# How to Reach Linguistic Consensus: A Proof of Convergence for the Naming Game

Bart De Vylder [a,*] Karl Tuyls [b]

[a] *Artificial Intelligence Lab, Vrije Universiteit Brussel, Belgium*

[b] *Institute for Knowledge and Agent Technology (IKAT), Universiteit Maastricht, The Netherlands*

**Abstract**

In this paper we introduce a mathematical model of naming games. Naming games have been widely used within research on the origins and evolution of language. Despite the many interesting empirical results these studies have produced, most of this research lacks a formal elucidating theory. In this paper we show how a population of agents can reach linguistic consensus, i.e. learn to use one common language to communicate with one another. Our approach differs from existing formal work in two important ways: One, we relax the too strong assumption that an agent samples infinitely often during each time interval. This assumption is usually made to guarantee convergence of an empirical learning process to a deterministic dynamical system. Two, we provide a proof that under these new realistic conditions, our model converges to a common language for the entire population of agents. Finally the model is experimentally validated.

*Key words:* Language Acquisition, Naming Games, Sampling-Response model, Amplification, Cultural Evolution, Horizontal Transmission

## 1 Introduction

Research in the domain of origins and evolution of language has received considerable attention in recent years for the obvious reason that language is one of the most prominent forms of communication between individuals. One way in which the origins and evolution of language can be studied, is using the paradigm of *language games*, see e.g. Steels (2003).

* Corresponding author.
*Email addresses:* `bartdv@arti.vub.ac.be` (Bart De Vylder), `k.tuyls@cs.unimaas.nl` (Karl Tuyls).

In this paper we present a proof of convergence of the naming game, which is a particular type of language game in which agents use proper names to refer to objects. The main property of the mathematical model we introduce, is the explicit modeling of the limited view individuals have of the different object-word associations used by other individuals in the population. Because of this, some agents will inevitably misjudge the population's state and change their behavior in a disadvantageous manner with respect to convergence. Yet, in this paper it is shown that the population will converge to the same language under rather general conditions. In particular, we prove that if the agents locally learn to reduce synonymy, this same behavior emerges at the collective, population level.

In language games in general, the focus is on the horizontal transmission of language within a population, which means that agents build up a language by interacting in a peer-to-peer fashion. The main difference with vertical transmission schemes inspired by biological evolution as discussed in Nowak et al. (2001, 1999); Komarova and Nowak (2001) is the lack of the notion of fitness: it is not assumed that better communication abilities result in more offspring and that children learn the language from their parents (or equivalently, from individuals with higher fitness). On the contrary, every individual is considered equally worth learning from. Another model using vertical transmission, the 'Iterated Learning Model', e.g. see Hurford and Kirby (2001); Smith (2004), assumes that the strongest force shaping a language is not its communicative function but its learnability. To study this, successive generations of agents are considered with a strict distinction between agents acting as a teacher and agents acting as a learner. This also contrasts with the horizontal transmission scheme in which an agent simultaneously influences and gets influenced by other agents' languages.

The horizontal transmission scheme has been used in research using computational models, e.g. Steels (2000, 2005, 1998, 2003); Steels and Belpaeme (2005); Zuidema and Westermann (2003), as well as in theoretical work by Cucker et al. (2004) and recently also in Matsen and Nowak (2004). Furthermore, in Lenaerts et al. (2005), the relation was drawn between the horizontal transmission scheme on the one hand and reinforcement learning and selection-mutation dynamics on the other hand. Most research has been conducted using computer simulations. While we believe this is a valuable tool to investigate issues concerning the origins and evolution of language, we think the research could greatly benefit from theoretical foundations for the following reason. There is always an intrinsic limitation in the explanatory power of a simulation. The observed phenomena can only be assumed to hold, based on some runs and under some specific parameter settings. Sometimes an extrapolation is possible. A theoretical model forces one to search for the essential features of a particular system which can explain a certain phenomenon. Nevertheless, relatively few theoretical results exist which confirm or explain the dynamics of linguistically interacting individuals (but see Cucker et al. (2004); Nowak et al. (2001, 1999); Komarova and Nowak (2001)). In this paper we try to make progress in this domain and present a model that guarantees that a population of agents, with a certain behavior, will certainly converge to a state, in which most or all of the agents use the same language.

The remainder of this document is structured as follows. The naming game in general is described in Section 2, together with the assumptions we make for the construction of the mathematical model. In Section 3 the translation of the naming game to the mathematical model, named the sampling-response model, is explained and compared with other mathematical models in the literature. Also the central concept of amplification is introduced. In Section 4, the validity of this model is examined through a comparison with the original game. Finally, we end with some conclusions. The appendices A through D elaborate on the mathematical issues of the model. They are referred to throughout the text.

## 2 The Naming Game: Introduction and Assumptions

The paradigm of language games suggests that language can arise culturally through verbal and non-verbal peer-to-peer interaction between agents which are situated in the same environment. One of the first and most basic language games is called the naming game (see e.g. Steels (1995)). This game investigates the way a population of language users can reach a consensus on how to name different objects. Hereby the words used are proper names: they refer directly to particular objects. A naming game is played between two agents which temporarily take up the role of speaker and hearer. They are situated in the same environment which contains a number of different objects. The speaker chooses one of the objects to be the topic and utters a word for it. Based on this name, the hearer guesses what the topic is, which determines whether the game is successful or not. In either case, after the game the hearer knows the topic (in case of failure the speaker points to it) and can use this information in future games.

In this framework, we identify a language with the associations an agent makes between words and objects, acting as a speaker and a hearer. We consider a population of agents which play naming games among each other. The question arises under what circumstances the agents will end up using the same language and whether this resulting language is effective and optimal. A language is effective if one can play successful naming games with it. This implies that every object must be associated with at least one word which identifies it unambiguously, this implies that homonymy must not occur. A language is optimal if it is parsimonious: no more words are used than necessary. This means that synonyms should not occur. Altogether, an effective, optimal language associates with each object one unique word [1] .

We will now further elaborate on this model in order to obtain a description suitable for mathematical analysis. One important assumption we make is that homonymy cannot occur as a result of the following observations: In the beginning, no agent knows a word for any object. If and only if an agent acts as a speaker and has to talk about an object he does not have a word for, he invents a new word. Thereby we assume that the event of a new word being the same as an existing one, is negligible. Furthermore, when an agent, acting as hearer, hears a word he does not know, he waits until the speaker points to the topic to decide with which object he will associate that word. These assumptions

[1] Although this is a strong simplification with regard to human language, in which synonymy and homonymy frequently occur, we believe these assumptions are a logical first step in studying the problem of reaching linguistic consensus. Moreover we think this simplified setting provides meaningful insights in the theoretical aspects of more complicated/realistic language settings.

imply that homonymy will not occur. In other words, the sets of words that are used for the different objects, are disjoint. This means that the dynamics of the synonyms used for each object are independent. Moreover, as we assume no bias in the objects, these dynamics are identical. Therefore it suffices to study the dynamics of the synonyms used for only one object, which is the approach taken in the rest of the paper.

Important to understand now is that, if the population has reached a point in which homonymy and synonymy are absent, the topic-guess of a hearer will always be correct. Also, the topic plus the name used by the speaker to refer to it, is sufficient for a hearer to update its internal state. So he does not explicitly need to use the correctness of its guess to update his state. Consequently, we can study the dynamics of the naming game ignoring the guessing part. As a result we can reduce a game to a one-way interaction in which a speaker points to an object and simultaneously utters a word for it.

With regard to the structure of the population, we make the assumption that every agent is equally likely to play a game with any other agent. While this assumption is probably not strictly necessary to reach convergence (as is analyzed in more detail in Cucker et al. (2004), yet in another setting), all agents must be somehow linguistically connected (one cannot expect two independent populations to start speaking the same language).

Now consider a population of agents, which communicates about one object by giving names to it. In order to estimate the number of different words that could arise during the naming game we make the following observation: a new word is only created when a speaker does not yet have a word for the object. Under the previous assumptions, this means that this speaker has never taken part in a naming game before, as speaker nor hearer. Eventually, every agent will have participated at least once in a naming game and from that time on, no new words are created anymore. The minimum number of words that can arise is 1: all agents, except one, are hearer in their first game. The maximum number of words is $N-1$ with $N$ the number of agents: all agents, except one, are speaker in their first game. If we assume that the speaker and hearer are chosen at random from the population, chances that an agent is a speaker in its first game equals 1/2. Therefore, the average number of words that will be created is $N/2$.

For the sake of completeness, we make an estimate of the number of games it takes to reach such a state. First, suppose that at every time step we randomly choose one agent from a population of $N$ agents (with replacement). Suppose that $i$ distinct agents have already been chosen at least once, then the chance of choosing a new agent equals $\frac{N-i}{N}$. So, the expected number of steps needed to go from $i$ to $i+1$ chosen agents, $t_i$, equals $\frac{N}{N-i}$. The expected number of steps to go from 0 to $N$ agents is then given by:

$$\sum_{i=0}^{N-1} t_i = \sum_{i=0}^{N-1} \frac{N}{N-i} = N\sum_{i=1}^{N} \frac{1}{i} \approx N(\log(N)+\gamma) \approx N\log(N),$$

with $\gamma$ the Euler-Mascheroni [2] constant. Let us now turn to the case where at each time step two agents, the speaker and hearer, are chosen at random. This process is equal to two times choosing one agent at random, except that the speaker and hearer are always

[2] The Euler-Mascheroni constant is the limit of the difference between the harmonic series and the natural logarithm: $\lim_{N\to\infty}((\sum_{i=1}^{N}\frac{1}{i}) - \log(N))$.

(a) $\Sigma_2$ (b) $\Sigma_3$

Fig. 1. The unit simplices $\Sigma_2$ and $\Sigma_3$.

different, which can only speed up the process. So, the expected number of games after which every agent has at least once participated, will be approximately

$$\frac{1}{2} N \log(N).$$

With this description of the naming game and the additional assumptions made, we are now ready to introduce our mathematical model, which is the topic of the next section.

## 3 The sampling-response model

In the remainder of the paper we will refer to the naming game as described above as the *turn-based* model. One of the most important properties is its stochasticity. This because the participators in a game are chosen at random but also the word used by a speaker may be chosen randomly according to some distribution. In this section we will transform this turn-based model to a model which is mathematically more tractable. We will name this new, deterministic model, the *sampling response* model.

### *3.1 Mathematical Preliminaries*

We first introduce some notations and definitions. Given $n$ elements which are randomly chosen with probabilities $(x_1, x_2, \ldots, x_n)$, there holds $x_1, x_2, \ldots, x_n \geq 0$ and $\sum_{i=1}^n x_i = 1$. We denote the set of all such probability distributions over $n$ elements as $\Sigma_n$ or simply $\Sigma$ if there is no confusion possible. $\Sigma_n$ is a $n-1$-dimensional structure called a simplex.[3] For example in Figure 1, $\Sigma_2$ and $\Sigma_3$ are shown. In the figures throughout the paper we use $\Sigma_3$, projected as an equilateral triangle as in Figure 1(b), but we drop the axes and labels.

[3] One degree of freedom is lost due to the constraint $\sum_{i=1}^n x_i = 1$.

Fig. 2. The set of possible word frequencies, given three words and a queue of length 6. Because there are three words these frequencies are elements of the 2-dimensional simplex $\Sigma_3$. For clarity, the numbers indicate the actual number of occurrences of each word, in order to obtain the frequencies these vectors have to be normalized, i.e. divided by 6.

### *3.2 Formulation of an aggregate sampling-response model*

Under the assumptions made in the previous section, we now have a population of agents which name a certain object using a certain set of $n$ words, say $\{w_1, w_2, \ldots, w_n\}$ which does not grow anymore. We now will explain the sampling-response model using an illustrative example of a language game with a particular kind of agent: the *queue-agent*. This agent just remembers the last $K$ words he heard. So the agent can be seen as a First-In-First-Out queue of length $K$: when acting as hearer, the agent drops the last word and adds the new word to the front of the queue. Furthermore the agent, when acting as speaker, chooses a word from this queue based on the frequencies [4] with which the different words occur. For example, suppose an agent has a buffer of length 6 and there are three words, $w_1$, $w_2$ and $w_3$ used in the population. If $m_i$ represents the number of times $w_i$ occurs in the queue, there are 28 possible compositions $(m_1, m_2, m_3)$: $(6, 0, 0), (5, 1, 0), (4, 1, 1), \ldots (0, 0, 6)$. Each such composition corresponds to certain frequencies $(x_1, x_2, x_3)$ for these words, e.g. $\frac{1}{6}(3, 1, 2) \approx (0.5, 0.17, 0.33)$. In Figure 2 these frequencies are shown as points on $\Sigma_3$. Based on these frequencies, the agent has to decide which word to use when he speaks or with which probabilities $(y_1, y_2, y_3)$ to choose between the words $w_1$, $w_2$ and $w_3$. Thus, the strategy an agent uses for this is in fact a function which maps its input $(x_1, x_2, x_3)$ to its output $(y_1, y_2, y_3)$. We name this function an agent's response function

$$s : \Sigma \to \Sigma.$$

An intuitively obvious strategy is to choose the word which occurred most. This is indeed a special case [5] of a class of response functions which are named *amplifying*. The next

[4] Throughout the paper we will use the term 'frequencies' to mean relative frequencies.
[5] Strictly speaking this is not correct, but it is the limit case for $m \to \infty$ in Example 5.

section explains in more detail the definition of an amplifying function. For now, it suffices to mention that one important property of an amplifying function is that the word with the highest frequency in the input, gains an even higher frequency in the output. As a consequence, an iteration of an amplifying map will result in one word predominating the others. Having said that, the derivation in this section holds for any response function $s$, hence also non-amplifying functions.

In the turn-based model, language games are played between two randomly chosen agents. During each game, the speaker chooses a word according to the content of its queue and this word enters the hearer's queue. From a hearer's point of view, it is not important which agent is the speaker, but only with what probabilities he will hear the different words. This probability distribution is the average of all[6] the agents' distributions for choosing between words and hence it is the same for each agent acting as hearer. We denote this distribution as $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$, with $\tau \in \Sigma$.

Now, in principle this $\tau$ might change after each game, because the state of the hearer could have changed which in turn can lead to a new speaking behavior of that agent. To arrive at an aggregate model, however, we will keep $\tau$ temporarily constant, just as long it takes for each agent to 'flush' its queue, i.e. to be $K$ times a hearer in a game. We call this an episode of games. After that, the new word distribution $\tau'$ is determined and we start all over again. One can argue that this modification will lead to a slower convergence, as the change in an agent's state can only be observed after an episode, while in the turn-based model it was immediate. Hence if we can prove convergence in this model, there is strong evidence the original model will converge as well.

The reason for keeping $\tau$ temporarily constant is that we can statistically predict what will be the content of the agents' queues after an episode and consequently predict the word frequencies in a queue. Therefore we define the *sampling set* $S$ as the set of all possible word frequencies in a queue (for example the set of frequencies displayed in Figure 2) and introduce the *sampling function*

$$w : \Sigma \to (S \to \mathbb{R}_{\geq 0}), \tag{1}$$

such that, given the word distribution on population level $\tau$, $w_\tau$ is a probability distribution over $S$ which models the outcome of a sampling of $\tau$. In particular, for the queue-agent introduced before this becomes a multinomial distribution:

$$w_\tau(\frac{1}{K}(m_1, m_2, \ldots, m_n)) \;=\; \frac{K!}{\prod_{i=1}^{n} n_i!} \prod_{i=1}^{n} \tau_i^{n_i} \tag{2}$$

For instance, suppose we have three words ($w_1$, $w_2$ and $w_3$), $\tau = (0.3, 0.2, 0.5)$ and a queue length $K$ equal to 6, then the probability of observing word frequencies $\frac{1}{6}(5, 0, 1)$, i.e. 5 $w_1$'s and 1 $w_3$ amounts to $6 \;\, 0.3^5 \;\, 0.5 = 0.00729$.

We will reasonably assume that the expected value of a sampling of $\tau$ yields $\tau$ itself. In

[6] For simplicity we allow an agent to speak to itself, although this happens rarely in large populations.

other words

$$\sum_{\sigma \in S} w_\tau(\sigma)\sigma = \tau \qquad \text{for all } \tau \in \Sigma \tag{3}$$

In case of the queue-agent with its associated multinomial distribution this requirement is met.

Now we have all the ingredients to construct a transition function for the population: Given the current word frequencies, $\tau$, in the population, we know through the sampling function $w$ with what probabilities the agents will observe the different possible word frequencies of the set $S$. Also, for each element of $S$, say $\sigma$, we know what the future speaking behavior (word frequencies) will be of an agent observing that element, namely $s(\sigma)$. So if we make the sum of these new frequencies, $s(\sigma)$, weighted with the probability $\sigma$ was encountered, $w_\tau(\sigma)$, we arrive at the average new state of the population $\tau'$, i.e. the new frequencies with which the different words are spoken in the population. Hence the transition function $t : \Sigma \to \Sigma$ is defined as follows:

$$t(\tau) = \sum_{\sigma \in S} w_\tau(\sigma) s(\sigma) \tag{4}$$

with $\tau' = t(\tau)$.

The main contribution of this paper is the proof (in appendix B) that if the agents behave amplifying, i.e. if $s$ is an amplifying function, that, under certain conditions on the sampling set $S$ and sampling function $w$, the amplification will be lifted toward the global transition function $t$. As already mentioned, this implies that a repeated application of the map $t$, describing the trajectory of the population's state, will result in the use of only one word, or $t^\infty(\tau) = (0, \ldots, 0, 1, 0, \ldots, 0)$. The conditions on $S$ and $w$ in general are stated explicitly in appendices A and B. In the particular case of the queue-agent these translate to the requirement that $K \geq 3$, i.e. the queue should be at least three long (see appendix D).

In addition, we will also investigate some non-amplifying response functions and their influence on the transition function $t$ and on the convergence of the sampling-response and turn-based model.

### *3.3 The response function*

An agent's response function is a function $s : \Sigma \to \Sigma$, which maps its observed word frequencies to the probabilities with which he uses the different words as a speaker. We now (informally) define several properties such a function might possess. For the formal definitions we refer to appendix A. A response function $s$ is

**symmetrical** if a permutation of the input frequencies results in the same permutation of the output frequencies,
**order preserving** if it preserves the order of the word frequencies
**weakly amplifying** if it increases the frequency of the most frequent word,
**amplifying** if the cumulative sum of the frequencies—sorted in decreasing order—increases from input to output.

All response functions we consider will be symmetrical, as this ensures that the agents don't have an ingrained preference among the words. Also, all response functions will be order preserving. On the other hand we will introduce both amplifying and non-amplifying functions. Further we have that an amplifying function is necessarily weakly amplifying, but the converse is not generally true. Finally, it is easy to see that an iteration of a (weakly) amplifying function will result in one word predominating the others (see also proposition 2 in Appendix A).

An example of an amplifying function is $s_{\mathrm{A}}$ defined by

$$[s_{\mathrm{A}}(\sigma)]_i = \frac{\sigma_i^m}{\sum_{j=1}^{n} \sigma_j^m} \tag{5}$$

for all $\sigma \in \Sigma$ and with $m \in \mathbb{R}$ and $m > 1$. This is proved in appendix C. The corresponding global transition function is written as $t_{\mathrm{A}}$.

To illustrate, suppose an agent has a buffer of length 12 and observed four words $w_1$ (= A), $w_2$ (= B), $w_3$ (= C) and $w_4$ (= D) respectively 1, 4, 5 and 2 times. Then these words occur with the following frequencies:

| | A | B | C | D |
|---|---|---|---|---|
| $\sigma_i$ | 0.083 | 0.333 | 0.417 | 0.167 |

We now apply the amplifying map (5) with $m = 2$: each frequency is squared and the result is normalized such that its sum equals 1 again.

| | A | B | C | D |
|---|---|---|---|---|
| $\sigma_i$ | 0.083 | 0.333 | 0.417 | 0.167 |
| $\sigma_i^2$ | 0.007 | 0.111 | 0.174 | 0.027 |
| $[s_{\mathrm{A}}(\sigma)]_i$ | 0.022 | 0.348 | 0.543 | 0.087 |

Note that the frequency of C, which occurred most often, increases, but that the frequencies of the other words can as well increase as decrease, e.g. B increased and A and D decreased. This transformation of the frequencies is symmetrical, order preserving (in both input and output we have the following order: C > B > D > A) and amplifying:

| | C | B | D | A |
|---|---|---|---|---|
| $\sigma$ sorted | 0.417 | 0.333 | 0.167 | 0.083 |
| $s_{\mathrm{A}}(\sigma)$ sorted | 0.543 | 0.348 | 0.087 | 0.022 |
| $\sigma$ sorted cumulative | **0.417** | **0.750** | **0.917** | **1** |
| $s_{\mathrm{A}}(\sigma)$ sorted cumulative | **0.543** | **0.891** | **0.978** | **1** |

As already mentioned, an important property of any (weakly) amplifying function is that a repeated application of it will result in one word to remain. Furthermore, we stated that (under some extra conditions) the amplifying property of $s$ is lifted to the global transition function $t$ as defined in (4). This implies that the amplification of $s$ is a sufficient condition for the convergence of the sampling-response model. One may however wonder whether it is also a necessary condition. After all, the amplification of $t$ is not a necessary condition for convergence, weakly amplification by itself is already sufficient.

While we do not have an answer to this question in general, we show by a counterexample that symmetry, order preservation and weakly amplification of the response function $s$ do not guarantee the weakly amplification, hence convergence, of $t$. For this we define a function $s_{\mathrm{W}}$ which increases the highest frequency and makes all other frequencies equal. [7] More precisely, let $\sigma^+$ be the maximal element in $\sigma$ and $\kappa(\sigma)$ be the number of elements in $\sigma$ which are maximal. Further let $\delta(\sigma)$ be the sum of the elements of $\sigma$ which are not maximal and $\alpha \in ]0,1[$ a constant, then we have

$$[s_{\mathrm{W}}(\sigma)]_i = \begin{cases} \sigma^+ + \frac{\alpha\delta(\sigma)}{\kappa(\sigma)} & \text{if } \sigma_i = \sigma^+ \\ \frac{(1-\alpha)\delta(\sigma)}{n-\kappa(\sigma)} & \text{otherwise} \end{cases} \tag{6}$$

This function is not amplifying because we have for example, with $\alpha = 0.1$, $s_{\mathrm{W}}((0.6, 0.4, 0)) = (0.64, 0.18, 0.18)$ whereby $0.64+0.18 < 0.6+0.4$. While $s_{\mathrm{W}}$ is weakly amplifying it is shown in section 4 that the resulting transition function $t_{\mathrm{W}}$ does not inherit this property.

Finally we also define the identity response function $s_{\mathrm{I}}$:

$$s_{\mathrm{I}}(\sigma) = \sigma \tag{7}$$

which corresponds to $s_{\mathrm{A}}$ with $m = 1$. For this function we can derive

$$t_{\mathrm{I}}(\tau) = \sum_{\sigma \in S} w_\tau(\sigma) s_{\mathrm{I}}(\sigma) = \sum_{\sigma \in S} w_\tau(\sigma)\sigma = \tau, \tag{8}$$

using (3). The global transition hence also becomes the identity function. In section 4 this response function will help us to clarify the relation between the turn-based model and the sampling-response model.

### *3.4 Properties*

#### *3.4.1 Finite sampling*

One way in which the process of agents playing naming games can be looked at, is that every agent tries to guess the population word frequencies $\tau$. This is done by taking a certain, limited number of samples. Based on the agent's guess of $\tau$, he then decides how to speak himself and so tries to come to an agreement with the other agents. Of course, if an agent can listen unlimitedly often to randomly chosen speakers, there is no limit on the accuracy with which $\tau$ can be estimated. Suppose that all the agents do this and then

[7] In fact such a function is not order preserving according to the definition given in appendix A, because frequencies which are different can become equal.

change their behavior such that they only use the word that occurred most frequently, then all agents will use the same word and the synonymy will vanish. Yet, this is not very realistic, as it requires an infinite number of games before coherence is reached. Therefore in the model presented, we dropped this strong assumption as it does not really reflect dynamic behavior in the real world. More precisely, we assumed that every agent decides on which word to use (or with which probabilities to use each word) only based on a limited number of interactions with other agents. This implies that the observations the agents make, vary and that for example the word observed most frequently can differ from agent to agent. This is an important difference with other mathematical models trying to capture the language dynamics of a population of agents ,e.g. Tuyls et al. (2003); Cucker et al. (2004). More precisely, Cucker et al. (2004) also models language evolution but the proof of convergence toward the same language relies on letting the number of samples go to infinity. [8]

#### *3.4.2 Deterministic transition function*

The transition function $t$ gives the expected new word frequencies, $\tau'$, given the current word frequencies $\tau$. This is an idealization with respect to the turn-based model in the sense that it assumes that every possible observation in $S$ will occur exactly with its predicted probability. As there are only a finite number of agents, the observed frequencies will instead fluctuate around their mean. Therefore, a trajectory of $\tau$ in the turn-based model will be stochastic instead of deterministic. Still, in Section 4 it is verified that this more realistic, stochastic system behaves qualitatively the same as its idealized counterpart, the sampling-response model, at least for amplifying response functions. The analysis of language dynamics in finite populations has also been studied in Komarova and Nowak (2003). The underlying model and assumptions made there are different from ours. First, in their work language evolution is studied using vertical transmission. In particular, children learn the language of their parents with a certain error rate. Second, though the set of choices (languages) on which the agents have to agree, is finite like in our model, the state of each agent is exactly one of these choices, whereas in our model an agent may use several alternatives with a certain preference. Finally, one of the main focuses of that paper was on the stationary distribution of the number of individuals speaking each of the different languages. We only address situations where such a stationary distribution corresponds to each agent speaking the same language, but on the other hand we focus more on the way in which a population evolves toward such a state.

#### *3.4.3 Horizontal versus vertical transmission*

The sampling-amplification model we proposed, is based on the naming game and meant to explain its dynamics. The main modification we made, in order to obtain an mathematically analyzable model, was the synchronous interaction of the agents which is further discussed in Section 4. While the naming game assumes horizontal transmission, the model with synchronous updates for the agents also resembles the vertical transmission model.

[8] At the same time, that model also assumes a more abstract, continuous space (of languages) on which the agents have to reach coherence, whereas in our model the set of choices is discrete and finite. This probably makes the assumption of unlimited sampling inevitable.

That is, one can interpret the transition function $t$ as the mapping of the language from one generation to the next. This interpretation is valid as long as agents in the new generation (children) learn their language from *randomly chosen* agents in the current generation (parents), thus irrespective of their fitness (e.g. success in playing language games).

## 4 Discussion

In the previous section we introduced a mathematical, deterministic model of the evolution of word frequencies in a population of agents. This model resulted in a transition function $t$ as defined in (4). Moreover, we stated and proved (in appendix B) that under a certain condition on the response function, i.e. amplification, and some other restrictions on the sampling set $S$ and sampling function $w$, this transition function induces a dynamic where only one word remains. We will now check whether this sampling-response model predicts the behavior of the original, stochastic turn-based model. Therefore we compare the evolution of word frequencies under the transition function $t$ (as defined in equation (4)) in the sampling-response model with the evolution of the average agents' speaking behavior in the turn-based model. This comparison is conducted in three settings, which differ in the agents' response function. In both the sampling-response and turn-based model we use queue-agents with $K = 3$. In the turn-based model the population consists of $A = 200$ agents. The three types of response functions used are $s_A$, $s_I$ and $s_W$ as defined in (5), (7) and (6) respectively. To make a visualization of the corresponding space $\Sigma_n$ possible, the number of words $n$ was restricted to three. Therefore, in the turn-based model, each slot in the agents' queues was initially filled at random with one three words. The results are shown in Figure 3.

The first response function used was $s_A$ with $m = 2$. In Figure 3(a), the transition function $t_A$ was iterated starting from different initial states near the central point $\tau_c = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. In Figure 3(b) the evolution of the average speaking behavior in the turn-based model is shown, for 50 runs. The average speaking behavior of the population was plotted every $KA$ games. This implies that the number of agent interactions between two connected points is the same in both model. The parameters are such that the conditions for amplification of $t_A$ are fulfilled: $s_A$ is amplifying as $m > 1$ and the sampling function $w$ is consistent because $K \geq 3$. Indeed, in Figure 3(a) one can see clearly the convergence toward one of the corners of the simplex (arrows were omitted for clarity). Figure 3(b) suggests that the turn-based model, however stochastic, shows the same tendency to converge toward one of the stable fixed points of the dynamic.

The second response function investigated was $s_I$, defined as $s_I(\sigma) = \sigma$, which in its turn implied $t_I(\tau) = \tau$. The corresponding speaking behavior of an agent can be understood as if he randomly selects a word from its queue. One may wonder whether this neutral dynamic of $t_I$ predicts the behavior of the turn-based system. Figures 3(c) and 3(d) show a comparison between the sampling-response system and the turn-based system in this case. In Figure 3(c), the absence of a dynamic is illustrated by the dots: in fact every point on the simplex is a neutral fixed point. With regard to the turn-based model in Figure 3(d), this neutrality translates apparently into a random walk on the simplex, which is

| Response Function | Sampling-response Model | Turn-based Model |
| --- | --- | --- |
| $s_{\mathrm{A}}$ | (a) | (b) |
| $s_{\mathrm{I}}$ | (c) | (d) |
| $s_{\mathrm{W}}$ | (e) | (f) |

Fig. 3. Comparison of the sampling-response model with the turn-based model for three different response functions: $s_{\mathrm{A}}$, $s_{\mathrm{I}}$ and $s_{\mathrm{W}}$. In all cases we use queue-agents with $K = 3$, three words and the plots show the evolution over $\Sigma_3$ of the word frequencies. In (b),(d) and (f) the population consists of $A = 200$ agents, and points are drawn every $KA = 600$ games.

shown by just one run in order not to clutter the image. This means that in this case the population will eventually still end up in a fixed point, as words will accidentally get lost from time to time until only one word remains.[9] . Thus, from the stationarity of the sampling-response model, we cannot conclude that the turn-based model will also not converge to the use of a single word. What the neutral dynamic in Figure 3(c) however does suggest, is that the time needed to reach a consensus will be relatively large in this case. This is verified further on.

The last response function we consider is $s_{\mathrm{W}}$. Figures 3(e) and 3(f) show a comparison between the sampling-response system and the turn-based system in this case. In Figure 3(e) one can see that despite $s_{\mathrm{W}}$ being order-preserving and weakly amplifying, $t_{\mathrm{W}}$ is not weakly amplifying (which would imply all trajectories to converge to one of the corners). On the contrary, the central point $\tau_c$ is a stable fixed point and an attractor for a large domain of $\Sigma_3$. In the turn-based model this means that trajectories will randomly wander around the central point as is shown in 3(f). Also in this case, the trajectory will eventually end up in a corner, for the same reason as before. We expect however that the time to reach such a state will be larger than in the case of $s_{\mathrm{I}}$ and much larger than in the case

[9] Geometrically, on $\Sigma_3$, losing a word means hitting an edge (from three to two words) or a corner (from two to one word) of the simplex.

Fig. 4. The average number of games needed to reach convergence as a function of the population size in the turn-based model. In the three graphs we have a population of queue-agents with $K = 3$ with different types of response functions: $s_{\mathrm{A}}$, $s_{\mathrm{I}}$ and $s_{\mathrm{W}}$. The plots show an average over 100 runs and the error bars show a 95% confidence interval.

of $s_{\mathrm{A}}$.

We verified the expected difference in convergence time between the response functions $s_{\mathrm{A}}$, $s_{\mathrm{I}}$ and $s_{\mathrm{W}}$ in the turn-based model. In Figure 4, the average number of games needed to reach consensus is shown as a function of the population size for the three cases. There is clearly a large difference in time to convergence between the three cases, with an increasing population size. The most important difference (in relative terms) exists between the amplifying response function $s_{\mathrm{A}}$ and the non-amplifying functions $s_{\mathrm{I}}$ and $s_{\mathrm{W}}$. Amplification hence dramatically increases the speed of convergence. Still, a considerable difference exists between $s_{\mathrm{I}}$ and $s_{\mathrm{W}}$, which is explained by their associated transition functions. While $t_{\mathrm{I}}$ allows a pure random walk on $\Sigma_3$, $t_{\mathrm{W}}$ superposes an attracting force toward $\tau_c$ on this random walk, which increases the time to escape from the central area of $\Sigma_3$.

## 5 Conclusion

The sampling-response model together with the central concept of amplification is a mathematical model specifically tailored to study and explain the language dynamics of a population of agents playing naming games. The main difference with existing mathematical work on learning processes in general and language dynamics in particular is the abandoning of infinite sampling in order to proof convergence, i.e. a population of agents reaching consensus on the use of a common language.

We experimentally validated our approach, by providing an amplifying function and a consistent sampling function. Moreover we compared the mathematical model, using the defined functions, with simulations under realistic conditions. Our experiments have shown a clear similar behavior with the conducted simulations, which implies a good predictive behavior of our sampling-amplification model. This can be of great significance in setting up a multi-agent system in which the agents first have to acquire a common language. In this manner good parameter settings, which fulfill the amplifying conditions and thus guarantee convergence to a common language, can be predicted beforehand by using this model.


## Acknowledgments

We thank Joachim de Beule for checking the math and for helpful discussions. Bart De Vylder is sponsored by the Fund for Scientific Research - Flanders (Belgium). Karl Tuyls is sponsored by a grant of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, nr: BSIK03024 (The Netherlands).


## A Definitions and Properties

In this section we provide the necessary definitions and propositions together with their proof to understand the proof of convergence of Section B.

The $i^{\text{th}}$ component of a vector $x$ is noted as $x_i$. In a compound expression $v(x)$ both $v_i(x)$ and $[v(x)]_i$ will be used.

The set of all permutation operators (or shortly permutations in what follows) on $n$ elements is written as $\mathbb{P}$. For a permutation $p \in \mathbb{P}$ and $x$ a vector of $n$ elements, $p(x)$ is the result of $p$ acting on $x$. $p_i(x)$ has the usual interpretation $[p(x)]_i$, yet $p_i$ itself will be used to give the index of the element that moves to position $i$, or $p((x_1, x_2, \ldots, x_n)) = (x_{p_1}, x_{p_2}, \ldots, x_{p_n})$.

The permutation that swaps two elements on position $i$ and $j$ is written as ${}_{i \leftrightarrow j}$.

**Definition** A map $v : \Sigma \to \Sigma$ is **symmetrical** iff

$$v(p(\sigma)) = p(v(\sigma)) \qquad \forall p \in \mathbb{P}, \forall \sigma \in \Sigma.$$

**Definition** An element $x \in \Sigma$ is **decreasing** iff $x_1 \geq x_2 \geq \ldots \geq x_n$ and is **strictly decreasing** if the inequalities are strict.

The subset of $\Sigma$ of decreasing elements is written as $\Sigma'$. We define $n$ elements of $\Sigma'$, $u^{(i)}$, $1 \leq i \leq n$, as follows:

$$u_j^{(i)} = \begin{cases} \frac{1}{i} \text{ for } j \leq i \\ 0 \text{ elsewhere} \end{cases} \tag{A.1}$$

and the set of these points $U = \{u^{(1)}, u^{(2)}, \ldots, u^{(n)}\}$. The set $\Sigma^* \subset \Sigma'$ is defined as $\Sigma^* = \Sigma' \setminus U$. An alternative characterization of $\Sigma^*$ is

$$\sigma \in \Sigma^* \Leftrightarrow \begin{cases} \sigma \text{ is decreasing} \qquad \text{and} \\ \exists m,\ m < n \text{ and } \sigma_m > \sigma_{m+1} > 0 \end{cases} \tag{A.2}$$

**Definition** A map $v : \Sigma \to \Sigma$ is **order preserving** iff

$$\sigma_i < \sigma_j \ \Rightarrow\ v_i(\sigma) < v_j(\sigma) \quad \forall \sigma \in \Sigma\ \forall i, j \tag{A.3}$$

It follows immediately that for an order preserving map $v$

$$\sigma_i = \sigma_j \quad \Leftrightarrow \quad v_i(\sigma) = v_j(\sigma). \tag{A.4}$$

**Definition** The function $c : \Sigma \to \mathbb{R}^n$ transforms an element $\sigma$ in its cumulative distribution vector $c(\sigma)$ defined as:

$$c_k(\sigma) = \sum_{i=1}^{k} \sigma_i.$$

Clearly, $c_n(\sigma) = 1$.

**Definition** A symmetrical, order preserving map $v : \Sigma \to \Sigma$ is **amplifying** iff

$$c_k(v(\sigma)) \geq c_k(\sigma) \quad \forall \sigma \in \Sigma',\ \forall k \tag{A.5}$$

with strict inequality if $\sigma \in \Sigma^*$, $k < n$ and $\sigma_{k+1} > 0$. The condition for $\sigma$ outside $\Sigma'$ follows from the symmetry of $v$.

It can be easily shown that for such a map must hold

$$v(\sigma) = \sigma \qquad \forall \sigma \in U. \tag{A.6}$$

Next, we define a partition of $\Sigma'$ in $n$ subsets $B_i$, $1 \leq i \leq n$, such that for $\sigma \in \Sigma'$:

$$\sigma \in B_i \Leftrightarrow \sigma_1 = \sigma_2 = \ldots = \sigma_i \quad \text{and} \tag{A.7}$$

$$\sigma_i > \sigma_{i+1}, \qquad \text{if } i < n. \tag{A.8}$$

**Proposition 1.** If $v : \Sigma \to \Sigma$ is an order preserving map, then the sets $B_i$, $1 \leq i \leq n$ are invariant under $v$.

*Proof.* The equalities and inequality in (A.7) and (A.8) are preserved because of (A.3) and (A.4). □

**Proposition 2.** Let $v : \Sigma \to \Sigma$ be a symmetrical, order preserving, amplifying map. If we consider the difference equation:

$$\sigma^{(j+1)} = v(\sigma^{(j)}), \quad \text{with } \sigma^{(0)} \in B_i$$

then

$$\lim_{j\to\infty} \sigma^{(j)} = u^{(i)}.$$

However, only $u^{(1)}$ is an asymptotically stable fixed point.

*Proof.* From proposition 1 we have $\sigma^{(j)} \in B_i$ for all $j$. Also $u^{(i)} \in B_i$. From (A.6) follows that $u_i$ is a fixed point and from (A.5) we know there can be no other fixed points. We now define the following function $V : B_i \to \mathbb{R}_{\geq 0}$

$$V(\sigma) = \frac{1}{i} - \sigma_1.$$

We have

(i) $V(u^{(i)}) = 0$, because $V(u^{(i)}) = \frac{1}{i} - u_1^{(i)} = 0$.
(ii) $V(\sigma) > 0$ for all $\sigma \in B_i \setminus \{u^{(i)}\}$. Because $\sum_{k=1}^{i} \sigma_k = 1$ would imply that $\sigma = u^{(i)}$, we deduce that $\sum_{k=1}^{i} \sigma_k < 1$. Further, as $\sum_{k=1}^{i} \sigma_k = i\sigma_1$ we obtain $\sigma_1 < \frac{1}{i}$.
(iii) $\Delta V(\sigma) = V(v(\sigma)) - V(\sigma) < 0$ in $\sigma \in B_i \setminus \{u^{(i)}\}$. We have $V(v(\sigma)) - V(\sigma) = \sigma_1 - v_1(\sigma) = c_1(\sigma) - c_1(v(\sigma)) < 0$, because $v$ is amplifying and the conditions for strict inequality in (A.5): $\sigma \in \Sigma^*$ and $\sigma_2 > 0$ are met ($\sigma_2 = 0$ would imply $\sigma = u_1$).

Hence, $V$ is a Lyapunov function on $B_i$ and therefore the fixed point $u^{(i)}$ has basin of attraction $B_i$. However, as an arbitrary small neighborhood of $u^{(i)}$, $i \geq 2$ contains elements outside of $B_i$, e.g.

$$(\frac{1}{i} + \epsilon, \frac{1}{i} - \frac{\epsilon}{i-1}, \ldots, \frac{1}{i} - \frac{\epsilon}{i-1}, 0, \ldots, 0) \in B_1, \quad \text{for } \epsilon > 0, \tag{A.9}$$

these $u^{(i)}$ are unstable fixed points. Therefore only $u^{(1)}$ is asymptotically stable. □

**Definition** A finite set $S \subset \Sigma$ is **symmetrical** iff

$$\sigma \in S \Leftrightarrow p(\sigma) \in S \qquad \forall p \in \mathbb{P}. \tag{A.10}$$

**Definition** A **sampling function** $w$ over a finite symmetrical set S is a function with signature $\Sigma \to (S \to \mathbb{R}_{\geq 0})$, such that for all $\tau \in \Sigma$, $w_\tau$ is a probability distribution over S with expected value $\tau$, or

$$\sum_{\sigma \in S} w_\tau(\sigma)\sigma = \tau. \tag{A.11}$$

In the following, we assume $S$ to be a finite symmetrical set.

**Definition** The binary relation $\succcurlyeq$ on $\Sigma$ is defined as $x \succcurlyeq y$ if $y_i$ is zero whenever $x_i$ is zero, or formally

$$x \succcurlyeq y \Leftrightarrow \forall i \; x_i = 0 \;\; \Rightarrow \;\; y_i = 0. \tag{A.12}$$

In addition we also define the relation $\doteqdot$ as

$$\begin{aligned} x \doteqdot y &\Leftrightarrow x \succcurlyeq y \wedge y \succcurlyeq x && \text{(A.13)} \\ &\Leftrightarrow \forall i \; x_i = 0 \;\; \Leftrightarrow \;\; y_i = 0. && \text{(A.14)} \end{aligned}$$

Obviously $\succcurlyeq$ is transitive and

$$x \succcurlyeq y \Rightarrow p(x) \succcurlyeq p(y) \quad \forall p \in \mathbb{P} \tag{A.15}$$

**Proposition 3.** With $w$ a sampling function over $S$

$$w_\tau(\sigma) > 0 \Rightarrow \tau \succcurlyeq \sigma \tag{A.16}$$

for all $\tau \in \Sigma$ and $\sigma \in S$.

*Proof (by contrapositive).* Let $\tau_i = 0$ and $\sigma_i > 0$. As $w$ is a sampling function we have

$$\tau_i = \sum_{x \in S} w_\tau(x)\, x_i \;\; = \sum_{x \in S \setminus \{\sigma\}} w_\tau(x)\, x_i + w_\tau(\sigma)\, \sigma_i \;\; = \;\; 0 \tag{A.17}$$

which can only be true if $w_\tau(\sigma) = 0$. □

**Definition** A sampling function $w$ over $S$ is **supportive** if also the converse of proposition 3 holds:

$$\tau \succcurlyeq \sigma \Rightarrow w_\tau(\sigma) > 0 \tag{A.18}$$

for all $\tau \in \Sigma$ and $\sigma \in S$.

**Proposition 4.** Let $w$ be a supportive sampling function over $S$, $\tau \in \Sigma$ and $p \in \mathbb{P}$ with $\tau \doteqdot p(\tau)$, then for all $\sigma \in S$

$$w_\tau(\sigma) > 0 \Rightarrow w_\tau(p(\sigma)) > 0 \tag{A.19}$$

*Proof.* From $\tau \doteq p(\tau)$ follows $p^{-1}(\tau) \doteq p^{-1}(p(\tau))$ and thus $\tau \doteq p^{-1}(\tau)$ (using (A.15)). If we assume $w_\tau(\sigma) > 0$ then proposition 3 implies $\tau \succcurlyeq \sigma$. As a result $p^{-1}(\tau) \succcurlyeq \sigma$ (using transitivity) and hence $\tau \succcurlyeq p(\sigma)$. Finally, this implies $w_\tau(p(\sigma)) > 0$ as $w$ is supportive. □

**Definition** A sampling function $w$ over S is **symmetrical** iff

$$w_\tau(\sigma) = w_{p(\tau)}(p(\sigma)) \qquad \forall\tau \in \Sigma,\ \forall\sigma \in S,\ \forall p \in \mathbb{P}. \tag{A.20}$$

In the following, we will use the following notations: $S' = S \cap \Sigma'$ and $S^* = S \cap \Sigma^*$.

**Definition** A symmetrical sampling function $w$ over $S$ is **covering** iff

$$\forall\tau \in \Sigma^*\ \exists\sigma \in S^*\ w_\tau(\sigma) > 0. \tag{A.21}$$

The requirements for $\tau$ outside $\Sigma^*$ follow from the symmetry of $w$.

**Proposition 5.** Let $w$ be a symmetrical, supportive sampling function over $S$. A necessary and sufficient condition for $w$ to be covering is that there exists an element $\sigma^* \in S^*$ such that $\sigma^*_i = 0$ for all $i > 2$.

*Proof.* In order to prove sufficiency let $\tau \in \Sigma^*$. As $\tau_1 > 0$ and $\tau_2 > 0$ ($\tau_2 = 0$ would imply $\tau = u^{(1)}$) and $w$ is supportive we have $w_\tau(\sigma^*) > 0$. Thus $w$ is covering. Conversely, if $w$ is covering, choose $\tau \in \Sigma^*$ with $\tau_i = 0$ for all $i > 2$. Let $\sigma'$ be an element of $S^*$ for which $w_\tau(\sigma') > 0$. From proposition 3 we infer $\sigma'_i = 0$ for all $i > 2$ and hence we can choose $\sigma^* = \sigma'$. □

**Definition** A symmetrical sampling function $w$ over $S$ is **consistent** iff

$$\tau_i \geq \tau_j\ \wedge\ \sigma_i \geq \sigma_j \quad\Rightarrow\quad w_\tau(\sigma) \geq w_\tau({}_{i\leftrightarrow j}(\sigma)) \tag{A.22}$$

for all $i, j$ with strict inequality if $\tau_i > \tau_j$, $\sigma_i > \sigma_j$ and $w_\tau(\sigma) > 0$.

Clearly, if one of the conjuncts in the left hand side of (A.22) is an equality then also the right hand side is: If $\sigma_i = \sigma_j$ then $\sigma = {}_{i\leftrightarrow j}(\sigma)$ and $w_\tau(\sigma) = w_\tau({}_{i\leftrightarrow j}(\sigma))$. Likewise, if $\tau_i = \tau_j$ then $\tau = {}_{i\leftrightarrow j}(\tau)$, $w_\tau(\sigma) = w_{i\leftrightarrow j(\tau)}(\sigma)$ and using the symmetry of $w$, $w_\tau(\sigma) = w_\tau({}_{i\leftrightarrow j}(\sigma))$. As a consequence, the requirement for consistency can be restated as

$$\tau_i > \tau_j\ \wedge\ \sigma_i > \sigma_j \quad\Rightarrow\quad w_\tau(\sigma) \geq w_\tau({}_{i\leftrightarrow j}(\sigma)) \tag{A.23}$$

for all $i, j$ with strict inequality if $w_\tau(\sigma) > 0$. Moreover, if $w$ is supportive we have the following

**Proposition 6.** A symmetrical, supportive sampling function $w$ over $S$ is consistent iff

$$\tau_i > \tau_j \ \wedge \ \sigma_i > \sigma_j \ \wedge \ \ w_\tau(\sigma) > 0 \quad \Rightarrow \quad w_\tau(\sigma) > w_\tau({}_{i\leftrightarrow j}(\sigma)) \tag{A.24}$$

*Proof.* The necessity of (A.24) follows immediately from (A.22) or (A.23). In order to prove the condition to be sufficient for consistence we have to show that whenever $\tau_i > \tau_j$, $\sigma_i > \sigma_j$ and $w_\tau(\sigma) = 0$, also $w_\tau({}_{i\leftrightarrow j}(\sigma)) = 0$. Now, because $w$ is supportive, $w_\tau(\sigma) = 0$ implies that there exists a $k$ such that $\tau_k = 0$ and $\sigma_k > 0$. Clearly $k \neq i$ as $\tau_i > 0$. If moreover $k \neq j$ then we have $[{}_{i\leftrightarrow j}(\sigma)]_k = \sigma_k > 0$. If, on the other hand $k = j$ then $[{}_{i\leftrightarrow j}(\sigma)]_k = \sigma_i > 0$, such that $[{}_{i\leftrightarrow j}(\sigma)]_k > 0$ in all cases. By proposition 3 we may then conclude $w_\tau({}_{i\leftrightarrow j}(\sigma)) = 0$. □

## B Main Result: Proof of Convergence

Before introducing and proving our main theorem we need to introduce two lemma's.

**Lemma 1.** For any $a, b \in \mathbb{R}^n$ with $\sum_{i=1}^n b_i = 0$,

if, for all $m$, $1 \leq m < n$,

$$a_m \geq a_{m+1} \qquad \text{and} \tag{B.1}$$

$$\sum_{i=1}^{m} b_i \geq 0 \tag{B.2}$$

then

$$\sum_{j=1}^{n} a_j b_j \geq 0. \tag{B.3}$$

If the inequalities (B.1) and (B.2) are simultaneously strict for at least one $m$, then (B.3) is also strict.

*Proof.* The stated follows immediately from the following identity:

$$\sum_{j=1}^{n} a_j b_j = \sum_{m=1}^{n-1} \Big( (a_m - a_{m+1}) \sum_{i=1}^{m} b_i \Big).$$

□

**Lemma 2.** For any $k$, $1 \leq k < n$, and $\mu, \nu \in \{1..n\}$ with $\mu \neq \nu$ and $e$ any function with domain $\mathbb{P}$, the following identity holds:

$$\sum_{i=1}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu}} e(p) - \sum_{i=1}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\nu}} e(p) = \sum_{i=1}^{k}\sum_{j=k+1}^{n}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} \Big(e(p) - e({}_{i\leftrightarrow j}\circ p)\Big) \tag{B.4}$$

*Proof.* Rewriting the first term in the left hand side of (B.4) we get

$$\begin{aligned}
\sum_{i=1}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu}} e(p) &= \sum_{i=1}^{k}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p) && \text{(B.5)}\\
&= \sum_{i=1}^{k}\Bigg(\sum_{\substack{j=1\\j\neq i}}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p) + \sum_{j=k+1}^{n}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p)\Bigg) && \text{(B.6)}\\
&= \sum_{\substack{i,j=1\\j\neq i}}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p) + \sum_{i=1}^{k}\sum_{j=k+1}^{n}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p), && \text{(B.7)}
\end{aligned}$$

in which the first term in (B.7) is symmetrical in $\mu$ and $\nu$. Likewise, we can derive the same expression, expect for $\mu$ and $\nu$ interchanged, for the second term of the left hand side of (B.4). Therefore, these first terms cancel each other out and we obtain

$$\begin{aligned}
\sum_{i=1}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu}} e(p) - \sum_{i=1}^{k}\sum_{\substack{p\in\mathbb{P}\\p_i=\nu}} e(p) &= \sum_{i=1}^{k}\sum_{j=k+1}^{n}\Bigg(\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} e(p) - \sum_{\substack{p\in\mathbb{P}\\p_i=\nu\\p_j=\mu}} e(p)\Bigg) && \text{(B.8)}\\
&= \sum_{i=1}^{k}\sum_{j=k+1}^{n}\sum_{\substack{p\in\mathbb{P}\\p_i=\mu\\p_j=\nu}} \Big(e(p) - e({}_{i\leftrightarrow j}\circ p)\Big) && \text{(B.9)}
\end{aligned}$$

□

**Theorem 1** (Main Result)**.** Given a map $s : \Sigma \to \Sigma$ which is symmetrical, order preserving and amplifying and given a consistent, symmetrical, supportive, covering sampling function $w$ over a symmetrical set $S$ then the map $t : \Sigma \to \Sigma$ defined by

$$t(\tau) = \sum_{\sigma\in S} w_\tau(\sigma)s(\sigma)$$

is also symmetrical, order preserving and amplifying.

*Proof of symmetry.* Given $\tau \in \Sigma$ and $p \in \mathbb{P}$, then

$$t(p(\tau)) = \sum_{\sigma\in S} w_{p(\tau)}(\sigma)\ s(\sigma) \ = \ \sum_{\sigma\in S} w_{p(\tau)}(p(\sigma))\ s(p(\sigma)) \ = \ \sum_{\sigma\in S} w_\tau(\sigma)\ p(s(\sigma)) \tag{B.10}$$

$$= p\Big(\sum_{\sigma\in S} w_\tau(\sigma)\ s(\sigma)\Big) \ = \ p(t(\tau)), \tag{B.11}$$

using a substitution $\sigma \to p(\sigma)$, the symmetry of $w$ and $s$ and the linearity of $p$. □

*Proof of order preservation.* Given $\tau \in \Sigma$ with $\tau_i > \tau_j$, we have to show that $t_i(\tau) > t_j(\tau)$. We have

$$t_i(\tau) - t_j(\tau) = \sum_{\sigma\in S} w_\tau(\sigma)(s_i(\sigma) - s_j(\sigma)) \tag{B.12}$$

$$= \sum_{\substack{\sigma\in S\\ \sigma_i>\sigma_j}} \left(w_\tau(\sigma) - w_\tau(i{\leftrightarrow}j(\sigma))\right)(s_i(\sigma) - s_j(\sigma)) \tag{B.13}$$

The ignored terms with $\sigma_i = \sigma_j$ do not alter this sum because then $s(\sigma)_i = s(\sigma)_j$ as $s$ is order preserving. To show that this sum is strictly positive we turn to

$$\tau_i - \tau_j = \sum_{\sigma\in S} w_\tau(\sigma)(\sigma_i - \sigma_j) \ > \ 0 \tag{B.14}$$

which implies that there exists at least one element of $S$, say $\sigma^*$, for which $\sigma^*_i - \sigma^*_j > 0$ and $w_\tau(\sigma^*) > 0$. This implies that $s_i(\sigma^*) - s_j(\sigma^*) > 0$, because $s$ is order preserving, and $w_\tau(\sigma^*) - w_\tau(i{\leftrightarrow}j(\sigma^*)) > 0$, because $w$ is consistent. Returning to (B.13) we may conclude that at least one term is strictly positive. □

*Proof of amplification.* Let $\tau \in \Sigma'$. We now have to prove that

$$\sum_{i=1}^{k}(t_i(\tau) - \tau_i) \geq 0, \tag{B.15}$$

with strict inequality if $\tau \in \Sigma^*$, $k < n$ and $\tau_{k+1} > 0$. We have

$$t_i(\tau) - \tau_i = \sum_{\sigma\in S} w_\tau(\sigma)s_i(\sigma) - \sum_{\sigma\in S} w_\tau(\sigma)\sigma_i \tag{B.16}$$

$$= \sum_{\sigma\in S} w_\tau(\sigma)(s_i(\sigma) - \sigma_i) \tag{B.17}$$

$$= \sum_{\sigma\in S} w_\tau(\sigma)d_i(\sigma) \tag{B.18}$$

with an auxiliary function $d(\sigma) = s(\sigma) - \sigma$. Note that $d$ is a symmetrical function and $\sum_{i=1}^{n} d_i(\sigma) = 0$. We now divide the summation domain $S$ into $n!$ summations over $S'$, using permutations to cover the original set:

$$\sum_{\sigma\in S} w_\tau(\sigma)d_i(\sigma) = \sum_{\sigma\in S'} \frac{1}{\rho(\sigma)} \sum_{p\in\mathbb{P}} w_\tau(p(\sigma))d_i(p(\sigma)) \tag{B.19}$$

with $\rho(\sigma) = \#\{p \in \mathbb{P} \mid p(\sigma) = \sigma\}$.

Rewriting the second sum in the right hand side of (B.19) using $d_i(p(\sigma)) = p_i(d(\sigma)) = d_{p_i}(\sigma)$ and sub-dividing the permutations we obtain

$$\sum_{p\in\mathbb{P}} w_\tau(p(\sigma))d_i(p(\sigma)) = \sum_{j=1}^{n} \sum_{\substack{p\in\mathbb{P}\\ p_i=j}} w_\tau(p(\sigma))d_{p_i}(\sigma) \tag{B.20}$$

$$= \sum_{j=1}^{n} \sum_{\substack{p\in\mathbb{P}\\ p_i=j}} w_\tau(p(\sigma))d_j(\sigma) \tag{B.21}$$

$$= \sum_{j=1}^{n} \Big[ \sum_{\substack{p\in\mathbb{P}\\ p_i=j}} w_\tau(p(\sigma))\Big] d_j(\sigma) \tag{B.22}$$

If we now return to (B.15) and apply (B.19) and (B.22) we get

$$\sum_{i=1}^{k}(t_i(\tau) - \tau_i) = \sum_{\sigma\in S'} \frac{1}{\rho(\sigma)} \sum_{j=1}^{n} \Big[ \sum_{i=1}^{k} \sum_{\substack{p\in\mathbb{P}\\ p_i=j}} w_\tau(p(\sigma))\Big] d_j(\sigma). \tag{B.23}$$

Now we use lemma 1 to prove that the right hand side of (B.23) non-negative in general and strictly positive under stronger assumptions on $\tau$. We identify $a_j$ with $\sum_{i=1}^{k} \sum_{\substack{p\in\mathbb{P}\\ p_i=j}} w_\tau(p(\sigma))$ and $b_j$ with $d_j(\sigma)$.

First we prove that (B.23) is always nonnegative. Condition (B.2) becomes

$$\sum_{l=1}^{m} d_l(\sigma) \ = \sum_{l=1}^{m}(s_l(\sigma) - \sigma_l) \ \geq \ 0, \tag{B.24}$$

which holds as a direct consequence of $s$ being amplifying and $\sigma \in S'$.

Regarding condition (B.1) and using lemma 2 with $e(p) = w_\tau(p(\sigma))$, $\mu = m$ and $\nu = m+1$, we have

$$a_m - a_{m+1} = \sum_{i=1}^{k} \sum_{\substack{p\in\mathbb{P}\\ p_i=m}} w_\tau(p(\sigma)) - \sum_{i=1}^{k} \sum_{\substack{p\in\mathbb{P}\\ p_i=m+1}} w_\tau(p(\sigma)) \tag{B.25}$$

$$= \sum_{i=1}^{k} \sum_{j=k+1}^{n} \sum_{\substack{p\in\mathbb{P}_n\\ p_i=m\\ p_j=m+1}} \Big( w_\tau(p(\sigma)) - w_\tau({}_{i\leftrightarrow j}(p(\sigma))) \Big). \tag{B.26}$$

Now, as $\tau \in \Sigma'$ and $i < j$ we have $\tau_i \geq \tau_j$. Also, $p_i(\sigma) \geq p_j(\sigma)$ as $p_i(\sigma) = \sigma_{p_i} = \sigma_m$, $p_j(\sigma) = \sigma_{m+1}$ and $\sigma \in S'$. Therefore, with $w$ consistent, we infer $w_\tau(p(\sigma)) \geq w_\tau({}_{i\leftrightarrow j}(p(\sigma)))$. Hence, (B.26) is always nonnegative.

In order to show (B.23) to be strictly positive if $\tau \in \Sigma^*$, $k < n$ and $\tau_{k+1} > 0$, there must exist at least one $\sigma$ and $m$ for which the inequality in (B.24) is strict and (B.26) is strictly positive. The former requirement is fulfilled if

$$\sigma \in S^* \tag{B.27}$$

$$\sigma_{m+1} > 0 \tag{B.28}$$

because $s$ is amplifying. The latter is fulfilled if

$$w_\tau(p(\sigma)) > 0 \tag{B.29}$$

$$\tau_i > \tau_j \tag{B.30}$$

$$\sigma_m > \sigma_{m+1}, \tag{B.31}$$

for at least one term in $i$, $j$ and $p$ in (B.26), because $w$ is consistent, $\sigma_m > \sigma_{m+1}$ is equivalent to $p_i(\sigma) > p_j(\sigma)$ and regarding proposition 6.

To show this, choose $\sigma \in S^*$ with $w_\tau(\sigma) > 0$. Such a $\sigma$ exists because $w$ is covering. Next, choose $m$ such that $\sigma_m > \sigma_{m+1} > 0$, which is possible regarding (A.2). With these choices, (B.27), (B.28) and (B.31) hold.

As for condition (B.30), let $j$ be the greatest index for which $\tau_j > 0$. Clearly $j \geq k+1$. Next, choose $i \leq k$ such that (B.30) holds, which is possible as $\tau \notin U$.

Regarding condition (B.29), we first show that $m+1 \leq j$. As $w_\tau(\sigma) > 0$ we have $\tau \succcurlyeq \sigma$. Considering $\sigma_{m+1} > 0$ this means that $\tau_{m+1} > 0$ and by definition of $j$ that $m+1 \leq j$. Next, we choose $p \in \mathbb{P}$, with $p_i = m$, $p_j = m+1$ and for which $p_l \leq j \Leftrightarrow l \leq j$ or in words the permutation must not mix elements below $j$ with elements strict above $j$. The existence of such a permutation is guaranteed by $m+1 \leq j$. Then, as $\tau \in \Sigma^*$ there holds $\tau_l > 0 \Leftrightarrow l \leq j$, such that $\tau \doteqdot p(\tau)$. Finally, by proposition 4 we may conclude that $w_\tau(p(\sigma)) > 0$. □

## C An amplifying map

As an example of a symmetrical, order preserving and amplifying map, consider $s : \Sigma \to \Sigma$, with $m \in \mathbb{R}$ and $m > 1$ defined as

$$s_i(\sigma) = \frac{\sigma_i^m}{\sum_{j=1}^n \sigma_j^m}. \tag{C.1}$$

In order to establish symmetry and order preservation, we have

$$s_i(p(\sigma)) = \frac{p_i(\sigma)^m}{\sum_{j=1}^n p_j(\sigma)^m} \;=\; \frac{\sigma_{p_i}^m}{\sum_{j=1}^n \sigma_j^m} \;=\; s_{p_i}(\sigma) \;=\; p_i(s(\sigma)) \tag{C.2}$$

and

$$\sigma_i < \sigma_j \;\Leftrightarrow\; \sigma_i^m < \sigma_j^m \;\Leftrightarrow\; \frac{\sigma_i^m}{\sum_{k=1}^n \sigma_k^m} < \frac{\sigma_j^m}{\sum_{k=1}^n \sigma_k^m} \tag{C.3}$$

With regard to amplification, we have to prove that, for $\sigma \in \Sigma'$,

$$\sum_{i=1}^k s_i(\sigma) - \sum_{i=1}^k \sigma_i \;\geq\; 0, \tag{C.4}$$

with strict inequality if $\sigma \in \Sigma^*$, $k < n$ and $\sigma_{k+1} > 0$. We have

$$\sum_{i=1}^k (s_i(\sigma) - \sigma_i) = \sum_{i=1}^k \left( \frac{\sigma_i^m}{\sum_{j=1}^n \sigma_j^m} - \frac{\sigma_i}{\sum_{j=1}^n \sigma_j} \right) \tag{C.5}$$

$$= \frac{1}{\sum_{j=1}^n \sigma_j^m} \sum_{i=1}^k \sum_{j=1}^n \sigma_i^m \sigma_j - \sigma_i \sigma_j^m \tag{C.6}$$

$$= \frac{1}{\sum_{j=1}^n \sigma_j^m} \sum_{i=1}^k \sum_{j=k+1}^n \sigma_i^m \sigma_j - \sigma_i \sigma_j^m \tag{C.7}$$

$$= \frac{1}{\sum_{j=1}^n \sigma_j^m} \sum_{i=1}^k \sum_{j=k+1}^n \sigma_i \sigma_j (\sigma_i^{m-1} - \sigma_j^{m-1}) \tag{C.8}$$

having used the symmetry in $i$ and $j$ in the summand of (C.6). As $\sigma$ is decreasing and $i < j$ it follows $\sigma_i \geq \sigma_j$ and hence $\sigma_i^{m-1} \geq \sigma_j^{m-1}$, such that (C.8) is nonnegative. If moreover $\sigma \in \Sigma^*$ and $\sigma_{k+1} > 0$ then there exist $i$,$j$ with $1 \leq i \leq k < j \leq n$ such that $\sigma_i > \sigma_j > 0$ and therefore (C.8) is strictly positive.

## D A consistent sampling function

We will now show that the sampling function resulting from randomly sampling the population a fixed number of times, $K$, with $K \geq 3$, is a symmetrical, supportive, consistent, covering sampling function. Suppose an agent, at each time step, listens to $K$ randomly chosen speakers or, in other words, receives $K$ samples chosen independently from a discrete probability distribution $\tau$. If we denote with $x_i$ the number of times the agent heard the $i^{\text{th}}$ word, the frequencies observed are given by

$$\sigma = \frac{1}{K}(x_1, x_2, \ldots, x_n). \tag{D.1}$$

These finite number of elements from $\Sigma$ define a symmetrical set $S$ of size $\binom{n+K-1}{K}$, over which the sampling function will be defined.

The resulting sampling function corresponds to a multinomial distribution:

$$w_\tau(\sigma) \;=\; w_\tau(\frac{1}{K}(x_1, x_2, \ldots, x_n)) \;=\; \frac{K!}{\prod_{i=1}^n x_i!}\prod_{i=1}^n \tau_i^{x_i} \tag{D.2}$$

with the convention $0^0 = 1$. Clearly, $w$ is symmetrical and supportive, as $w_\tau(\sigma) = 0$ implies $\tau_i = 0$ and $x_i > 0$ for some $i$. To demonstrate that $w$ is consistent, we make use of proposition 6 and assume $\tau_i > \tau_j$, $x_i > x_j$ or equivalently $\sigma_i > \sigma_j$ and $w_\tau(\sigma) > 0$. We then have to show that $w_\tau(\sigma) > w_\tau({}_{i\leftrightarrow j}(\sigma))$. If $w_\tau({}_{i\leftrightarrow j}(\sigma)) = 0$ this holds, otherwise we have

$$\frac{w_\tau(\sigma)}{w_\tau({}_{i\leftrightarrow j}(\sigma))} \;=\; \frac{\tau_i^{x_i}\tau_j^{x_j}}{\tau_i^{x_j}\tau_j^{x_i}} \;=\; \left(\frac{\tau_i}{\tau_j}\right)^{x_i - x_j} \;>\; 1. \tag{D.3}$$

Considering the element $\frac{1}{K}(K-1, 1, 0, \ldots, 0)$ from $S$, which is in $S^*$ if $K \geq 3$, we may conclude from proposition 5 that $w$ is also covering. The problem with $K = 1$ or $K = 2$, is that we have respectively $S = \{(1,0), (0,1)\}$ and $S = \{(1,0), (\frac{1}{2}, \frac{1}{2}), (0,1)\}$, in case of two possible words (with more words the reasoning is similar). In both cases there are no observations—word frequencies—which can be amplified. We obviously have $s\left((1,0)\right) = (1,0)$ and $s\left((0,1)\right) = (0,1)$, but also $s\left((\frac{1}{2}, \frac{1}{2})\right) = (\frac{1}{2}, \frac{1}{2})$ if $s$ is symmetrical. Hence $s(\sigma) = \sigma$ for all $\sigma \in S$ and

$$t(\tau) = \sum_{\sigma \in S} w_\tau(\sigma)s(\sigma) = \sum_{\sigma \in S} w_\tau(\sigma)\sigma = \tau, \tag{D.4}$$

using (3).