Naming Games

Bas Cornelissen

August 9, 2017

How can a population negotiate a shared language without central coordination? This is the terrain of naming games, the second class of agent-based models. In local, horizontal interactions, agents 'align' their language until they reach coherence. We discuss several alignment strategies, some of which return in later chapters, and conclude with a proof suggesting that a stable, single-word language always emerges. The model used therein is the stepping stone for the next chapter, where we connect naming games to Bayesian models of iterated learning.

Naming games (NG) or language games were pioneered in the 90s by Luc Steels and colleagues. The view of language that motivated their work was similar to the views expressed in the iterated learning literature. As Steels (1995) puts it, "language is an autonomous adaptive system, which forms itself in a self-organising process". However, language games approach the adaptive system from a different angle than iterated learning. The development of linguistic structure is not primarily driven by transmission, as Kirby and others proposed, but "by the need to optimise *communicative success*" (p. 319, my italics). The central question takes the form (Steels 2011): how can a convention of some sort (lexical, grammatical, or otherwise) emerge in and spread through a population as a result of local communicative interactions, that is, without central coordination? So if iterated learning is a model of *vertical* language evolution, then the naming games model *horizontal* language evolution.

One of the first studies to explore this, Steels (1995), used a game in which (software) agents negotiated a spatial vocabulary. Equipped with a primitive perceptual apparatus, the agents learned to identify each other by name or spatial position in a shared simulated environment. Later research extended this approach to embodied robotic agents, grounding their 'language' in the physical world. These *grounded naming games* (Steels 2012; Steels 2015) introduce additional complexities pertaining to the perceptual and motor systems of the robots. We focus on non-grounded games, which can be divided into two branches. The first is centred around the *minimal naming game*, studied extensively using methods from statistical physics. The second extended the first naming games to more complex and possibly realistic linguistic scenarios. This chapter discusses and compares both branches. Of particular interest is the kind of dynamics one can expect from these models. We therefore conclude with the proof by

De Vylder and Tuyls (2006) suggesting that naming games always converge to a stable, single-word language.

The basic naming game

Picture a group of people encountering a colourless green object for which they do not have a name. Of even worse, suppose they don't have a shared language at all. Confused, I suppose, they furiously shout out names for the object. But can they gradually align their vocabularies by carefully attending to what the others are saying, until they have agreed on a word for the object — *gavagai*, perhaps?

Frivolities aside, this is the essence of the naming game. It imagines a population of N agents in a shared environment filled with objects, which the agents try to name. At the start of the game, there is no agreement whatsoever about the names of the objects. Every agent has an inventory of names for the objects (a lexicon), which is adjusted after every round with the goal of increasing communicative success. In every round, two randomly selected agents interact, one as speaker, one as the hearer, according to the following script (Wellens 2012):

1. The speaker selects one of the objects which is to serve as the *topic* of the interaction. She[1] produces a name for the object, either by using one of the names she already knew, or by inventing a new name.
2. The hearer receives the word, interprets it and points to the object he believes was intended.
3. The speaker indicates whether she agrees or disagrees, in that way signalling whether communication was successful.
4. Both the speaker and hearer can update their inventories.

The script is a broad outline and concrete implementations are more specific. How, for example, does the speaker select a word in step 1? The typical assumption is that the speaker uses her own experience as a proxy of the hearer's inventory and opts for a signal she would likely interpret correctly herself. This is a so called *obverter* strategy (Oliphant and Batali 1996). Or more importantly, how do the speaker and hearer update their lexicons after the encounter? Here, the sky is the limit. Does the speaker update her lexicon, or the hearer, or both? What happens after successful communications, what after failure? In years of research, one particular script emerged, which is discussed below. It also became clear that whichever update strategy is used, it must improve the *alignment* between the lexicons Steels (2011). That means that the probability that a future encounter will be successful is increased. Such strategies thus reinforce successfully communicated words and this often installs a winner-takes all dynamics which, in the end, leads to a (unique) shared convention. This is best seen in the so called *minimal naming game*.

1 'Gender' is only introduced to convieniently disambiguate the intended agent: the speaker (she) or the *h*earer (he). This even puts the 'men' in the role of listener — which I belief is sometimes regarded to be the appropriate role.

A. Failed communication

SPEAKER	HEARER	⟹	SPEAKER	HEARER
Gavagai	Spam		**Gavagai**	Spam
Cofveve	Foo		Cofveve	Foo
Spam			Spam	**Gavagai**

B. Successful communication

SPEAKER	HEARER	⟹	SPEAKER	HEARER
Gavagai	**Spam**		**Spam**	**Spam**
Cofveve	Foo			
Spam				

FIGURE 1 The updates of the minimal naming game illustrated. If communication fails, the hearer adds the word uttered by the speaker (bold) to its vocabulary. After a success, both empty their vocabularies and keep only the communicated word.

Figure inspired by Wellens (2012).

The minimal strategy

The *minimal naming game* was introduced by statistical physicist Andrea Baronchelli (2006) and simplifies earlier naming game in several respects (Baronchelli, Felici, et al. 2006). First, it assumes that homonymy cannot occur. Homonymy can only be introduced when a speaker invents a *new* word for an object that happens to have been used already to name another object. If the space of possible new words is large enough, we can safely assume that invented words are unique and homonymy will be absent. Secondly, one can assume, without loss of generality, that there is only one object. If there is no homonymy, the update in step 4 will never affect words used for a different object. The competition between the synonyms for a particular object is thus completely independent from other objects. As a result, the dynamics of a naming game with multiple objects is fully determined by the dynamics of a game with a single object.

In the minimal naming game, the inventory of every agent is a list of words. In step 1, the speaker select one word uniformly at random from her inventory. The update in step 4 distinguishes two cases.

- **Success.** If the hearer knows the word, communication is successful. Both hearer and speaker remove all *other* words from their inventories, yielding two perfectly aligned inventories with one single word.
- **Failure.** If the hearer does not know the word, communication fails and the hearer adds the word to his lexicon.

Figure 1 illustrates how the inventories of agents change after failed and successful communication. The dynamics of the games can be studied by collecting several statistics (cf. Baronchelli 2017; Wellens 2012), typically with a certain resolution (e.g. after every 10 rounds). Concretely, we measure the following:

- **(Probability of) communicative success** $p_s(t)$. The probability that an interaction at time t is successful. These probabilities are estimated by averaging this binary variable over many runs.
- **Total word count** $N_{\text{total}}(t)$ The total number of words used in the population at time t. Some authors prefer to divide it by the population size to get the average number of words per agent.
- **Unique word count** $N_{\text{unique}}(t)$. The number of unique words used in the population at time t.

Due to the stochasticity of the games, individual runs vary substantially and can obscure underlying regularities. Conversely, the behaviour of a single run can suggest

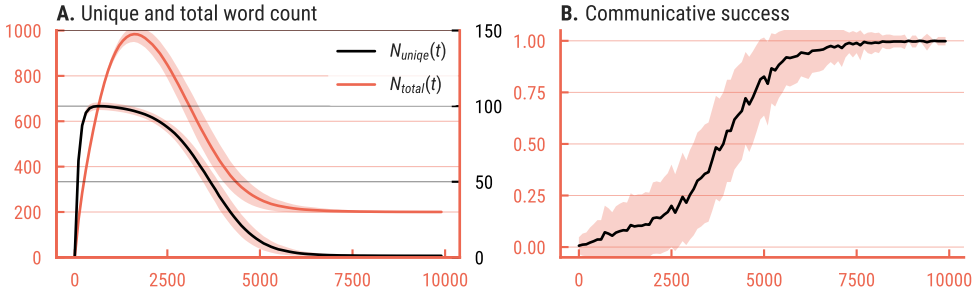

FIGURE 2 The dynamics of the minimal naming game. An sharp transition leads to convergence and the emergence of consensus.

MNG01 Results shown for $N = 200$; avg. of 300 runs, 1 std. shaded.

regularities that do not generalise. For that reason, we study the average behaviour of the games, obtained by averaging over many simulation runs.

PHENOMENOLOGY The minimal naming game goes through three distinct phases, as illustrated in figure 2. In the first phase, most interacting agents will have empty vocabularies and thus invent new words. This results in a sharp increase of the number of unique words N_{unique} in the population. In the second phase, no new words are invented, but the invented words spread through the population. Alignment is still low and words will rare rarely be eliminated, so N_{total} keeps growing. In the third phase, after the peak of N_{total}, this changes. Interactions are increasingly likely to be successful, leading to a sharp increase in communicative success and a drop in N_{total} as more and more words are eliminated. This also results in the characteristic S-shaped curve of p_{success}. Eventually the population reaches coherence in the absorbing state where all agents share one unique word and reach perfect communicative success ($N_{\text{unique}} = 1$, $N_{\text{total}} = N$ and $p_{\text{success}} = 1$).

The game has two important properties, that one might call *effectiveness* and *efficiency*. The resulting communication system is *effective* because agents learn to communicate successfully, and *efficient* in the sense that agents do not memorise more words than strictly necessary (one, in this case). A simple argument shows that the minimal naming game almost always reaches an efficient and effective stable state (Baronchelli, Felici, et al. 2006). At any point in the game, there is a positive probability of reaching coherence in $2(N-1)$ steps: pick one speaker and let her speak to all other $N-1$ agents twice. The first time, a hearer might still have to adopt the word, but after the second interaction only one word will remain in his inventory. If p is the probability of this (unlikely) sequence of interactions, the probability that it has not occurred after $k \cdot 2(N-1)$ steps is less than $(1-p)^k$, which decreases exponentially in k. With probability 1, the population will thus reach coherence as $k \to \infty$. The argument is somewhat unsatisfactory as it does not reveal anything about the dynamics: how fast is the convergence, for example?

SCALING RELATIONS AND NETWORK STRUCTURE To obtain a better insight in the dynamics, one can adopt a methodology commonly used in statistical physics and look at *scaling relations*. The question is then how certain quantities, like convergence time, *scale* with the size of the system, i.e. the number of agents. To that end, two critical points are identified: the time t_{conv} where the game reaches coherence and the time

$t_{\max}$ at which point $N_{\text{total}}(t)$ reaches its maximum. It turns out that these quantities depend on the population size N in a power-law fashion (Baronchelli, Felici, et al. 2006; Loreto et al. 2011):

$$t_{\text{conv}},\ t_{\max},\ N_{\text{total}}(t_{\max}) \propto N^{\alpha} \qquad \text{where } \alpha \approx 1.5 \tag{1}$$

Now note that $N_{\text{total}}(t_{\max})/N$ is the maximum number of words each agent has to store on average — the maximum memory load, perhaps. Baronchelli (2017) concludes that "the cognitive effort an agent has to take, in terms of maximum inventory size, *depends on the system size* and, in particular, diverges as the population gets larger" (Baronchelli 2017, italics in original). Although interesting, I would be hesitant to concede that linguistic activity in a small language community requires less cognitive effort than the same activity in a larger community.

Besides the scaling effects, the role of the network structure of the population has been studied extensively (see Baronchelli 2017, for an overview). In the classical naming game any two agents can interact — there is *homogeneous mixing* — corresponding to a fully connected social network. Varying the topology (to e.g. more realistic small-world networks, Dall'Asta et al. 2006) strongly influences the dynamics. This is reflected by different scaling relations, but not by convergence per se: the population still negotiates a unique word — as long as the networks remains connected, of course.

Lateral inhibition strategies

The minimal strategy is somewhat opportunistic in that it forgets all other words after a successful encounter. It has been suggested that subtler alignment mechanisms might yield faster convergence times: so called *lateral inhibition strategies* (see Wellens 2012 ch. 2, for an overview). The name is ultimately derived from biology, where excited neurons can be found to *inhibit* neighbouring neurons. Similarly, lateral inhibition strategies decrease the chance of using competing words again. To that end, they assign a *score* to every word. If a word is communicated successfully, its score is increased, and the scores of competitors are decreased or *inhibited*. The production mechanism must also accounts for the scores, typically by producing the highest-scoring word.

The (basic) lateral inhibition strategy was first formulated in Steels and Belpaeme (2005) and is described by five nonnegative parameters (Wellens 2012)[2]

$$\delta_{\text{inc}},\quad \delta_{\text{inh}},\quad \delta_{\text{dec}},\quad s_{\text{init}},\quad s_{\max}. \tag{2}$$

After a success, both agents increase the score of the communicated word by δ_{inc} and decrease scores of competitors by δ_{inh}. After a failure, the hearer adopts the word with score s_{init} and the speaker decreases the score by δ_{dec}. Whenever a score drops below (or equals) 0 the word is removed, and scores can never grow larger than $s_{\max}$. Other inhibition strategies have also been used and will be discussed in chapter **??**.

The minimal strategy is a special case of the lateral inhibition strategy, for $\delta_{\text{inc}} = \delta_{\text{dec}} = 0$ and $\delta_{\text{inh}} = s_{\text{init}} = 1$ (see also table 1). With those parameters new words get score 1 and this score is never further increased. It *can* be inhibited, by 1, which leads to to immediate removal. In this strategy, the scores thus play a purely administrative

2 Wellens (2012) only uses δ's in $(0, 1)$, but this general formulation allows the inclusion of the frequency strategy.

TABLE 1 Parameter settings for four different strategies, whose behaviour is shown in figure 3. Note that equivalent parametrisations also exist; see main text for details.

	δ_{inc}	δ_{inh}	δ_{dec}	s_{init}	s_{max}
MINIMAL STRATEGY	0	1	0	1	1
LAT. INHIBITION STRATEGY 1	1	1	0	1	∞
LAT. INHIBITION STRATEGY 2	0.1	0.5	0.1	0.5	1
LAT. INHIBITION STRATEGY 3	0.1	0.2	0.2	0.5	1
FREQUENCY STRATEGY	1	0	0	1	∞

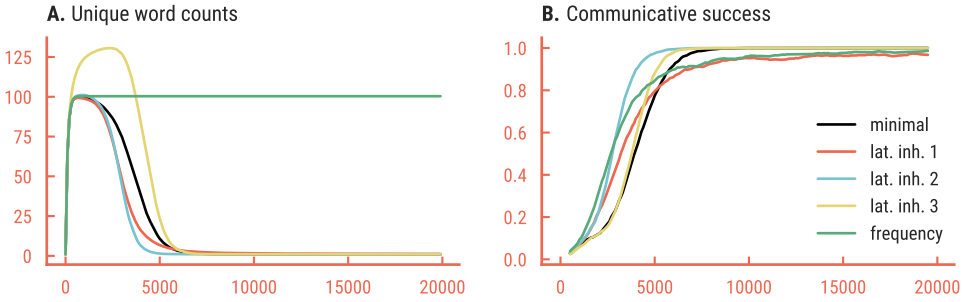

FIGURE 3 Comparison of the four naming game strategies in table 1. The the unique word count and communicative success show that all strategies reach communicative success. The stable language for the frequency strategy is not efficient.

LINGO1 Results shown for $N = 200$; avg. of 300 runs. p_{success} is a rolling average over a centered window of 1000 iterations.

role. A strategy where scores play a larger role, is the *frequency strategy* which counts how often every word has been encountered. This strategy however exhibits no form of lateral inhibition. The minimal strategy and frequency strategy thus mark two extremes: the former has the strongest possible form of lateral inhibition, the latter none. Between these endpoints lie the proper lateral inhibition strategies.

I want to discuss three fairly different LI strategies here: LI strategy 1 is a strategy that returns in chapter **??**; strategy 2 is taken from Wellens (2012); and strategy 3 is a variation thereof. The parameters are listed in table 1 and figure 3 shows the dynamics. First of all note that the dynamics of N_{unique} can strongly differ for different strategies (subfigure A). If for example $\delta_{\text{inh}} = \delta_{\text{dec}}$ as in LI strategy 3, many more words can be invented. But eventually this strategy gives rise to an efficient language. So do all other strategies, except that the frequency strategy results in a maximally inefficient languages where all agents know all words. Since agents only use the most frequent word, perfect communicative accuracy is still attained, as is the case for the other strategies.

These are just five strategies, but what does the rest of the strategy space look like? In appendix **??** I systematically explore a larger part of the space, following Wellens (2012). I indeed find that δ_{inh} interpolates between the minimal and frequency strategy. Further, relatively large δ_{inc} can lead to temporary stabilisation at a non-equilibrium state, until inhibition takes over the stable state is reached. However, I should note that I do not replicate Wellens's finding that the frequency converges faster than the minimal strategy (see also figure 3), and have not been able to reconstruct why. Although the behaviour might vary initially, the long-term behaviour is unaffected: convergence to a single-word language.

In sum, all strategies discussed allow the population to solves the naming problem and leads to effective communication within the population. Any form of lateral inhibition dampens competing words, a result of which agents eventually forget all but one word. The frequency strategy is the only discussed strategy that is not *efficient* in

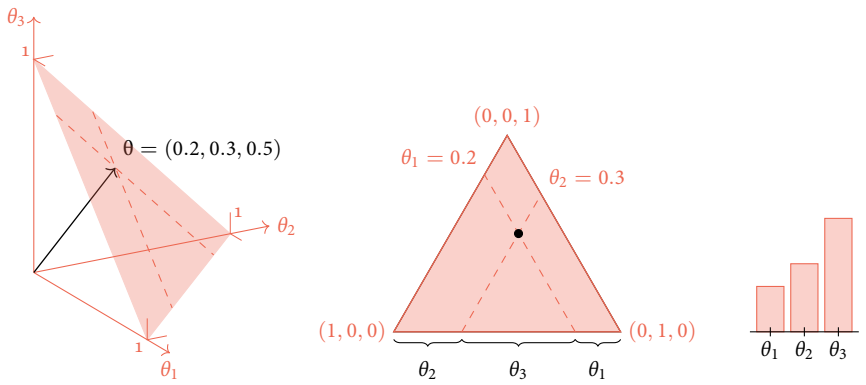

FIGURE 4 A discrete distribution θ over three values corresponds to a point in the 2-simplex, a triangular slice of $\mathbb{R}^3$ (left). The simplex can be embedded in the plane (middle), so that every point in the triangle determines a distribution (right).

this sense. For different parameter settings communicative success can increase earlier, later or even stabilise temporarily, but will eventually be reached nonetheless. Indeed, it seems that "adding a scoring mechanism yields only marginal improvements in terms of communicative and alignment success" (Wellens 2012, p. 23)[3] Why, one wonders, is the convergence so robust?

Proof of convergence

To the best of my knowledge, De Vylder and Tuyls (2006) provided the only analytical result indicating that non-minimal naming games converge to a shared, single-word language. The results apply to a variant of the game, which makes similar simplifications as the minimal naming game: there is no homonymy and only a single object. It moreover starts 'later' in the game, when all agents have already engaged in an interaction and no new words are invented. At this point, there are K unique words $w_1, \ldots w_K$ in the game and the authors assume none of these is ever removed — very much like the frequency strategy. Similarly, speakers use observed frequencies to determine which word they will produce. For production strategies that reinforce or amplify the most frequent word, the authors are able to prove convergence to a single-word language. However, their proof applies to a *deterministic* model, the *sampling-response model*, and De Vylder and Tuyls use simulations — not a proof — to argue that their results generalise to the actual *stochastic*, *turn-based model*. I will present the deterministic, *sampling-response* model in some detail, partly because it is the stepping stone for the next chapter.

PRELIMINARIES First of all, we need to introduce the *simplex*: the space of discrete probability distributions. A probability distribution over K words is described by a vector $\theta = (\theta_1, \ldots, \theta_K)$ such that all θ_k are positive, and they together sum to 1, i.e. $\sum_k \theta_k = 1$. Note that the last entry, θ_K, is determined by the others and constraint $\sum_k \theta_k = 1$. Probability vectors therefore lie in a $(K-1)$-dimensional slice of $\mathbb{R}^K$. This slice is known as the $(K-1)$-*simplex* Δ^{K-1}, or simply Δ if no confusion can arise. The 2-simplex corresponds to a triangle, as illustrated in figure 4.

The model proposed by De Vylder and Tuyls (2006) considers a population of N

3 That is, for the basic naming game, since Wellens (2012) finds that in more complicated games, subtle update mechanisms *can* be beneficial.

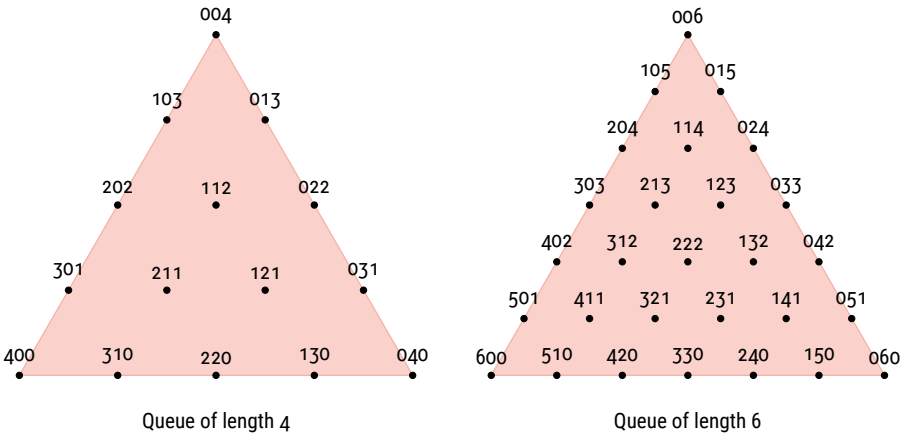

FIGURE 5 All possible frequencies of 3 words in a queue of length 4 (left) and 6 (right) form a discrete subset of the simplex. The corresponding relative frequencies are the 'languages' used by agents in the sampling-response model. Frequencies (a, b, c) are labeled abc.

Figure inspired by (De Vylder and Tuyls 2006).

agents who keep a queue of the last Q words they have observed.[4] A speaker will utter a word based on the relative frequencies of the words in her queue. Formally, we write $\mathbf{c} = (c_1, \ldots, c_K)$ for the vector of *counts*, i.e. c_k the the number of k's in the queue. The counts correspond to (relative) frequencies $\theta = (\theta_1, \ldots, \theta_K)$ where $\theta_k = c_k/Q$. The point $\theta = (0.2, 0.3, 0.5)$ in figure 4 for example depicts the frequencies of $K = 3$ words in a queue of length $Q = 10$ with 2 occurrences of w_1, 3 of w_2 and 5 of w_3. By 'frequencies' θ we from now on mean *relative* frequencies and we also call θ the *language* of an agent. The frequencies lie in a discrete subset Δ_Q of the simplex which depends on the size of the queue Q (see figure 5).

Given a language, a *response function* r determines with what probability each word is uttered. Consider for example the response function r that puts all mass on the most frequent word. In our example with $\theta = (0.2, 0.3, 0.5)$ this means that $r(\theta) = (0, 0, 1)$, so the probability of uttering w_3 is $p(x = w_3 \mid \theta) = 1$. More generally, $r : \Delta \to \Delta$ maps the language θ_A of agent A to a *word distribution* $\pi_A := r(\mathbf{z}_A)$, such that the probability of uttering word $x = w_k$ is

$$p(x = w_k \mid A) = \pi_{A,k}, \qquad \text{where } \pi_{A,k} = \big[r(\theta_A)\big]_k \tag{3}$$

THE SAMPLING-RESPONSE MODEL It is not easy to analyse this game directly. Consider how the language θ of a hearer changes during an interaction. The only thing that matters is the probability of hearing a word, not which speaker uttered it. We obtain those probabilities by averaging over all possible speakers (for simplicity, agents are allowed to speak to themselves),

$$p(x = w_k) = \bar{\pi}_k, \qquad \text{where } \bar{\pi} = \frac{1}{N}\sum_{A=1}^{N} \pi_A, \tag{4}$$

and call this average word distribution the *aggregate languages* as it aggregates the languages of all agents. Since the language of the hearer changes in every round, the aggregate language $\bar{\pi}$ also varies from round to round. To obtain a analysable model, De Vylder and Tuyls (2006) nonetheless assume it temporarily remains constant. In

4 The notation of De Vylder and Tuyls (2006) maps to ours as follows: $n \rightsquigarrow K$, $K \rightsquigarrow Q$, $m_i \rightsquigarrow c_i$, $x_i \rightsquigarrow q_i$, $s(k) \rightsquigarrow \pi_k$ $\Sigma \rightsquigarrow \Delta$, $\sigma \rightsquigarrow \theta$ (mostly), and $\tau \rightsquigarrow \bar{\pi}$.

the resulting *sampling-response* model all agents interact synchronously in successive *episodes*. During an episode all agents simultaneously receive Q utterances, drawn from the aggregate language $\bar{\pi}$. One episode therefore corresponds to multiple rounds of the original turn-based game, enough to ensure that all agents have 'flushed' their queues, i.e. acted as a hearer Q times. Indeed, the analogy is not perfect, but deliberately so.

Importantly, *the sampling-response model is deterministic* and analysing how an agent's language changes during an episode becomes much easier. Concretely, if $\bar{\pi}$ is the aggregate language during an episode t, the probability of observing frequencies θ_t is the probability of observing the corresponding counts c_t amongst Q independent draws from $\bar{\pi}_t$. A multinomial probability, that is, so the sampling-response model takes the form

$$\theta_t \mid \bar{\pi}_t \sim \text{Multinomial}\big(c_t \mid \bar{\pi}_t\big), \qquad c_t = Q \cdot \theta_t, \tag{5}$$

$$x_t \mid \theta_t \sim \text{Categorical}(r(\theta_t)) \tag{6}$$

We can use this to compute the word distribution of agent A *after* episode t:

$$\begin{aligned} \pi_{A,k}^{(t+1)} &= p(x_{t+1} = w_k \mid A, \bar{\pi}_t) && (7)\\ &= \sum_{\theta \in \Delta_Q} p(x_{t+1} = w_k \mid \theta) \cdot p(\theta \mid \bar{\pi}_t) && (8)\\ &= \sum_{\theta \in \Delta_Q} [r(\theta)]_k \cdot \text{Multinomial}(c \mid \bar{\pi}_t). && (9) \end{aligned}$$

Note that the word distribution does not depend on A. This implies that the next aggregate language is also

$$\bar{\pi}_{t+1} = \sum_{\theta \in \Delta_Q} r(\theta) \cdot \text{Multinomial}(c \mid \bar{\pi}_t). \tag{10}$$

This defines a deterministic transition $\bar{\pi}_t \mapsto \bar{\pi}_{t+1}$ from the aggregate language in one episode to the next.

CONVERGENCE In summary, in episode t of the sampling-response model, all agents simultaneously hear Q words drawn from the aggregate language $\bar{\pi}_t$. At the end of the episode, all agents have updated their language, resulting in a new aggregate language $\bar{\pi}_{t+1}$. The new language is a deterministic function of $\bar{\pi}_t$, but also depends on the response function r. De Vylder and Tuyls (2006) showed that under certain conditions $\bar{\pi}_t$ converges to an aggregate language with only a single word:

$$\lim_{t \to \infty} \bar{\pi}_t = (0, \ldots, 0, 1, 0, \ldots, 0). \tag{11}$$

Perhaps the most important condition was that the response function must be *amplifying*. That means, roughly, that the response function increases the probability of producing the most frequent word (with respect to its frequency). We have already seen the prime example of an amplifying function: the function that *exponentiates* a distribution (see figure **??**):

$$r_\eta(\theta) = \frac{1}{\sum_{k=1}^{K} \theta_k^\zeta} \cdot \left(\theta_1^\zeta, \ldots, \theta_K^\zeta\right), \qquad \zeta > 1 \tag{12}$$

With this response function, the population would eventually adopt a language with only one word.

Conclusions

Naming games try to understand how self-organisation can lead to the emergence of a shared vocabulary. To reach coherence, agents have to align their vocabularies after every encounter, for which various strategies can be used. The strategies discussed in this chapter — the minimal, frequency, and lateral inhibition strategies — all lead to the emergence of a consensus. As long as the alignment strategy implements some kind of competition damping, for example in the form of lateral inhibition, the resulting language is effective and agents remember no more words than strictly required. The frequency strategy was the exception, where agents remember all words but nevertheless reach full communicative success.

The naming games discussed in this chapter are the simplest, but, it seems, most important models in the literature. By dropping various assumptions, different games have been obtained. One could for example allow homonymy, in which case coherence can then still be reached by damping competing homonyms and synonyms. This introduces a kind of *mapping-uncertainty* (Wellens 2012) (which word corresponds to which object?) that will return in chapter **??**. This problem is stronger in so called *Guessing games* where the speaker is not allowed to indicate the object the hearer, who has to *guess* the intended object. Rather than simplifying the basic naming game, it can also be extended. Recently, Steels (2015) for example proposed the *syntax game*, where agents communicate n-ary relationships rather than words. Both the script and underlying mechanism are in the end very simimlar to the original naming game. This also seems to hold for work that has adopted *fluid construction grammars* to extend the representational capacities, hoping to move closer towards natural language (see e.g. Steels (2016) for an overview). Since this thesis concerns itself with the dynamics of the underlying game, such extensions have been left out.

The underlying, long-term dynamics of naming games seems rather clear. Where Bayesian iterated learning found a convergence to the prior, in naming games one finds convergence to single-word, coherent stable language, if the alignment mechanism somehow amplifies the highest-scoring word. So much both the proof of De Vylder and Tuyls (2006) and experimental results suggest. The proof applies only to a deterministic variant of the naming game, and it remains an open problem to show convergence for stochastic naming games. The wide range of experimental results suggests this should be possible — at least as long as the rules of the game are respected. As soon as the rules are changed, convergence can break. Baronchelli, Dall'Asta, et al. (2007) for example introduced a parameter β regulating the probability with which agents update their inventories. They find that for values of β below some critical point, multiple words can survive in the population.

The desiderata formulated in the last chapter were clearly motivated by iterated learning models, and might not be directly relevant for naming games. Nevertheless, it should be pointed out that naming games give rise to stable languages **??**, are to some extend analysable **??** and appear to be robust to population structure **??**. They include

various strategies **??**, but the resulting behaviour is always similar terms of long-term behaviour. One desideratum the naming game does clearly not fulfil is the explicit representation of the learning biases. In fact, agents have hardly any cognitive makeup, but this is addressed in the next chapter.

Bibliography

Baronchelli, Andrea (2006). "Statistical mechanics approach to language games". PhD thesis. Università di Roma "La Sapienza".

— (2017). "A gentle introduction to the minimal Naming Game". In: pp. 1–24. DOI: 10.1075/bjl.30.08bar. arXiv: 1701.07419.

Baronchelli, Andrea, Luca Dall'Asta, et al. (2007). "Nonequilibrium phase transition in negotiation dynamics". In: *Physical Review E* 76.5, p. 051102. DOI: 10.1103/PhysRevE.76.051102. arXiv: 0611717 [cond-mat].

Baronchelli, Andrea, Maddalena Felici, et al. (2006). "Sharp transition towards shared vocabularies in multi-agent systems". In: *Journal of Statistical Mechanics: Theory and Experiment* 2006.06, P06014–P06014. DOI: 10.1088/1742-5468/2006/06/P06014. arXiv: 0509075 [physics].

Dall'Asta, L et al. (2006). "Agreement dynamics on small-world networks". In: *Europhysics Letters (EPL)* 73.6, pp. 969–975. DOI: 10.1209/epl/i2005-10481-7. arXiv: 0603205 [cond-mat].

De Vylder, Bart and Karl Tuyls (2006). "How to reach linguistic consensus: A proof of convergence for the naming game". In: *Journal of Theoretical Biology* 242.4, pp. 818–831. DOI: 10.1016/j.jtbi.2006.05.024.

Loreto, Vittorio et al. (2011). "Statistical physics of language dynamics". In: *Journal of Statistical Mechanics: Theory and Experiment* 2011.04, P04006. DOI: 10.1088/1742-5468/2011/04/P04006.

Oliphant, Michael and John Batali (1996). "Learning and the Emergence of Coordinated Communication". In: *Center for research on language newsletter* 11.1, pp. 1–46. DOI: 10.1.1.27.2287.

Steels, Luc (1995). "A Self-Organizing Spatial Vocabulary". In: *Artificial Life* 2.3, pp. 319–332. DOI: 10.1162/artl.1995.2.319.

— (2011). "Modeling the cultural evolution of language". In: *Physics of Life Reviews* 8.4, pp. 339–356. DOI: 10.1016/j.plrev.2011.10.014.

— (2012). "Introduction. Self-organization and selection in cultural language evolution". In: *Experiments in Cultural Language Evolution*, pp. 1–37. DOI: 10.1075/ais.3.02ste.

— (2015). *The Talking Heads experiment: Origins of words and meanings*. Ed. by Luc Steels and Remi van Trijp. Computational Models of Language Evolution. Language Science Press. DOI: 10.17169/langsci.b49.75.

Steels, Luc (2016). "Agent-based models for the emergence and evolution of grammar". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1701, p. 20150447. DOI: 10.1098/rstb.2015.0447.

Steels, Luc and Tony Belpaeme (2005). "Coordinating perceptually grounded categories through language: A case study for colour". In: *Behavioral and Brain Sciences* 28.04, pp. 469–529. DOI: 10.1017/S0140525X05000087.

Wellens, Pieter (2012). "Adaptive Strategies in the Emergence of Lexical Systems". PhD thesis. Vrije Universiteit Brussel.