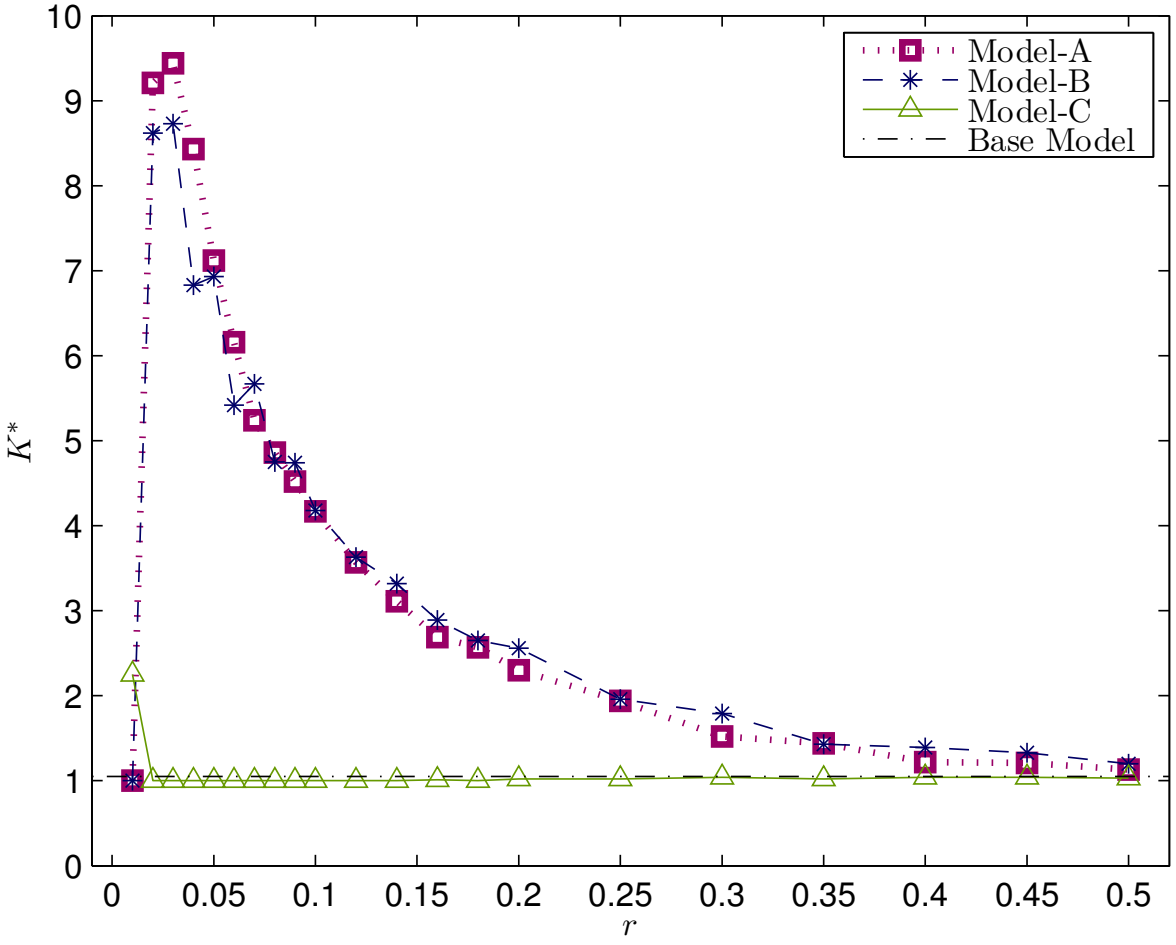


Figure 5.2. Community counts for simulations of $N = 100$, $M = 8$, S $= 8$, $Q = 4$.

On the other hand, our models bring along some restrictions, therefore it is expected to see some decrease in comprehension values. Let's consider the minimum requirements for a comprehension that is as successful as overall comprehension obtained in Base Model. In Figure 5.3, we see that, comprehensions in Models A and B are poor for low r values. As r gets higher, we see that W^* also approaches to more reasonable values. This observation indicates that there is some threshold value for r, let's say r_c, for $r < r_c$, detected sub-languages are worse-off. r_c value can be found as the intersection point between plot line of Base Model and plot curves of Models A and B in Figure 5.3. Hence our first result is that,

$$r_c^A \approx 0.12 \quad \text{and} \quad r_c^B \approx 0.25$$

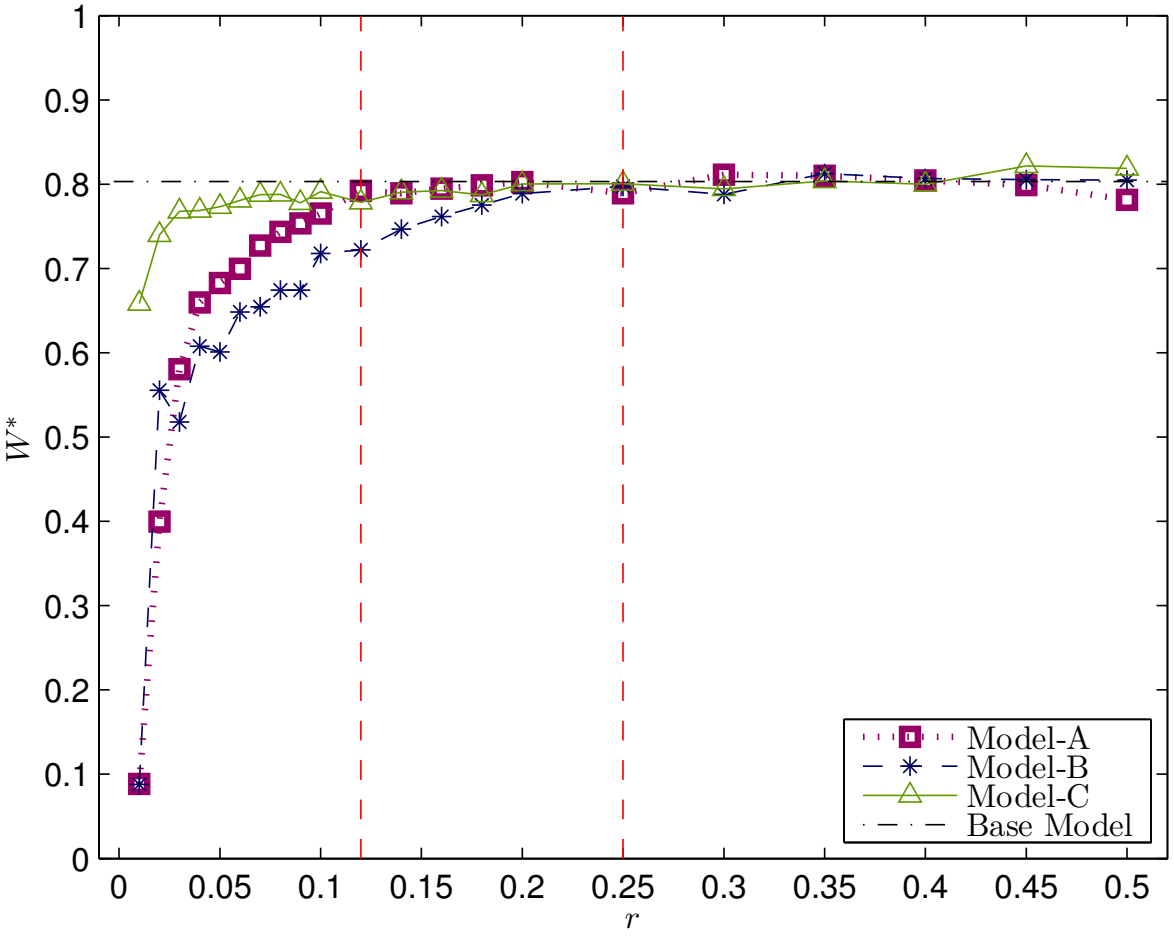


Figure 5.3. Within community comprehensions for simulations of $N = 100$, $M = 8$, $\mathrm{S} = 8$, $Q = 4$.

where r_c^A and r_c^B are threshold values for Models A and B, respectively.

As expected, Base Model and Model-C resulted in one community, indicating emergence of one global language in Figure 5.2. On the other hand, Models A and B showed dramatic increase in within community comprehension values, whereas we found more than one communities of sub-languages.

Although Models A and B are very different from each other in structure, we observe in Figure 5.2 and Figure 5.3 that resulting number of emerged sub-communities and the average comprehension value in these communities are approximately the same for both. This is quite interesting since in Model-A agents are expected to choose the

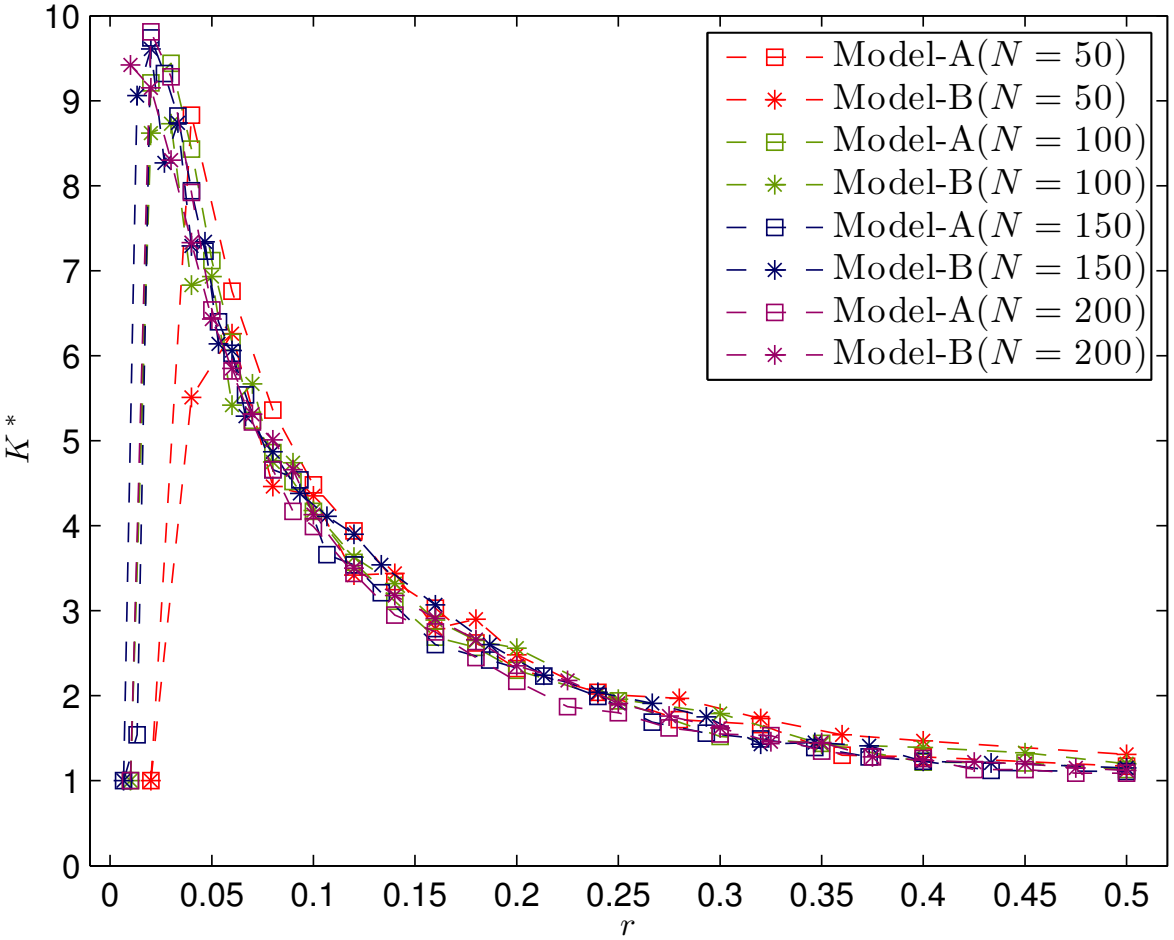


Figure 5.4. Cluster counts of Models A and B for different population sizes.

language-wise closest candidates, whereas in Model-B they just pick the physically closest candidates, which are decided by structural organization of the population. As a result, population ends up with similar communities in terms of comprehension and size in experiments with both models.

Moreover, regarding Models A and B we found that the optimum community count K^* is independent of N, yet dependent to r. In Figure 5.4, we presented the simulation results for different population sizes. As we can see, in Models A and B the resulting optimum community counts remained the same for corresponding r for different population sizes. Thus, relation

$$K^* \propto r^{-\gamma}$$

exists between r and K^*. One direct conclusion can be that the number of sub-language communities in a population can be understood and controlled via the ratio of neighborhood size to population size. Note that, we did not share the results of all experiments for different population sizes $N \in \{50, 100, 150, 200\}$ here to avoid repetition, since the observations stayed the same. Rest of the simulation results can be found in Appendix B.

Finally, we observe best comprehension values in Model-C, along with the emergence of one global language in most of the cases. This result can be explained by the fact that in Model-C agents are essentially freer to select any agent. Thus, this results in a situation where every part of the population have a chance to transmit their languages. As a result, emerged language is a product of everyone, therefore it can be globally communicated with.

6. CONCLUSION

6.1. Future Work

The models we used only cover very basic form of the process and far away from analyzing many complex details of language. Various additions could be made to the model. First of all, we have assumed that each individual inherits their language from one teacher in a very specific way. Different types of learning processes have been reported in Reference [10]. For example, evolution of language can be perceived as a cultural process where some group of people are responsible for the transfer [13]. That is more than one teacher could be assigned to each child.

Another alternative approach is that changing the comprehension types. In this study, we seek mutual comprehension between teacher and parent in selection process. However, we can argue that one way comprehension may result in different evolutionary result. In this sense, we can use different comprehension types such as; (i) teacher is the one who understands the parent best $F(i \to t)$ and (ii) teacher is the one who the parent understands best $F(t \to i)$ instead of mutual comprehension $F(i \leftrightarrow t)$, and compare emerged languages as a result.

Now, let's assess the quality of the community detection algorithm that we used in our experiments. Even though k-means is a widely used heuristics, we may need much more specialized form of community detection algorithms. Clustering algorithms are essentially NP-hard, therefore heuristic algorithms that fits best to our problem can be investigated [39].

One noticeable problem with the usage of k-means algorithm is that it does not guarantee that size of the communities will be the same. Suppose the partition has only 2 clusters with 1 agent and $N-1$ agents. In this case 1 agent cluster dominates $W(\mathbb{P}_K)$. Indeed we encountered some communities that are very small in size in the experiments. There is a need for a kind of normalization with cluster sizes. One simple

example to the modified objective function is below:

$$W(\mathbb{P}_K) = \frac{1}{K} \sum_{i=1}^{K} |\mathcal{C}_i| \; W(\mathcal{C}_i)$$

thus, we give less weight to the small cluster sizes. Other alternative approaches that take into community sizes can be looked into. In our experiments, we used a modified version of k-means algorithm to cluster agents into several groups according to their languages. This algorithm may be revisited in terms of quality and performance. In summary, after implementation of this algorithm, agents in the same group are expected to have closer languages in terms of comprehension. Here, language-wise closeness may be considered in detail.

Let's consider mutual comprehension $F(i \leftrightarrow j)$ between two different agents i and j. Mathematically we know that $\max F(i \leftrightarrow j) = 1$ and $\min F(i \leftrightarrow j) = 0$. However, we did not encounter these two cases in our experiments, since it requires many conditions to exist in the language. Additionally, as we reported before in Equation 3.1, comprehension value to be in between $[0, 1/M]$ is not very different than having no comprehension at all. On the other hand, since some ambiguity is acceptable in language, for instance $F(i \leftrightarrow j) \approx 0.8$ is not so bad. Furthermore, in case where ambiguity is very common in community even $F(i \leftrightarrow j) \approx 0.6$ could be considered good. This knowledge makes it problematic to decide whether two agents should be put in the same community or not.

We did not include average inter community mutual comprehension value I^* in the clustering algorithm. It can be seen in Appendix E that inter community mutual comprehension values are not distinctive with values close to 0.1. That is, communities are already very distinct from each other, therefore only within community comprehension values determine the quality of languages.

In this study, we designed experiments to see their effect on emerged languages. However, in reality each individual has different strategy in teacher selection, thus

some strategies may not survive against others. Therefore, it would be reasonable to approach the problem as game theoretical application. To do this, we can simply design a game where four different strategies play the game with payoff of W^*. Then we can report the surviving strategies.

Also note that, we tried to model the fundamental forms of selection mechanisms. On the other hand, there are many other limitations that can effect selection of teachers, such as labor roles, class, gender and racial differences. Specifically in Model-B we have discussed territorial differences and we use ring lattice as a spatial organization. Any other graph network could be an alternative, and could result in different form of sub-community formations.

6.2. Conclusion

We view this study as a contribution towards understanding the nature of the evolution of language groups. The origin of language is investigated in many studies through Darwinian Frameworks, although the mathematical models are rare. In Reference [10], the mathematical premises for the emergence of language are given. In this study, we extended these frameworks to investigate the emergence of diversity in languages.

Specifically, we investigated teacher selection mechanisms, assuming that teachers are supposed to be close to parents socially or physically. As a result we showed the emergence of sub-languages. To achieve this, we modified and used k-means community detection algorithm. Furthermore, we found a significant correlation between the size of neighborhood of parents and the number of these sub-languages. The nature of this correlation might be investigated in detail with many other possible extensions in the future.

REFERENCES

1. Pinker, S. and P. Bloom, "Natural Language and Natural Selection", *Behavioral and Brain Sciences*, Vol. 13, No. 04, pp. 707–727, 1990.

2. Ewens, W., *Mathematical Population Genetics*, Springer, NewYork, 2 edn., 2004.

3. Smith, W. J., *The Behavior of Communicating*, Harvard University Press, 1980.

4. Szathmáry, E., J. M. Smith *et al.*, "The Major Evolutionary Transitions", *Nature*, Vol. 374, No. 6519, pp. 227–232, 1995.

5. Seyfarth, R. and D. Cheney, "The assessment by vervet monkeys of their own and another species' alarm calls", *Animal Behaviour*, Vol. 40, No. 4, pp. 754–764, 1990.

6. Lieberman, P., *The Biology and Evolution of Language*, Harvard University Press, 1984.

7. Hauser, M. D., *The Evolution of Communication*, MIT Press, 1996.

8. Terrace, H. S., L.-A. Petitto, R. J. Sanders and T. G. Bever, "Can an Ape Create a Sentence", *Science*, Vol. 206, No. 4421, pp. 891–902, 1979.

9. Deacon, T. W., *The Symbolic Species: The Co-evolution of Language and the Brain*, WW Norton & Company, 1998.

10. Nowak, M. A., J. B. Plotkin and D. C. Krakauer, "The Evolutionary Language Game", *Journal of Theoretical Biology*, Vol. 200, No. 2, pp. 147–162, 1999.

11. Krakauer, D. C., "Selective Imitation for a Private Sign System", *Journal of Theoretical Biology*, Vol. 213, No. 2, pp. 145–157, 2001.

12. Hurford, J. R., "Biological Evolution of the Saussurean Sign as a Component of

the Language Acquisition Device", *Lingua*, Vol. 77, No. 2, pp. 187–222, 1989.

13. Nowak, M. A. and D. C. Krakauer, "The Evolution of Language", *Proceedings of the National Academy of Sciences*, Vol. 96, No. 14, pp. 8028–8033, 1999.

14. Cavalli-Sforza, L. L. and F. Cavalli-Sforza, *The Great Human Diasporas: The History of Diversity and Evolution*, Basic Books, 1995.

15. Oliphant, M., "The Dilemma of Saussurean Communication", *BioSystems*, Vol. 37, No. 1, pp. 31–38, 1996.

16. Greenberg, J. H., *Language, Culture, and Communication*, Vol. 2, Stanford University Press, 1971.

17. Chomsky, N., *Language and Thought*, Anshen transdisciplinary lectureships in art, science, and the philosophy of culture, Moyer Bell, 1993.

18. Fodor, J. A., *The Elm and the Expert*, MIT Press, 1994.

19. Greenfield, P. M. and E. S. Savage-Rumbaugh, "Imitation, Grammatical Development, and the Invention of Protogrammar by an Ape.", *Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc*, pp. 235–258, 1991.

20. Brown, R., *A First Language: The Early Stages.*, Harvard University Press, 1973.

21. Sachs, J., "Communication Development in Infancy", *The development of language (4th ed., pp. 40–68). Boston: Allyn & Bacon*, 1997.

22. Newport, E. L., "Maturational Constraints on Language Learning", *Cognitive Science*, Vol. 14, No. 1, pp. 11–28, 1990.

23. Fontanari, J. F. and L. I. Perlovsky, "A Game Theoretical Approach to the Evolution of Structured Communication Codes", *Theory in Biosciences*, Vol. 127, No. 3, pp. 205–214, 2008.

24. Pop, C. M. and E. Frey, "Language Change in a Multiple Group Society", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 88, pp. 1–9, 2013.

25. Tzafestas, E., "On Modeling Proto-Imitation in a Pre-associative Babel", *International Conference on Simulation of Adaptive Behavior*, pp. 477–487, Springer, 2008.

26. Tzafestas, E. S., "Spaces of Imitation", *Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on*, pp. 794–799, IEEE, 2009.

27. Grim, P., T. Kokalis, A. Alai-Tafti, N. Kilb and P. St Denis, "Making Meaning Happen", *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 16, No. 4, pp. 209–243, 2004.

28. Zangwill, N., "Metaphor as Apropriation", *Philosophy and Literature*, Vol. 38, No. 1, pp. 142–152, 2014.

29. Lenaerts, T., B. Jansen, K. Tuyls and B. De Vylder, "The Evolutionary Language Game: An Orthogonal Approach", *Journal of Theoretical Biology*, Vol. 235, No. 4, pp. 566–582, 2005.

30. Steinbacher, M., M. Steinbacher and M. Steinbacher, "Interaction-Based Approach to Economics and Finance", *Complexity in Economics: Cutting Edge Research*, pp. 161–203, Springer, 2014.

31. DeMarzo, P. M., J. Zwiebel and D. Vayanos, "Persuasion bias, social influence, and uni-dimensional opinions", *Social Influence, and Uni-Dimensional Opinions (November 2001). MIT Sloan Working Paper*, Vol. 1, No. 4339-01.

32. Golub, B. and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds", *American Economic Journal: Microeconomics*, Vol. 2, No. 1, pp. 112–149, 2010.

33. Acemoglu, D., A. Ozdaglar and A. ParandehGheibi, "Spread of (mis) information in social networks", *Games and Economic Behavior*, Vol. 70, No. 2, pp. 194–227, 2010.

34. Acemoglu, D., M. A. Dahleh, I. Lobel and A. Ozdaglar, "Bayesian learning in social networks", *The Review of Economic Studies*, Vol. 78, No. 4, pp. 1201–1236, 2011.

35. Jadbabaie, A., P. Molavi, A. Sandroni and A. Tahbaz-Salehi, "Non-Bayesian social learning", *Games and Economic Behavior*, Vol. 76, No. 1, pp. 210–225, 2012.

36. Nowak, M. A., *Evolutionary Dynamics: Exploring the Equations of Life*, Cambridge, MA: Belknap of Harvard University Press, 2006.

37. Dall'Asta, L., A. Baronchelli, A. Barrat and V. Loreto, "Agreement Dynamics on Small-World Networks", *EPL (Europhysics Letters)*, Vol. 73, No. 6, p. 969, 2006.

38. Duda, P. O. and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

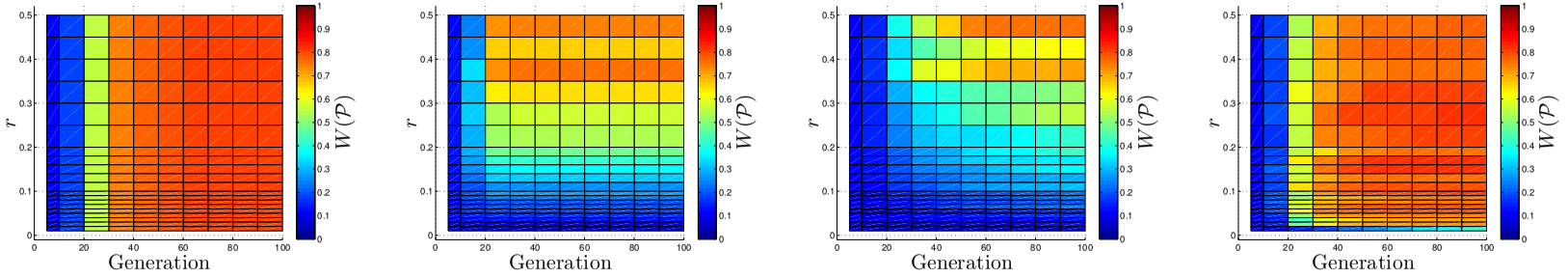
39. Fortunato, S., "Community Detection in Graphs", *Physics Reports*, Vol. 486, No. 3, pp. 75–174, 2010.

APPENDIX A: DIFFERENT GENERATION COUNTS

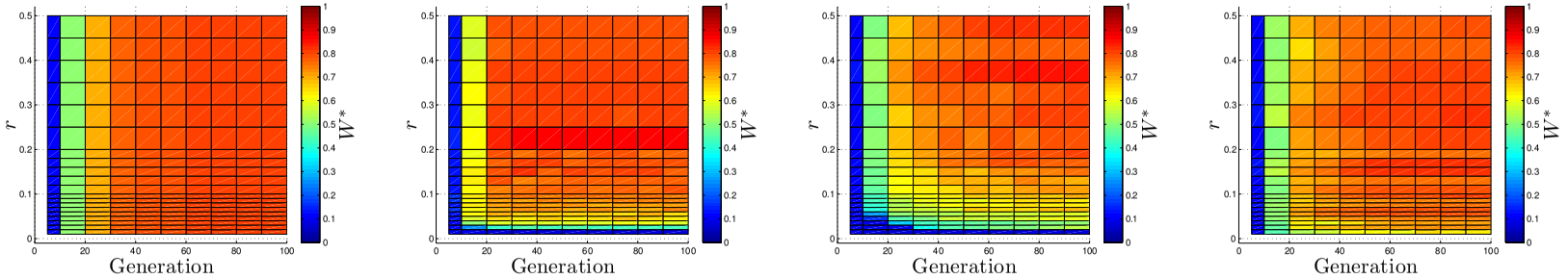
In this appendix, we presented the simulation results for different generation counts. Specifically, results for generations before 100 and results for generations before $1,000$ can be found in Figure A.1 and Figure A.2 respectively. We selected population size $N = 100$, meaning count $M = 8$, signal count $S = 8$ and sampling size $Q = 4$ as parameters. These simulation results are average of 10 repetitions.

In summary, we listed the results for different result types; overall comprehension $W(\mathcal{P})$, optimal mutual comprehension W^*, and optimal cluster count K^* for different models; Base Model, Model-A, Model-B, and Model-C. The figures are organized as follows. Figures sharing the same row contain certain result type for different models, whereas figures sharing the same column contain different simulation result types for certain model.

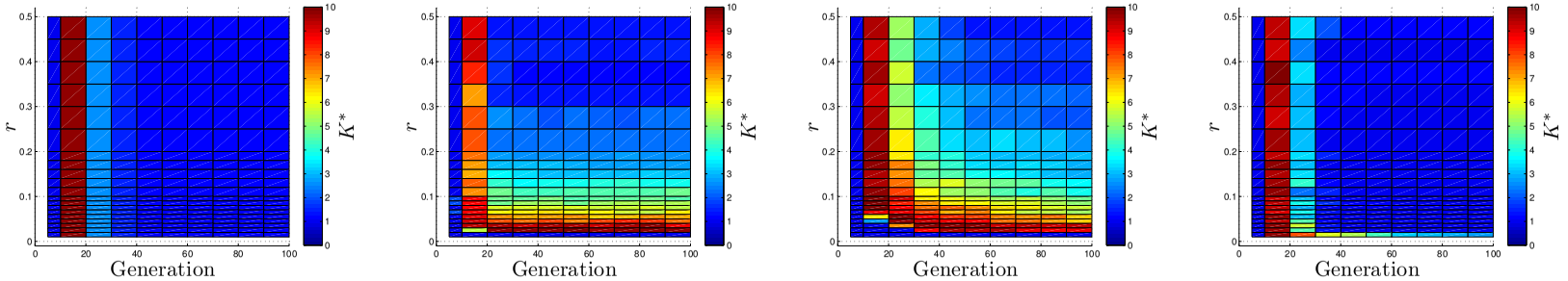
As we can see in Figure A.1a, Figure A.1b and Figure A.1d overall comprehension $W(\mathcal{P})$ remains stationary after generation 40. In Figure A.1c we see that $W(\mathcal{P})$ continues to change for some generations later, until generation 400 as we see in Figure A.2c. In results shared 2nd and 3rd rows, some fluctuation can be detected. In corresponding simulations we used k-means algorithm, thus this type of uncertainty in the result is expected since k-means is heuristic algorithm. Regarding generations until $1,000$, we could not detect any significant change in the results.



(a) $W(P)$ in Base Model. (b) $W(P)$ in Model-A. (c) $W(P)$ in Model-B. (d) $W(P)$ in Model-C.

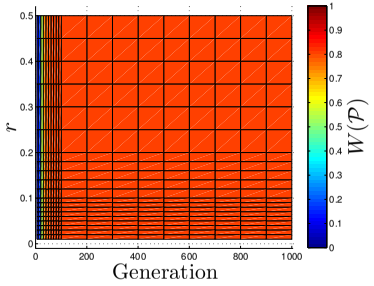


(e) W^* in Base Model. (f) W^* in Model-A. (g) W^* in Model-B. (h) W^* in Model-C.

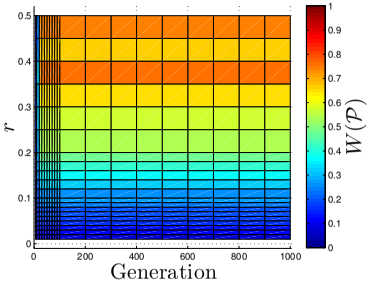


(i) K^* in Base Model. (j) K^* in Model-A. (k) K^* in Model-B. (l) K^* in Model-C.

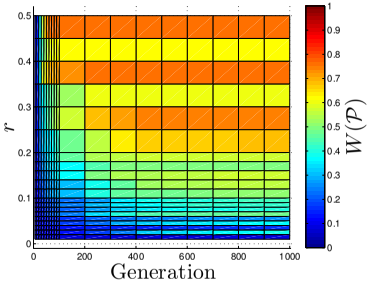
Figure A.1. Change of values through generations [0, 100] for simulations of $(N,M,S,Q) = (100,8,8,4)$.



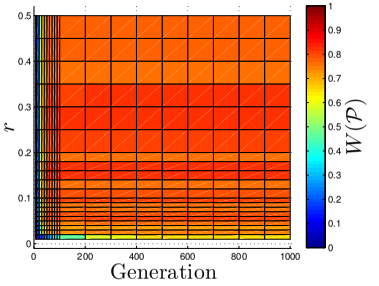
(a) $W(P)$ in Base Model.



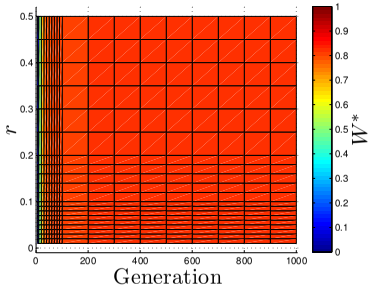
(b) $W(P)$ in Model-A.



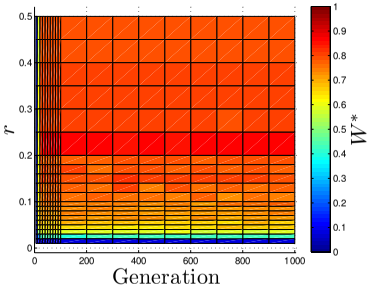
(c) $W(P)$ in Model-B.



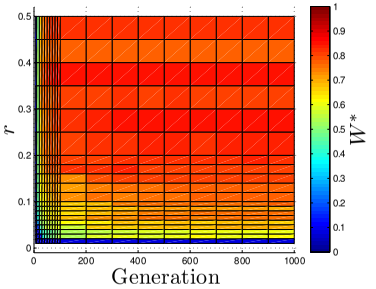
(d) $W(P)$ in Model-C.



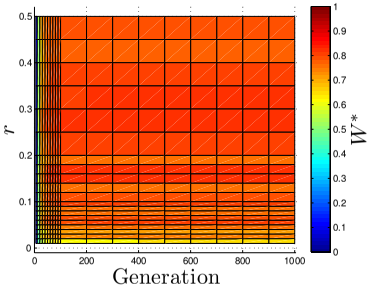
(e) W^* in Base Model.



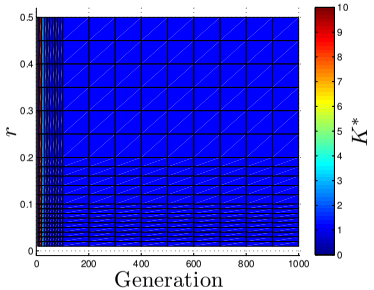
(f) W^* in Model-A.



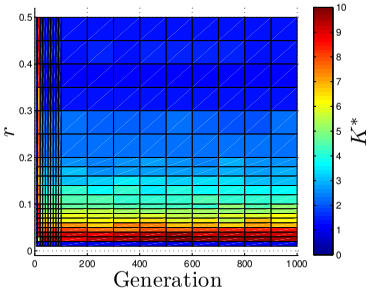
(g) W^* in Model-B.



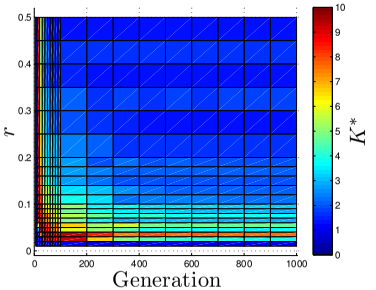
(h) W^* in Model-C.



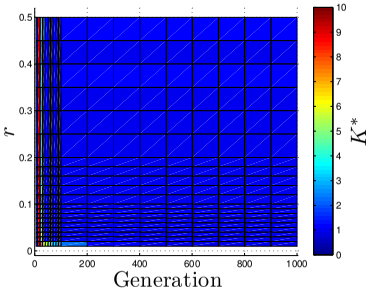
(i) K^* in Base Model.



(j) K^* in Model-A.



(k) K^* in Model-B.



(l) K^* in Model-C.

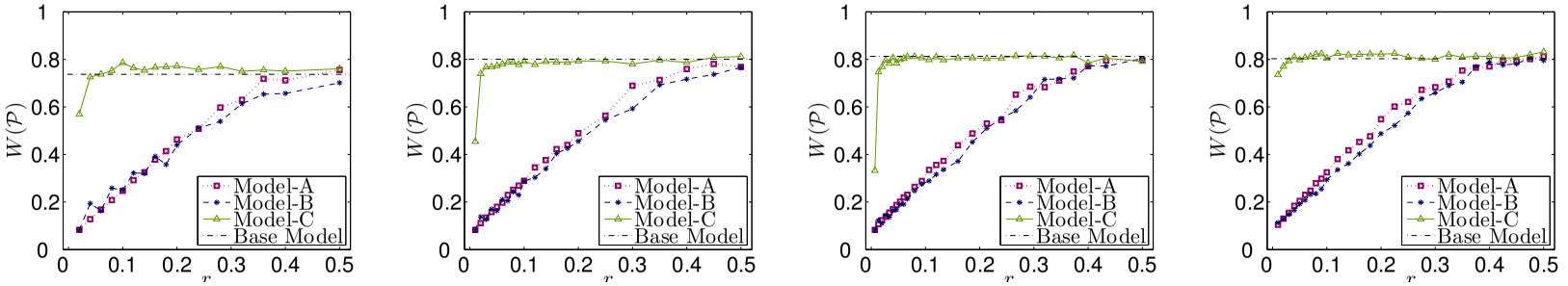
Figure A.2. Change of values through generations [0, 1000] for simulations of $(N,M,S,Q) = (100,8,8,4)$.

APPENDIX B: DIFFERENT POPULATION SIZES N

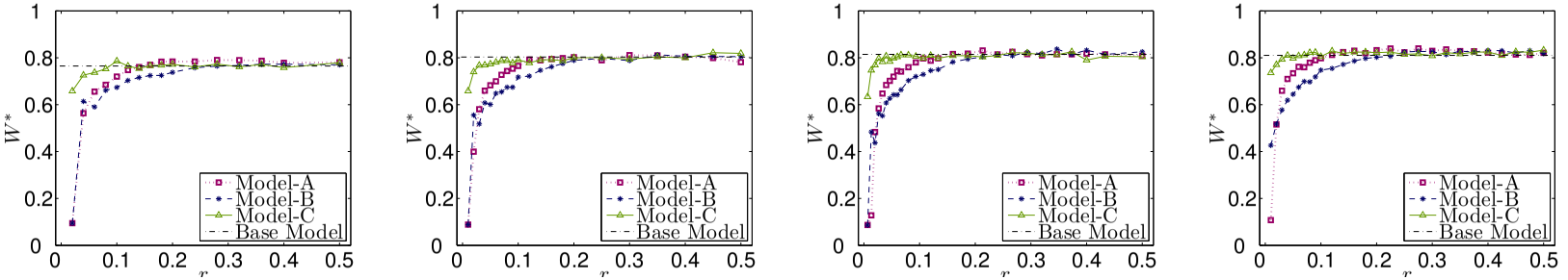
In this appendix, we presented the simulation results for different population sizes. Specifically, simulation results for $N = 50, 100, 150, 200$ can be found in Figure B.1. We selected meaning count $M = 8$, signal count $S = 8$ and sampling size $Q = 4$ as parameters. These simulation results are average of 100 repetitions.

In summary, we listed the results for different result types; overall comprehension $W(\mathcal{P})$, optimal mutual comprehension W^*, and optimal cluster count K^* for different population sizes. The figures are organized as follows. Figures sharing the same row contain certain result type for different population sizes, whereas figures sharing the same column contain different result types for certain population size.

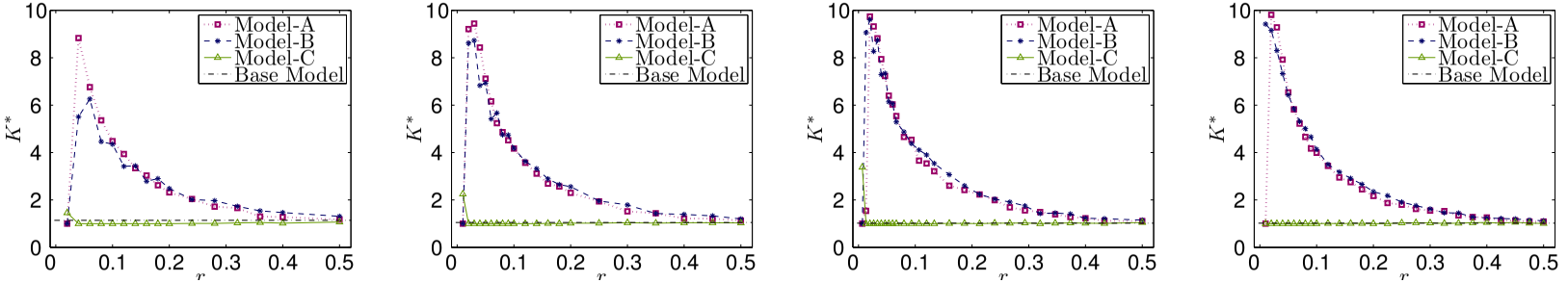
As we can see in these figures, there is no significant change in results for different population sizes. This finding supports our argument about the optimum community count K^* being independent of N. Furthermore, corresponding overall comprehension $W(\mathcal{P})$ and optimal mutual comprehension W^* seem to be not effected by the change in N, too.



(a) $W(P)$ in $N = 50$. (b) $W(P)$ in $N = 100$. (c) $W(P)$ in $N = 150$. (d) $W(P)$ in $N = 200$.



(e) W^* in $N = 50$. (f) W^* in $N = 100$. (g) W^* in $N = 150$. (h) W^* in $N = 200$.



(i) K^* in $N = 50$. (j) K^* in $N = 100$. (k) K^* in $N = 150$. (l) K^* in $N = 200$.

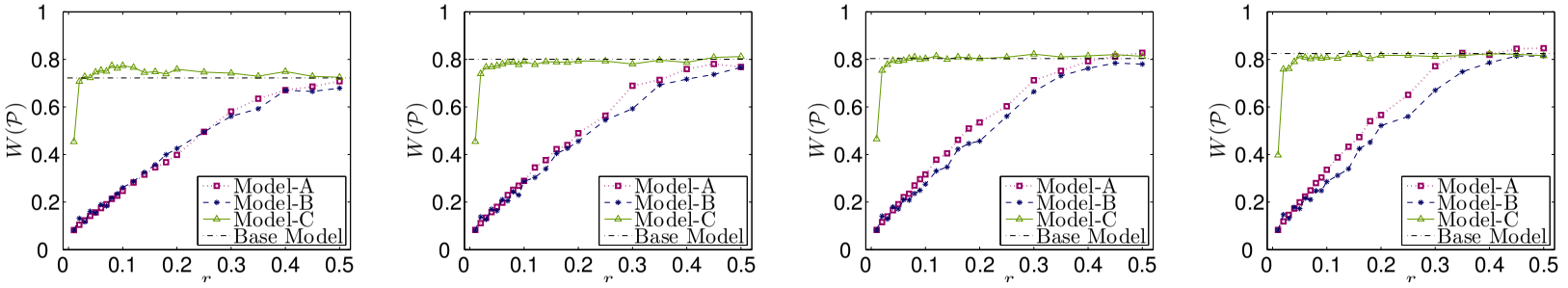
Figure B.1. Simulation results of different population sizes for $(M,S,Q) = (8,8,4)$.

APPENDIX C: DIFFERENT SAMPLING SIZES Q

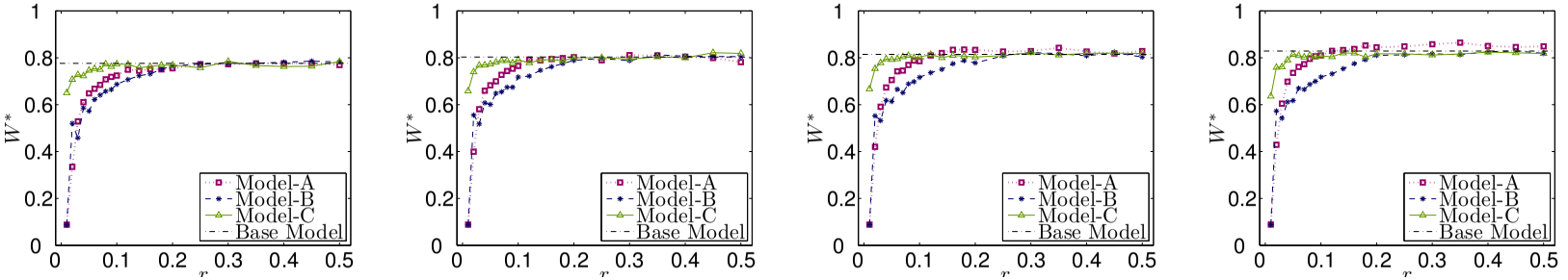
In this appendix, we presented the simulation results for different sampling sizes. Specifically, simulation results for $Q = 1, 4, 7, 10$ can be found in Figure C.1. We selected population size $N = 100$, meaning count $M = 8$ and signal count $S = 8$ as parameters. These simulation results are average of 100 repetitions.

In summary, we listed the results for different result types; overall comprehension $W(\mathcal{P})$, optimal mutual comprehension W^*, and optimal cluster count K^* for different sampling sizes. The figures are organized as follows. Figures sharing the same row contain certain result type for different sampling sizes, whereas figures sharing the same column contain different result types for certain sampling size.

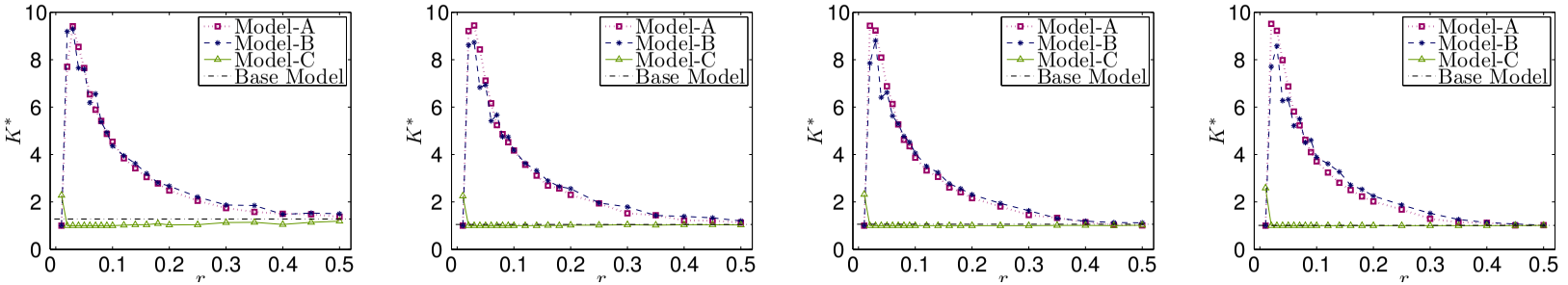
As we can see in these figures, results are uplifted as Q is changed from 1 to 4. On the other hand, for $Q = 4, 7, 10$ there is no significant change in the results.



(a) $W(P)$ in $Q = 1$. (b) $W(P)$ in $Q = 4$. (c) $W(P)$ in $Q = 7$. (d) $W(P)$ in $Q = 10$.



(e) W^* in $Q = 1$. (f) W^* in $Q = 4$. (g) W^* in $Q = 7$. (h) W^* in $Q = 10$.



(i) K^* in $Q = 1$. (j) K^* in $Q = 4$. (k) K^* in $Q = 7$. (l) K^* in $Q = 10$.

Figure C.1. Simulation results of different sampling sizes for $N = 100$, $M = 8$, $S = 8$.

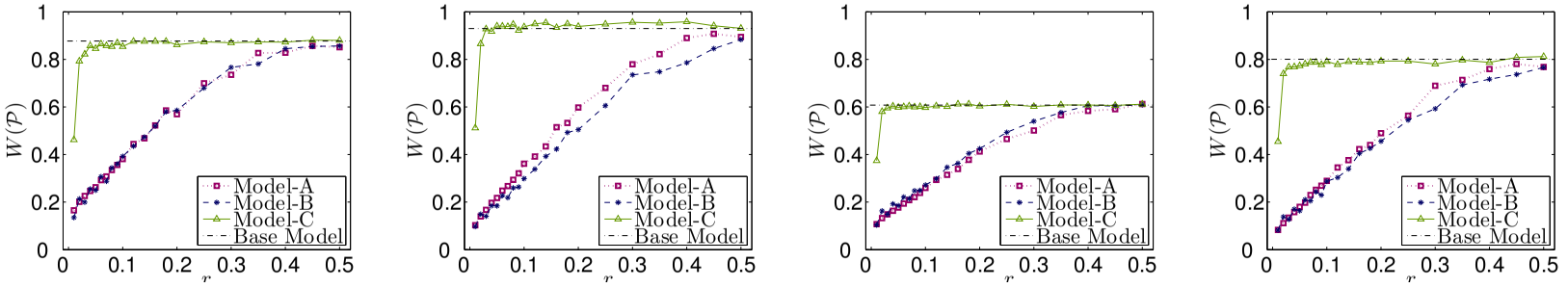
APPENDIX D: DIFFERENT (M, S) PARAMETERS

In this appendix, we presented the simulation results for different meaning and signal counts. Specifically, simulation results for $(M, S) = (5, 5), (5, 8), (8, 5), (8, 8)$ can be found in Figure D.1. We selected population size $N = 100$ and sampling size $Q = 4$ as parameters. These simulation results are average of 100 repetitions.

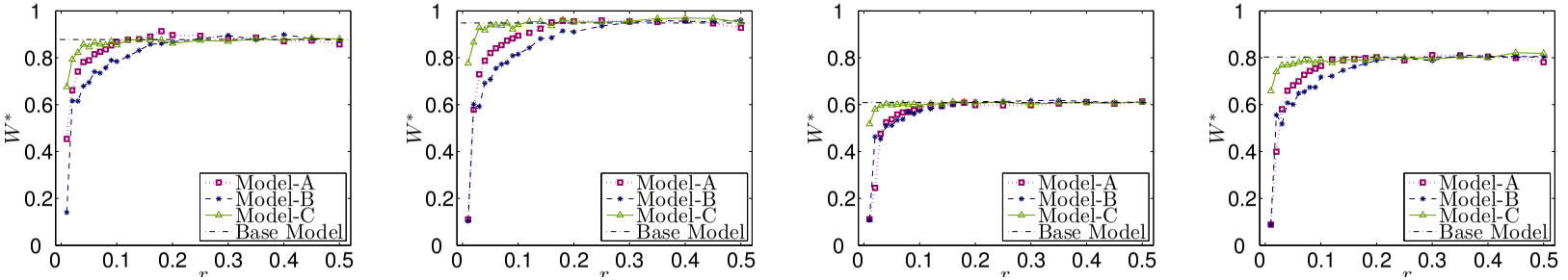
In summary, we listed the results for different result types; overall comprehension $W(\mathcal{P})$, optimal mutual comprehension W^*, and optimal cluster count K^* for different (M, S) pairs. The figures are organized as follows. Figures sharing the same row contain certain result type for different (M, S) pairs, whereas figures sharing the same column contain different result types for certain (M, S) pair.

As we can see in these figures, the ordering of the results from high to low according to comprehension values is $(M, S) = (5, 8)$, $(M, S) = (5, 5)$, $(M, S) = (8, 8)$ and $(M, S) = (8, 5)$. As we mentioned in Section 3.2, ambiguity causes comprehension value to be worse. Therefore it is expected to see very low comprehension value in case where $M > S$ which is covered by $(M, S) = (8, 5)$ and higher comprehension values when $M < S$ which is covered by $(M, S) = (5, 8)$.

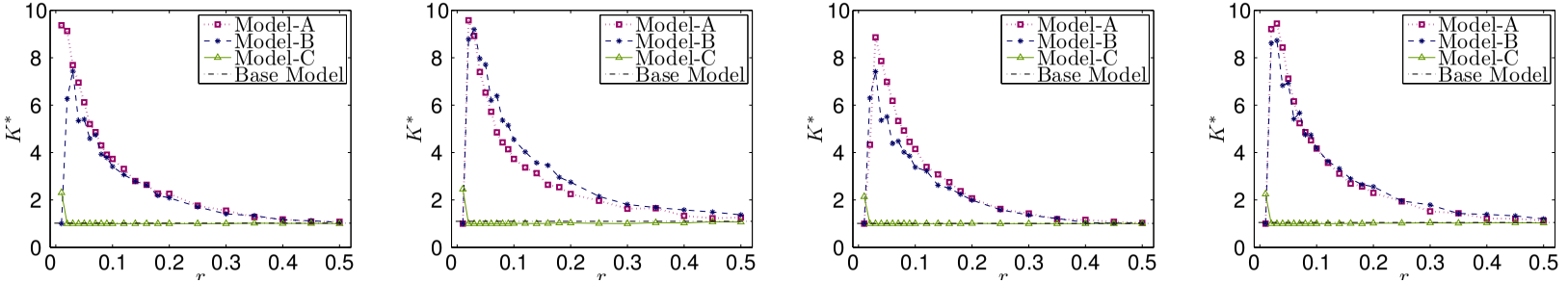
On the other hand we observed better comprehension values in $(M, S) = (5, 5)$ case than $(M, S) = (8, 8)$ case. This is caused by the selection of sampling size. As we mentioned in Section 3.4 and Appendix C, as sampling size Q gets higher, comprehension gets better, since more signals can be transmitted from teacher to the child and there is less chance for some signals to get lost. When M, S gets lower, it causes the similar situation described above where Q gets higher.



(a) $W(P)$ in $(M, S) = (5, 5)$. (b) $W(P)$ in $(M, S) = (5, 8)$. (c) $W(P)$ in $(M, S) = (8, 5)$. (d) $W(P)$ in $(M, S) = (8, 8)$.



(e) W^* in $(M, S) = (5, 5)$. (f) W^* in $(M, S) = (5, 8)$. (g) W^* in $(M, S) = (8, 5)$. (h) W^* in $(M, S) = (8, 8)$.



(i) K^* in $(M, S) = (5, 5)$. (j) K^* in $(M, S) = (5, 8)$. (k) K^* in $(M, S) = (8, 5)$. (l) K^* in $(M, S) = (8, 8)$.

Figure D.1. Simulation results of different meaning and signal counts for $N = 100$, $Q = 4$.

APPENDIX E: AVERAGE INTER-COMMUNITY COMPREHENSION I^*

In this appendix, we presented the average inter community comprehension values. We selected population size $N = 100$, meaning count $M = 8$, signal count $S = 8$ and sampling size $Q = 4$ as parameters. These simulation results are average of 100 repetitions.

As we can see in Figure E.1, we observe that $I^* <= 1.5$ for each models. We know that $W_{\mathrm{r}}(\mathcal{P}) = 1.25$ from Equation 3.2. Thus, we can conclude that average mutual comprehension between detected communities are found approximately random.

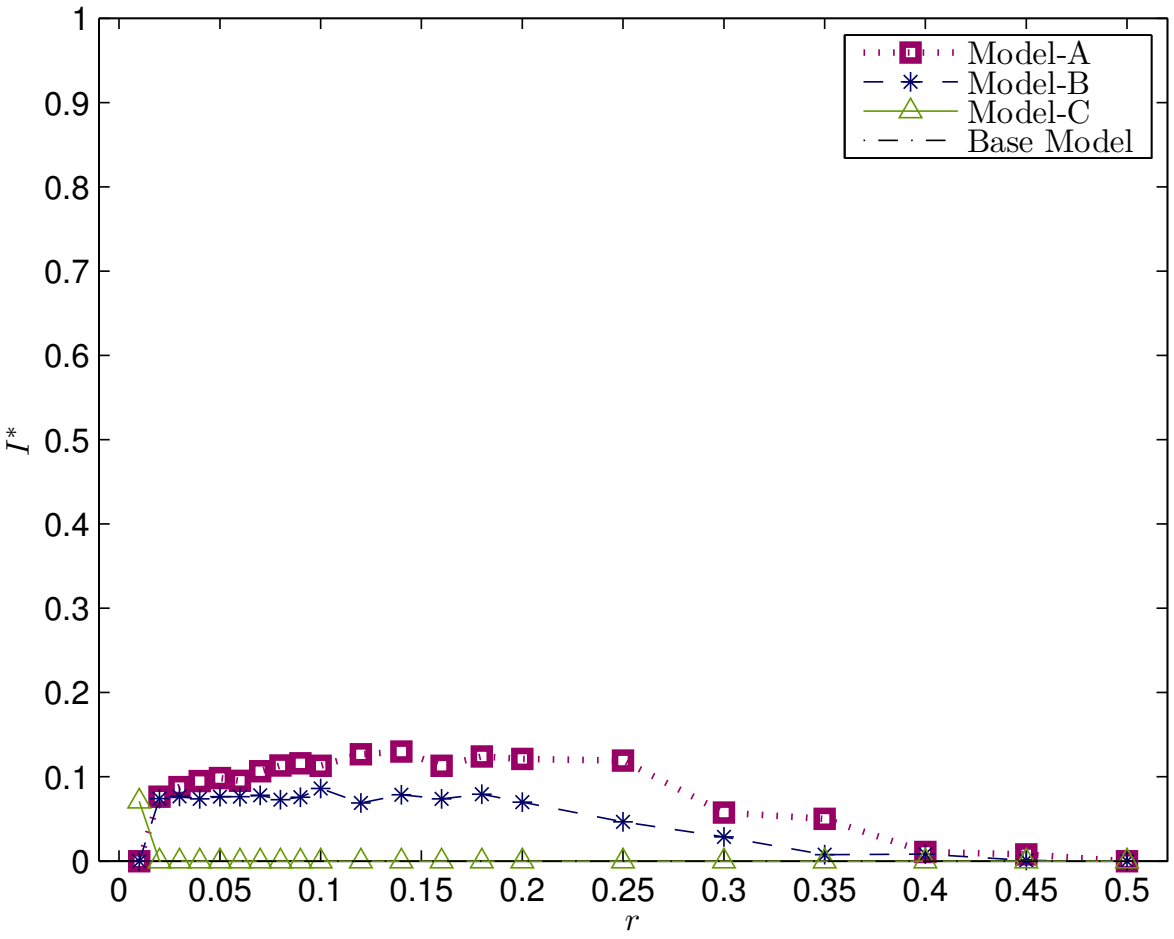


Figure E.1. Average inter-community comprehension value for simulations of $(N, M, S, Q) = (100, 8, 8, 4)$.