Learning with neighbours

Emergence of convention in a society of learning agents

Roland Mühlenbernd
University of Tübingen

Abstract. I present a game-theoretical multi-agent system to simulate the evolutionary process responsible for the pragmatic phenomenon *division of pragmatic labour* (DOPL), a linguistic convention emerging from evolutionary forces. Each agent is positioned on a toroid lattice and communicates via *signaling games*, where the choice of an interlocutor depends on the *Manhattan distance* between them. In this framework I compare two learning dynamics: *reinforcement learning* (RL) and *belief learning* (BL). An agent's experiences from previous plays influence his communication behaviour, and RL agents act in a non-rational way whereas BL agents display a small degree of rationality by using *best response dynamics.* The complete system simulates an evolutionary process of communication strategies, which agents can learn in a structured spatial society. The significant questions are: what circumstances could lead to an evolutionary process that doesn't result in the expected DOPL convention; and to what extent is interlocutor rationality necessary for the emergence of a society-wide convention à la DOPL?

Keywords: Multi-agent system, division of pragmatic labour, signaling games, learning dynamics, communication strategies, simulation of an evolutionary process

Abbreviations: DOPL – division of pragmatic labour; BL – belief learning; NE – Nash equilibrium; RD – replicator dynamics; RL – reinforcement learning

1. Introduction

During a party you accidentally eavesdrop on a group of colleagues talking about suspicious sounding topics. By chance you pick up the expression "John went to (the) jail", and you're not certain if the speaker has really used the word "the". What difference does this word make? Obviously a deciding one! With this word you would infer the literal interpretation that John went to the jail *building*, maybe to visit a prisoner, but without it you would infer the prototypical interpretation that John himself has been jailed. This example is an instance of the pragmatic rule *division of pragmatic labour* (DOPL), originated by Horn (1984). This rule says that a simple (unmarked) expression describes a prototypical case, whereas a complex (marked) expression describes a rare case. Horn presented many examples obeying his rule and depicting conventional language use.

Recently, game-theoretical accounts characteristically based on Lewisean (1969) *signaling games* have gained popularity as a means to examine pragmatic phenomena and linguistic conventions. Lewis showed that a convention can be seen as a successful, self-reinforcing communication strategy. According to these Lewisean signaling games, a convention is, in rough approximation, a combination of strategies that form a strict *Nash equilibrium* (NE). But for a multitude of phenomena modelled via signaling games there exist more than one NE. But which one depicts the convention? A host of work has been done to find a solution concept for this task of *equilibrium selection*, especially by using evolutionary game theory, and therein the popular *replicator dynamics* (RD) account: in this framework, for example, Foster & Young (1990) argued in support of the *stochastically stable* equilibrium and van Rooy (2004a) argued that the *strictly efficient* strategy is the unique equilibrium. All of these solution concepts show that for signaling games parametrized in line with DOPL,[1] the only convention is a communication strategy obeying Horn's rule. But is one or another of these solution concepts explanatory enough to convince us that this convention will always emerge?[2] What happens if we modify the RD account by extending update dynamics and network structure? Are there conditions which avoid an emergence of a convention obeying Horn's rule? This all leads me to the first significant research question of this paper:

> *What circumstances could lead to an evolutionary process that doesn't result in an equilibrium strategy of Horn's rule as sole convention?*

In this line of articles that treated pragmatic phenomena as linguistic conventions derived through signaling games, van Rooy (2004a) was exclusively concerned with the emergence of DOPL in language use. He remarks that there evolved two traditions for the explanation of this emergence, which Blutner & Zeevat (to appear) called the *diachronic view* and the *synchronic view*. Horn himself had the diachronic view in mind: He affirmed that this rule is a convention as a result of language change and traces back to evolutionary forces. In contrast, the synchronic view (Levinson (2000), Parikh (1991), Jäger & Ebert (2009)) apprehends language use according to DOPL not as a convention but as a result of online processing concerning rationality considerations. Whereas most examples depicting language use according to DOPL call for a diachronic treatment, some call for a synchronic treatment. But it isn't always possible to make a clear dichotomy for every example. Today's interpretation as result of rational online processing could

[1] In the next section I'll show how to set signaling games parameters in the way that the game depicts a DOPL situation.

[2] Indeed a convention obeying Horn's rule cover the majority of examples you can find in the literature. But you can also find examples depicting a conventions obeying the opposite of Horn's rule, which, as we will see in the next chapter, is called anti-Horn. Such examples are given e.g. by Schaden (2008)

be tomorrow's convention. This potential connection leads me to the second significant research question:

> *To what extent is interlocutor rationality necessary for the emergence of DOPL as a society-wide convention?*

With this paper I want to answer these questions by situating signaling games in a society of learning agents interacting on a simple social network. By modifying the structural features of the network as well as by changing the agents' learning dynamics, I want to examine the circumstances responsible for non-emergence of communication strategies according to Horn's rule. To find out the extent to which rationality of interlocutors is necessary for the emergence of DOPL, I compare two learning accounts: The first is *reinforcement learning*, which depicts non-rational, obstinate behaviour. The second is *belief learning* via *fictitious play* combined with *best response dynamics*, which depicts a more empathetic behaviour and a small degree of rationality.

2. Signaling games in evolution

To model strategic language use, I'll use signaling games, as introduced by Lewis (1969). According to his general idea, a common communicative situation between a sender and a receiver is defined in the following way: The sender has a private information state $t \in T$ and has to choose a message $m \in M$ to communicate it. The receiver has to construe this message by choosing an interpretation $a \in A$ as response to it.

For a signaling game to examine Horn's rule two information states are necessary, one for the prototypical and one for the rare case: $T = \{t_p, t_r\}$. Furthermore, there are two messages, an unmarked and a marked one: $M = \{m_u, m_m\}$. The interpretations correspond to the information states: $A = \{a_p, a_r\}$. To distinguish the prototypical and the rare cases, a probability function $Pr : T \to \mathbb{R}$ is necessary, which describes how likely it is that a state t is topic of communication. Thus the probability of the prototypical case must be higher than the probability of the rare case: $Pr(t_p) > Pr(t_r)$. In my model I adapt two different DOPL games, one with the values $Pr(t_p) = .6$ and $Pr(t_r) = .4$, which I call the *weak* DOPL game and one with the values $Pr(t_p) = .7$ and $Pr(t_r) = .3$, which I call the *strong* DOPL game. A distinction between marked and unmarked expression is achieved by a cost function $c : M \to \mathbb{R}$, whereby the marked message is more complex and therefore more expensive than the unmarked one: $c(m_m) > c(m_u)$. I use the values: $c(m_u) = .1$, $c(m_m) = .2$ for both DOPL games. The reason for such a function is that high complexity of a signal may lead to low utility, as Jäger (2008) pointed out. And as we'll see next the cost value dimishes the utility value.

Successful communication means that the receiver construes the interpretation appropriate to the information state the sender wanted to communicate. In this case, the outcome of both players has a utility value of 1, otherwise 0 for failure. To be exact, this value is reduced by the complexity and therefore by the cost value of the used message. Taken together we get a *utility function* for both players, as shown in equation (1).

$$U(t_i, m_j, a_k) = \begin{cases} 1 - c(m_j) & \text{if } i = k \\ 0 - c(m_j) & \text{else} \end{cases} \tag{1}$$

This utility function depicts interlocutors with coinciding interests, namely to communicate successfully. This means that the sender's information state should correspond to the receiver's interpretation. To maximize utility both players should coordinate their actions. The sender's communication behaviour can be described as a pure strategy $S \in [T \to M]$, the receiver's behaviour as a pure strategy $R \in [M \to A]$. In the present example, each participant has four pure strategies to play: Horn strategy (S_H for the sender, R_H for the receiver), anti-Horn strategy (S_{aH}, R_{aH}), Smolensky strategy[3] (S_S, R_S) and anti-Smolensky strategy (S_{aS}, R_{aS}) as depicted in Figure 1.

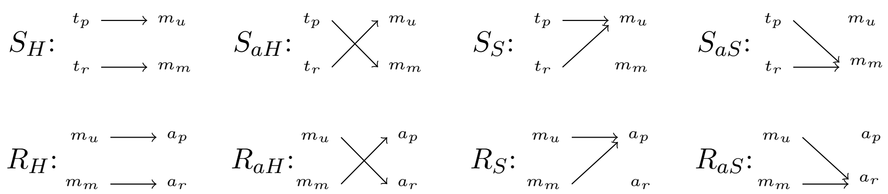


Figure 1. All possible pure sender and receiver strategies

For both participants it holds that behaving according to Horn's rule means playing the Horn strategy, whereas the anti-Horn strategy displays the opposite behaviour. Playing the Smolensky strategy means for the sender to choose the cheapest message in any case and for the receiver to choose the prototypical interpretation state in any case, whereas playing anti-Smolensky strategy means acting exactly the other way around. It is possible to compute how well a sender and a receiver strategy would go together by computing the *expected utility* $EU(S_i, R_j)$ of such a strategy pair (S_i, R_j):

[3] The Smolensky strategy is named in reference to Tesar & Smolensky (1998), who applied this strategy as a starting strategy for agents in their simulations with regards to the assumption that previous generations possibly weren't aware of the complex message.

$$EU(S_i, R_j) = \sum_{t \in T} Pr(t) \times U(t, S_i(t), R_j(S_i(t))) \qquad (2)$$

Figure 2 depicts the expected utilities for all 16 strategy combinations, the left table for the previously defined strong DOPL game and the right table for the standard Lewis game with uniform probabilities and no message costs. The standard solution concept for finding the optimal strategy pair in such a game is the *Nash equilibrium* (NE): a strategy pair (S_i, R_j) for which no player can increase his EU by switching to another strategy. For the strong as for the weak DOPL game there are three Nash equilibria: $E_H = (S_H, R_H)$, which depicts communication according Horn's rule, $E_{aH} = (S_{aH}, R_{aH})$, where both behave exactly in the opposite way and $E_S = (S_S, R_S)$. Only E_H and E_{aH} are strict Nash equilibria, and exclusively E_H is the Pareto efficient NE: the NE with the highest EU. Nevertheless there is no reason why agents shouldn't remain in E_{aH}. If both play anti-Horn and they know the other is doing so, there is no motivation for both to switch to another strategy. Thus we can imagine how both interlocutors would behave if they were settled in an NE. But what kind of processes lead participants to a particular NE?

To find an answer to this question there are two things to do. First, under the assumption that an NE depicts a convention that is common knowledge among a society, it is necessary to extend this two-player game to a multi-agent system where agents can communicate with different dialogue partners. Second, to simulate an evolutionary process, agents should play this game repeatedly and adapt their play accordingly. Thus it is possible that they will change their strategy because of experiences of past played rounds. With these desiderata, our signaling game should be integrated in an *evolutionary game theory* (EGT) account. An abundance of research in recent years has endeavored to combine linguistic phenomena and EGT, especially by integrating the

	R_H	R_{aH}	R_S	R_{aS}
S_H	.87	-.13	.57	.17
S_{aH}	-.17	.83	.53	.13
S_S	.6	.2	.6	.2
S_{aS}	.1	.5	.5	.1

	R_1	R_2	R_3	R_4
S_1	1	0	.5	.5
S_2	0	1	.5	.5
S_3	.5	.5	.5	.5
S_4	.5	.5	.5	.5

Figure 2. Expected utilities for all strategies combinations. *Left table:* Values for the predefined strong DOPL game. *Right table:* Values for the standard Lewis game with uniform probabilities and no message costs.

replicator dynamics (RD) (Taylor & Jonker, 1978), which became the standard model in the field.

A refinement of NE under an EGT point of view is the *evolutionary stable strategy* (ESS). An ESS is a strategy which, if adopted by a population of agents, cannot be invaded by any initially insubstantial alternative strategy. Unfortunately, as van Rooy (2004b) pointed out, even this refinement doesn't lead to only one solution in a DOPL game, because E_H as well as E_{aH} are both ESS's. Benz, Jäger & van Rooy (2005) showed that if a society starts with a setting where all strategies are played with equal probability, under RD it ends up in a state where all agents adhere to E_H.

These results show the superiority of E_H compared with E_{aH} in classical EGT setups. But it is necessary to examine this task in a more realistic account in two ways. First of all, agents should be more sophisticated in that they undergo a learning process rather than a reproduction process like e.g. RD. And according to Huttegger & Zollman (2011), who argue that concepts like ESS seem to have no connection to games with integrated learning models, the solution concepts mentioned above are not in the least bit appropriate. Second, while in all these accounts agents communicate randomly with every possible participant, I use a more realistic structure where the distance of agents on a lattice influences the probability of communication.

3. Update dynamics & Learning accounts

Response rules governing agent behaviour determine the *update dynamics* for repeated games. In a multi-agent system where agents communicate via signaling games, as both sender and receiver, each agent has a sender response rule $\rho_s : T \to \Delta(M)$ that determines the probability of sending a message $m \in M$ for a given state t and a receiver response rule $\rho_r : M \to \Delta(A)$ that determines the probability of choosing interpretation $a \in A$ to construe a received message m. Combining this with a learning account means that agents' responses are influenced by their previous experiences. In the following two subsections I will introduce the two learning accounts *reinforcement learning* (RL) and *belief learning* (BL) with the corresponding response rules agents use to make their decisions.

3.1. Reinforcement Learning

Reinforcement learning can be captured by a simple model based on urns, also known as *Pólya urns*, where the probability of making a particular decision is proportional to the number of balls in an urn. The distribution of the content of urns in each step is a result of previous decisions. Such an urn model for learning processes in games was first applied by Roth & Erev (1995), its combination with signaling games was, inter alia, studied by Skyrms (2010) and can be modelled in the following way: The sender has an urn Ω_t for each state $t \in T$ and each sender urn contains balls b_m for different messages $m \in M$. The number of balls b_m in urn Ω_t at time τ is designated with $m(\Omega_t)_\tau$, the overall number of balls in urn Ω_t at time τ with $|\Omega_t|_\tau$. If the sender is faced with a state t at time τ, he draws a ball b_m from urn Ω_t and sends message m. Thus the sender rule at time τ is the following:

$$\rho_s(t,m)_\tau = \frac{m(\Omega_t)_\tau}{|\Omega_t|_\tau} \tag{3}$$

In compliance with that rule, the receiver has an urn Ω_m for each message $m \in M$ with balls b_a for different actions $a \in A$. The number of balls b_a in urn Ω_m at time τ is designated with $a(\Omega_m)_\tau$, the overall number of balls in urn Ω_m at time τ with $|\Omega_m|_\tau$. If the receiver wants to construe a message m received at time τ, he draws a ball b_a from urn Ω_m and uses the interpretation a, which leads to the following receiver rule at time τ:

$$\rho_r(m,a)_\tau = \frac{a(\Omega_m)_\tau}{|\Omega_m|_\tau} \tag{4}$$

Up to now I have defined the response rules, but not the architecture of the learning process as defined by the update rules. The standard RL account works in the following way: after an agent has made his decision he can observe if communication was successful or not. If a sender observes that the message m he used to communicate state t didn't lead to successful communication, he simply puts the ball back in urn Ω_t. But if communication was successful, he puts not only this ball in urn Ω_t, but a second one. In this way behaviour that leads to success is reinforced because it increases the probability that b_m is drawn in the next round. The same procedure is used for the receiver urns.

For the given DOPL settings things are a little bit more complicated. Because of message costs, (un)successful communication doesn't necessarily lead to the same utility value. If you communicate successfully with the cheap message m_u, the utility value is .9, whereas successful communication with m_m leads to a utility value of .8. Because the utility values have only one decimal place, we obtain an integer by multiplying the utility value by 10. This integer can be used to set the number of balls added or subtracted. E.g. when a sender who communicates successful by sending m_m gets a utility value of .8, his appropriate urn is scaled up by 8 balls b_{m_m}. On the other hand, if unsuccessful communication leads to a utility value of -.2, the content of this urn is reduced by two balls b_{m_m}. To put things formally: If state t is chosen by nature at time τ, message m is the type of ball the sender draws from Ω_t at time τ and action a is the type of ball the receiver draws from Ω_m at time τ, then $\Delta_\tau = U(t, m, a) \times 10$ is the urn update value at time τ. The following equations describe the urn update rule for the sender after drawing b_m from Ω_t:

$$m(\Omega_t)_{\tau+1} = \max[m(\Omega_t)_\tau + \Delta_\tau, 10] \tag{5}$$

Analogously the following equation describes the urn update rule for the receiver after drawing b_a from Ω_m:

$$a(\Omega_m)_{\tau+1} = \max[a(\Omega_m)_\tau + \Delta_\tau, 10] \tag{6}$$

Notice that the $\max[x, 10]$-structure in both update rules ensures that there are at least ten balls of each message in every sender urn and accordingly at least ten balls of each interpretation in every receiver urn. This is a special extension in my model and guarantees that there is always the possibility for each kind of ball to be drawn, even if the probability becomes minute.

Another extension comparing with the standard account is called *lateral inhibition* (Franke & Jäger (to appear)). In this extension there is a *suppressor* value ε and after successful communication with state t, message m and action a the number of balls $b_{m'}$ $\forall m' \neq m$ in sender urn Ω_t will be decreased by ε. Accordingly the number of balls $b_{a'}$ $\forall a' \neq a$ in receiver urn Ω_m will be decreased by ε. That means that successful communication leads not only to increasing the balls involved but also to decreasing the balls not involved. For my model I set $\varepsilon = 3$.

3.2. Belief Learning

While in the RL account agents make no rational decisions, in this account agents display a small degree of rationality in the way that they play a best response by maximizing their *expected utilities*. I call $EU_s(m|t)$ the sender's expected utility for sending message m in state t and $EU_r(a|m)$ the receiver's expected utility for construing message m with interpretation a. Now a sender who wants to communicate state t will use the message m which maximizes his expected utility $EU_s(m|t)$. Accordingly, a receiver who received message m will construe it with the interpretation a which maximizes his expected utility $EU_r(a|m)$. If there are more choices, which maximize those EU's, then each choice is equiprobable. To put it formally: if $MAX(t) = \arg\max_m EU_s(m|t)$

is the set of messages where each one maximizes the sender's EU for a given state t and $MAX(m) = \arg\max_a EU_r(a|m)$ is the set of actions where each one maximizes the receiver's EU for a given message m, then we have the following sender and receiver response rules[4]:

$$\rho_s(t,m) = \begin{cases} \frac{1}{|MAX(t)|} & \text{if } m \in MAX(t) \\ 0 & \text{else} \end{cases} \tag{7}$$

$$\rho_r(m,a) = \begin{cases} \frac{1}{|MAX(m)|} & \text{if } a \in MAX(m) \\ 0 & \text{else} \end{cases} \tag{8}$$

The sender's expected utility $EU_s(m|t)$ returns the utility value the sender can expect for sending message m in state t. But this expected value depends on what he believes the receiver would play. His *belief about the receiver* $B_r(a|m)$ is a function returning the probability that the receiver construes message m with a. Given this belief the sender's expected utility is defined in the following way:

$$EU_s(m|t) = \sum_{a \in A} B_r(a|m) \times U(t,m,a) \tag{9}$$

The receiver's expected utility $EU_r(a|m)$ returns the value the receiver can expect for construing a received message m with interpretation a. Thus he needs to have a *belief about the sender* $B_s(t|m)$ that returns the probability that the sender is in state t by sending message m. Accordingly the receiver's expected utility is defined as follows:

$$EU_r(a|m) = \sum_{t \in T} B_s(t|m) \times U(t,m,a) \tag{10}$$

Now where do these beliefs come from? The BL account of my model engenders a process of acquiring these beliefs by observation. Concretely, a player's belief is a mixed strategy representing all the interlocutor's observed past plays. E.g. assume sender and receiver had the same kind of communicative situation many times before and that function $\sigma(m)$ returns the number of times the sender has sent message m to the receiver. Likewise $\sigma(a|m)$ returns the number of times the receiver has interpreted a received message m with a. Because of these observations the sender has the following belief $B_r(a|m)$ about the receiver:

$$B_r(a|m) = \begin{cases} \frac{\sigma(a|m)}{\sigma(m)} & \text{if } \sigma(m) > 0 \\ 1/|A| & \text{else} \end{cases} \tag{11}$$

[4] arg max stands for the *argument of the maximum*, that is to say, the set of points of the given argument for which the value of the given expression attains its maximum value: $\arg\max_x f(x) = \{x|\forall y : f(y) \leq f(x)\}$

In the same way an evaluation of the receiver's observations about the sender's behaviour leads to belief $B_s(t|m)$ about the sender:

$$B_s(t|m) = \begin{cases} \frac{\sigma(t|m)}{\sigma(m)} & \text{if } \sigma(m) > 0 \\ 1/|T| & \text{else} \end{cases} \tag{12}$$

Notice that both equations contain the condition that the denominator $\sigma(m)$ must be greater than zero. This is not only to avoid a division by zero. It also has a descriptive reason: if there has never been a communicative situation by using message m then both participants cannot have beliefs through past observations. In this case, the probabilities for this message are uniformly distributed, for the sender given by a uniform distribution over all possible interpretations $a \in A$ ($1/|A|$) and for the receiver accordingly over all possible states $t \in T$ ($1/|T|$).

In a repeated game after every communication situation the sender's belief $B_r(a|m)$ and the receiver's belief $B_s(t|m)$ as well will be updated. Hence the belief about the interlocutor's strategy results from previous communications with his dialogue partners. This account is a simple realization of Brown's (1951) *Fictitious Play*.

All in all, both presented learning accounts are defined by a response rule and an update rule. While the update rule of RL only depends on the utility value of a communicative situation, the update rule of BL considers the behaviour of the interlocutor to form a belief about him. This means, that BL is a conscious, empathetic learning process whereas RL can be explained as an unconscious or obstinate learning process. Furthermore, as the response rule shows, BL demands rationality in the sense that the interlocutor has to find the best response for a given belief, whereas RL agents make decisions randomly without any need for rational deliberation.

3.3. Bounded memory

A feature of both learning accounts, and especially of RL, is the fact that learned behaviour manifests itself very early and ingrains itself in the dynamics. Barrett & Zollman (2009) show that forgetting experiences increases both the dynamics of the system and the probability of an optimal language evolving. They introduced different learning accounts based on RL extended by different types of forgetting. I extend both RL and BL accounts with a simple forgetting rule, informally described as following:

As we know both learning accounts' response rules at time τ depend on a history of updates $H = \{u_1, u_2, \ldots, u_\tau\}$, where u_i is an update at time i via an appropriate update rule. An update for RL is specified by the urn update value, whereby an update for BL is specified by changing the values of $\sigma(a|m)$, $\sigma(t|m)$, $\sigma(m)$. This update history H records the information necessary to undo each update u_i. Now we can define a *memory size* μ for an agent: If an agent is at time τ, then undo all updates u_i with $i < \tau - \mu$. Thus all updates that happened more than μ time steps ago are cancelled and therefore have no influence on the response rule. In other words, an agent can't remember that they ever happened; i.e. he has forgotten them.

4. Communication in a social structure

Before I go on describing my model I want to introduce some notation. As agents in my model behave as both sender and as receiver in each communication step, an agent's behaviour is characterized by a pure strategy pair (S, R), which van Rooij (2008) called a *language*. Because there are 16 different combinations of pure strategies for a game with 2 states, 2 messages and 2 actions, there are 16 different languages. In the left table of Figure 3 you can see all languages for the standard Lewis game. The languages L_1 and L_2 are the strict Nash equilibria and shape what Lewis (1969) called a signaling system. I will call these languages *signaling languages*. All the languages L_{p_i} I call *pooling languages* because the agent's sender strategy and/or receiver strategy is a pooling strategy.[5] And the languages L_{m_1} and L_{m_2} are *miscommunication languages*. We will see that for the DOPL games the first three of the four strategies of the diagonal of this language table are of special interest. As you can see in the right table of Figure 3 they are denoted with L_H for *Horn language*, L_{aH} for *anti-Horn language* and L_S for *Smolensky language*.

	R_1	R_2	R_3	R_4
S_1	L_1	L_{m_1}	L_{p_1}	L_{p_2}
S_2	L_{m_2}	L_2	L_{p_3}	L_{p_4}
S_3	L_{p_5}	L_{p_6}	L_{p_7}	L_{p_8}
S_4	L_{p_9}	$L_{p_{10}}$	$L_{p_{11}}$	$L_{p_{12}}$

	R_H	R_{aH}	R_S	R_{aS}
S_H	$\mathbf{L_H}$	L_{m_1}	L_{p_1}	L_{p_2}
S_{aH}	L_{m_2}	$\mathbf{L_{aH}}$	L_{p_3}	L_{p_4}
S_S	L_{p_5}	L_{p_6}	$\mathbf{L_S}$	L_{p_8}
S_{aS}	L_{p_9}	$L_{p_{10}}$	$L_{p_{11}}$	$L_{p_{12}}$

Figure 3. Languages as strategy combination an agent can use. *Left table:* languages for the standard Lewis game with uniform probabilities and no message costs. *Right table:* languages for the predefined DOPL games.

[5] A sender's pooling strategy is one in which the sender chooses the same message for all states. Accordingly a receiver's pooling strategy is one in which the receiver chooses the same interpretation for each message.

In line with the claim I made at the end of the second section, Zollman (2005) argued for a more realistic framework than the existing RD accounts where each agent communicates with any other at random. In contrast to this he analysed signaling games in a multi-agent system with a spatial structure: the agents have a fixed position on a lattice and can only communicate with their direct eight neighbours. To avoid edge phenomena the lattice is mapped on a torus, thus every agent has exactly eight neighbours. Zollman used the following update dynamics: in each round every agent acts both as sender and receiver. After a communication step each agent observes the scores of all his neighbours and if a neighbour scored better than him the agent switches to the strategy of the best neighbour. These dynamics can have an interpretation as cultural evolution by imitation, but also as a limited case of reinforcement learning as Skyrms (2009) pointed out. Thus Zollman's model is a step in the right direction, but by integrating RL or BL in my account I want to simulate a more sophisticated and especially more protracted learning process than just imitating the best.

Furthermore Zollman didn't examine signaling games with a DOPL setting, but the standard Lewis game with uniform probabilities for all states and no message costs. As you can see in the right table of Figure 2, this game has two equally good Nash equilibria $E_1 = (S_1, R_1)$ and $E_2 = (S_2, R_2)$, which depicts the signaling languages L_1 and L_2 in the left table of Figure 3. Now Zollman's simulations led to the following result: Nearly every trial ended up in a population with only signaling language players. But instead of all agents playing the same language, the map is distributed in stable regions of L_1 and L_2 players. Thus this account leads to the evolution of *regional meaning* where an agent's strategy depends on his location in the society.

Wagner (2009) picked up Zollman's account and expanded the analyses in different ways, and two of his amendments are of special interest here. First, Wagner examined cases with non-uniform state probabilities, a situation like in the DOPL settings. Second, he compared neighbourhood communication with more complex network structures. He showed that the state probabilities play a role for the stability of regions and that pooling languages can become stable for particular settings. Furthermore the stability of a region strongly depends on the interactions around its frontiers. Corresponding to this result, simulation runs of my system revealed that for particular game settings frontier structure can be more important for the extinction of languages than the quality of the strategies.

Whereas a population communicating randomly via Lewisean signaling games learns one global signaling language, Zollmann showed that neighbourhood communication leads to multiple signaling languages.

To differentiate both accounts in the way an agent is assigned with a communication partners, I introduce a measure such that each account is placed at the extreme points of its scale. To be exact, such a measure should have the property that for random communication its value is 0, for direct neighbourhood communication it is maximal. I'll call this measure the *degree of locality*. If I define the distance d between two agents on an $n \times n$ toroid lattice as the shortest *Manhattan distance*, then the following condition should hold: the higher the degree of locality, the more probable it is that an agent chooses a communication partner with a small distance d.

To integrate such a measure in my account, an agent's decision for choosing a communication partner will consist of two steps: first he chooses a distance value d, and second he chooses an agent with distance d at random. While each agent has 8 direct neighbours and therefore 8 potential communication partners of distance 1, he has 16 potential communication partners of distance 2, 24 potential communication partners of distance 3, et cetera. In general each agent has $8 \times d$ potential communication partners of distance d, where $1 \leq d \leq \lceil n/2 \rceil - 1$ for an $n \times n$ toroid lattice. Now assume that the probability to choose distance d is given by the function $P_\gamma(d) = \frac{8 \times d / d^\gamma}{\eta}$, where η is a normalizer[6] and γ represents the degree of locality. For example if γ is 0, agents choose their partners completely randomly, but if γ goes to ∞, the agents communicate only with their direct neighbourhood.

Figure 4 shows the probability distributions for different γ-values. As you can see, for $\gamma = 0$ the probability of choosing a distance d increases linearly with the distance, and because the number of agents with distance d increases linearly with the distance, each possible communication partner is chosen with the same probability independently of the distance. In other words, this choice behaviour depicts random communication. For $\gamma = 8$ the probability of choosing a communication partner with distance 1 is $P_8(1) \approx 0.992$ for a maximal distance of 10. Thus for $\gamma = 8$ the probability is almost 1 that agents choose a direct neighbour as a partner. This choice behaviour therefore is close to neighbourhood communication and approximates it by increasing γ.

Of interest here is the overall system's behaviour for different γ-values. As I already mentioned, the behaviour of choosing a completely random partner or choosing one in an agent's direct neighbourhood are extreme points for the degree of locality. I want to know how the degree of locality affects the emergence of signaling language(s) in the

[6] I set $\eta = \sum_{d=1}^{\lceil n/2 \rceil - 1} 8 \times d/d^\gamma$ to guarantee that $P_\gamma(d)$ is a probability measure, in other words that holds: $\forall d \in D : 0 \leq P_\gamma(d) \leq 1$ and $\sum_{d \in D} P_\gamma(d) = 1$ with $D = \{1, 2, \ldots \lceil n/2 \rceil - 1\} \subseteq \mathbb{N}$

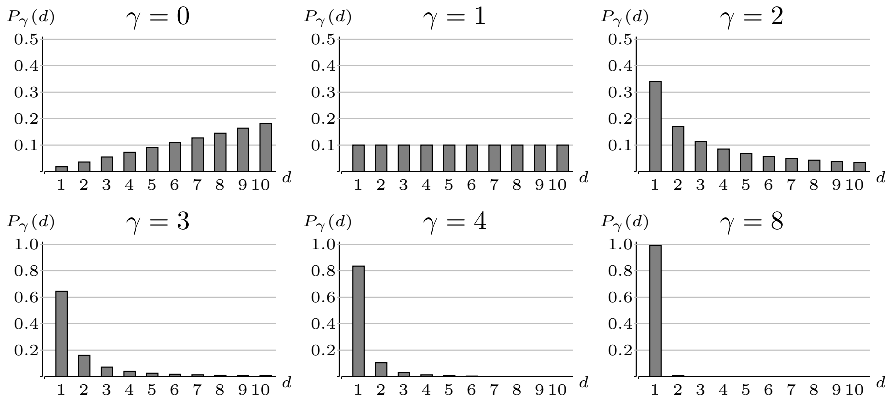


Figure 4. Degree distributions with a maximal degree of 10 for different γ-values. $\gamma = 0$ depicts random communication. By increasing the γ-value, agents' behaviour approximates neighbourhood communication. For $\gamma = 8$ the probability to choose a direct neighbour is ca. 0.992.

multi-agent system; in other words, I want to examine the society's behaviour in between those two extreme points.

5. Simulations and results

The agents in my model are placed on an $n \times n$ toroid lattice. Communicating via signaling games, they act as both sender and receiver while using the aforementioned response and update rules. A simulation run consists of consecutively executed communication steps. In one communication step each agent Z_i performs the following substeps:

1. Z_i chooses a distance d with probability $P_\gamma(d)$ and then a random partner Z_j of distance d
2. Z_i is assigned with a random state t according to probability $Pr(t)$
3. Z_i sends message m determined by $\rho_s(t, m)$ to partner Z_j
4. Z_j construes message m with interpretation a, determined by $\rho_r(m, a)$
5. Z_i and Z_j observe the result of the communication and update their respective beliefs or urns

It is possible to compute the agreement between an agent's current strategy and a pure strategy and therefore a language L_x via a measure known as the *Hellinger similarity*[7]. The current strategy of an RL agent is identical to his urn's setting, whereas the current strategy of a BL agent is uniquely defined by his beliefs. I used the Hellinger similarity to describe how close the current strategy is to one of the 16 languages depicted in Figure 3. The Hellinger similarity can range between 0 and 1, where 1 is given for identical strategies. If the Hellinger similarity between the current strategy and one of the languages is above a given threshold, I declared the agent's strategy to be close enough to be called the appropriate language; i.e. I allege that the agent has learned or is using this language, otherwise I assert that he hasn't learned a language yet.[8] Further conventions for the simulations are:

- Unless further noted, all simulations are performed with 400 agents placed on a 20×20 toroid lattice.
- Without further remarks, the memory size for agents with bounded memory is 100.
- BL agents start with an empty set of experiences, RL agents start with 100 balls of each type per urn.
- The experiments include simulations for the Lewis game and for the two predefined DOPL games. By describing agents' behaviour with particular languages, the appropriate names of these languages are depicted in both tables of Figure 3.
- By saying an agent is playing language L_x or that he is an x-player, I underscore that the Hellinger similarity of his current strategy is close enough to the appropriate language x. In the experiments the Hellinger similarity to the appropriate language must be higher than the threshold value of 0.8
- In general, simulation runs stopped when every agent learned one of the languages. Some cases required runs of up to 2,000 steps before a convention became stable. Longer simulations showed that simulations beyond 2,000 steps were stable enough so as not to require further iterations.

Table I gives an overview for the following experiments. There are four *Part A* experiments for all four possible combinations of player type and memory type. In each experiment I measured the influence

[7] The Hellinger distance is a standard measure for distances between probability distributions P and Q over the same set X, defined as $H(P,Q) = \sqrt{1 - \sum_{x \in X} \sqrt{P(x) \times Q(x)}}$. Hellinger similarity is defined as $1 - H(P,Q)$.

[8] The same way of proceeding was used by Franke & Jäger (to appear).

Table I. Overview for the different experiments

	Part A	**Part B**
Experiment 1: BL agents & unbounded memory	- different γ-values	- different popultation size
Experiment 2: RL agents & unbounded memory	- different γ-values	
Experiment 3: RL agents & bounded memory	- different γ-values	
Experiment 4: BL agents & bounded memory	- different γ-values	- different memory size

of the degree of locality γ on the resulting society. Furthermore for BL agents with unbounded memory I measured the influence of population size and for BL agents with bounded memory I measured the influence of memory size.

5.1. The influence of the social structure

In the first experiment 1A, I examined the influence of the degree of locality γ for BL agents with unbounded memory. First, I let them play the standard Lewis game to see how γ influences the emergence of multiple equilibria by starting simulations with different γ-values. The basic result is that each simulation run ended up either in a society where every agent learned the same signaling language or where the lattice is split into local signaling languages of both types. The runs resulting in the split societies produced regions separated by borders of players using no language, as depicted in the left picture of Figure 5. The phenomenon of such border players is a consequence of the fact that both signaling languages are highly incompatible.[9] Thus border players never learn a language, but rather switch between different strategies, torn between both signaling languages. Nevertheless, whenever one unique society-wide signaling language emerged, all agents learned it, and whenever multiple local signaling languages emerged, only the border players failed to learn one of them. These basic results spurred the examination of how the degree of locality impacts the probability of multiple local signaling languages emerging.

[9] As comprehensible by taking a look at Figure 2, the expected utility of a combination of opposed signaling strategies is 0 for the Lewis game $((S_1, R_2), (S_2, R_1))$ and negative for a DOPL game $((S_H, R_{aH}), (S_{aH}, R_H))$.

For different γ-values between 0 and 9.5 Figure 6 shows the percentage of 25 trials producing a society with multiple local signaling languages. The results indicate that the probability for the emergence of multiple languages depends on the degree of locality γ. Remember, with $\gamma = 0$ we have random communication and expect only one global signaling language L_1 or L_2 to emerge, whereas for a high γ-value we're close to neighbourhood communication and expect multiple local signaling languages L_1 and L_2 to emerge. The result for the Lewis game in Figure 6 shows that for $\gamma < 2$, every trial resulted in a society with only one global signaling language, but for $\gamma \geq 3$, every trial led to a society with both signaling languages. In the range $2 \leq \gamma < 3$, the percentage of trials ending with two signaling languages increased with γ. All in all, these trials show that the probability of multiple signaling languages emerging increases with respect to the degree of locality γ.

I used the same procedure for the weak DOPL game. The basic takeaway was that either the whole society learned the Horn language as a unique convention or that the society split into Horn and anti-Horn players, again divided by border players. As you can see in Figure 6, for $\gamma \leq 3$ all of the trials eventuated with only Horn players, whereas for $\gamma \geq 4$ the percentage of trials that ended with a society of both Horn and anti-Horn players was around 80%. In the range $3 < \gamma < 4$, the percentage of trials that rendered a society with anti-Horn players as well increased with the γ-value. For each trial resulting in a society with both Horn and anti-Horn players, the number of anti-Horn players was always small, on average around 30 players. An exemplary pattern of such a final society is depicted in the right picture of Figure 5.

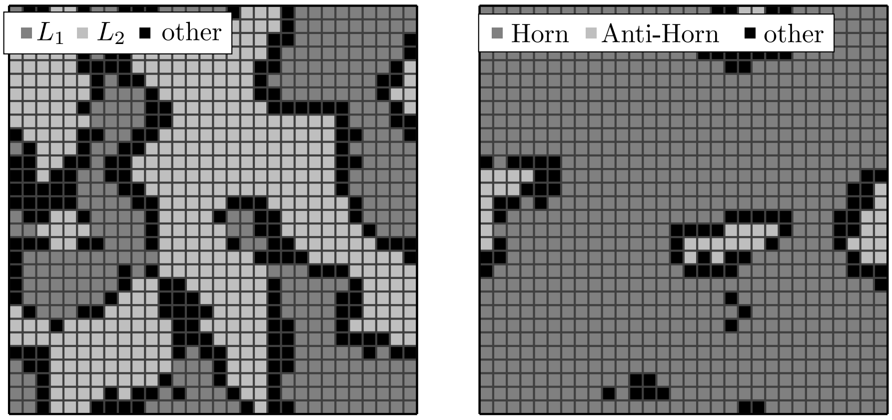


Figure 5. Exemplary distribution of players with different strategies. *Left picture:* Result of a Lewis game. *Right picture:* Result of a weak DOPL game.

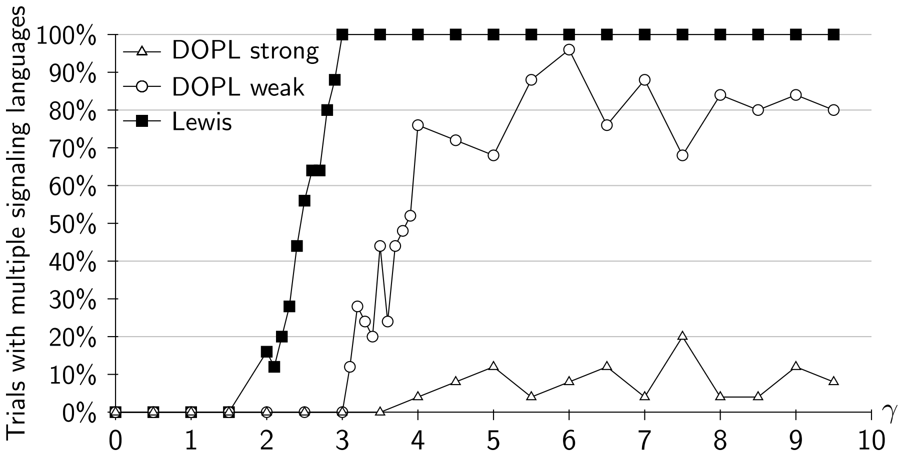


Figure 6. The percentage of trials resulting in a society with multiple signaling languages. For the Lewis game, every trial culminated with both signaling languages if $\gamma \geq 3$. For the weak DOPL game, about 80% of the trials produced both Horn and anti-Horn players if $\gamma \geq 4$. And for the strong DOPL game less than 20% of the trials resulted with both Horn and anti-Horn players.

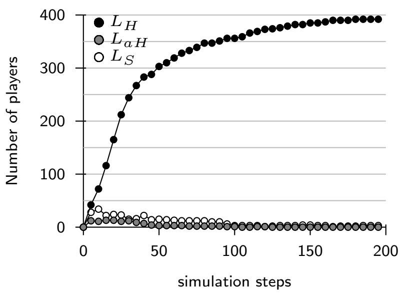
Finally, I used the same procedure for the strong DOPL game. As you can see the higher difference of the probabilities $Pr(t_p)$ and $Pr(t_r)$ had a deep impact on the emergence of anti-Horn players. Here there was no emergence of anti-Horn players for $\gamma < 4$. Even for $\gamma \geq 4$ the percentage of trials that finished with both Horn and anti-Horn players was only around 10%; furthermore, the average size of those local anti-Horn player groups was around 10 players. For the strong DOPL game, the emergence of anti-Horn players was highly improbable, and if they did arise, the group was minuscule.

Important to note is that the degree of locality γ was not the only factor influencing the probability that societies resulting in Horn *and* anti-Horn players would emerge. The overall number of agents also influenced this probability. The left picture of Figure 7 shows an evolution course of 200 steps of a trial for the strong DOPL game where no stable group of anti-Horn players emerged in the end. Observe that at the beginning of the trial such a group emerged but was driven to early extinction because it was not large enough.

For lattices with a higher number of agents, the probability is obviously higher that emerging groups of anti-Horn players are large enough to survive, flourish, and stabilize. To examine this fact I started experiment 1B, which simulated different lattices sizes of 5×5, 10×10, 15×15, 20×20, 25×25 and 30×30 agents over 15 trials in each case (fixed $\gamma = 9$). I then compared the probabilities of societies where anti-Horn players would emerge. As you can see in the right picture of Figure 7 for the weak DOPL game, the following holds: the probability of anti-Horn players emerging increases with the number of agents. For the strong DOPL game with $n \leq 15$, anti-Horn players do not emerge. This only occurs in around 10% of all trials if $n \geq 20$. While this result is still tentative, further investigation should give a more detailed analysis of these dependencies. Nevertheless, we see that the probability that stable groups of anti-Horn players emerge generally increases with the number of agents.

Now let's wrap up the results of experiment 1. For a multi-agent system with BL agents playing a DOPL game, the following holds: the higher the degree of locality γ and/or the higher the number of agents on the map, the higher the probability of the emergence of stable groups of anti-Horn players. Regardless of this, however, this probability is still very low in the strong DOPL game. As you can imagine, changing the game settings to still stronger values (higher distance of probability values and/or higher difference of message costs) will force the probability of those groups emerging to zero. Now let's take a look at how the degree of locality influences another species of agents, the much more unsophisticated RL agents.

In experiment 2, I used the same settings as in experiment 1A, but replaced BL agents by RL agents. A basic result is that the societies of RL agents are less uniform than those of BL agents. As evident in the line of experiment 1, BL agents have a strong drive for using exactly one signaling language for partners with whom they frequently communicate. Ergo, the resulting society always either uses one unique signaling language or the whole lattice is split into groups of both



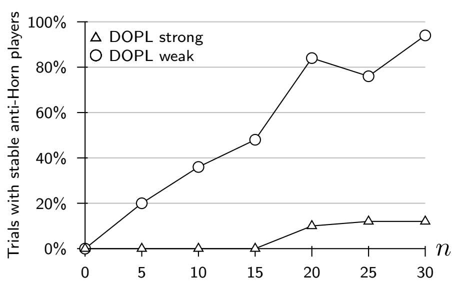


Figure 7. Left picture: A data plot for a strong DOPL game on a 20×20 lattice. Anti-Horn players emerge temporary but are driven to extinction because their group is too small. *Right picture:* Statistics for different lattice sizes with $\gamma = 9.0$. The emergence of stable groups of anti-Horn players also depends on the overall number of agents. By increasing n for $n \times n$ lattices, the percentage of trials ending up with anti-Horn players emerging also increases in the weak DOPL game.

signaling languages, where only agents on the borders of those groups cannot stabilize. Of such agents, we could say that *they fell between the cracks.* As we will see in my final conclusion this drive is a result of their high flexibility. In contrast the RL agents are much more inert after an initial phase, as they zero in on a language that is not necessarily a signaling language.

The results for RL agents playing the Lewis game contrasts to the BL agents' results in the following way: first, agents learned both signaling languages L_1 and L_2 in each trial independently of the γ-value. Second, only a fraction of the agents learned signaling languages. The left picture of Figure 8 shows a resulting pattern for 400 RL agents with $\gamma = 7.0$. Here the degree of locality γ has no influence of the fact that only one or both signaling languages emerge, but it does influence the number of agents who learn a signaling language. The right picture of Figure 8 shows the number of agents (averaged over 15 trials) who learned a signaling or a pooling language. As you can see for $0 \leq \gamma \leq 4$, the number of agents who learned a signaling language increases with the γ-value and for $\gamma > 4$ the number levels off at around 180 signaling players. In fact, signaling players stabilize in groups, and the higher the degree of locality γ, the stronger those groups grow. But because the RL agents become more inert as the simulation runs longer, the growing eventually stops, and most of the agents stabilize at a learned pooling language or have learned no language at all. Even for a high γ-value less than half of the agents learned a signaling language.

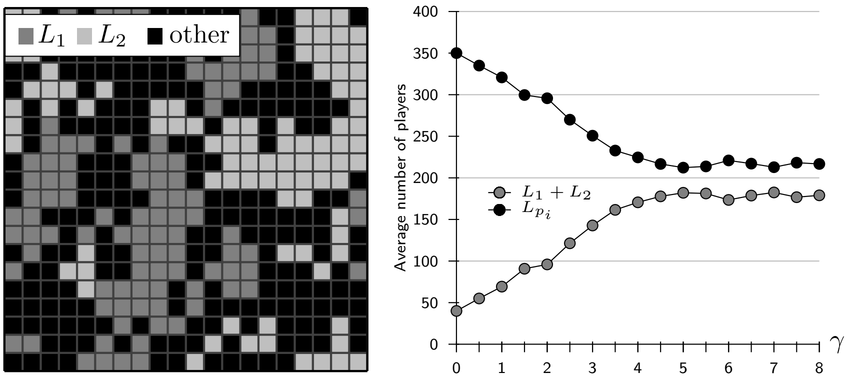
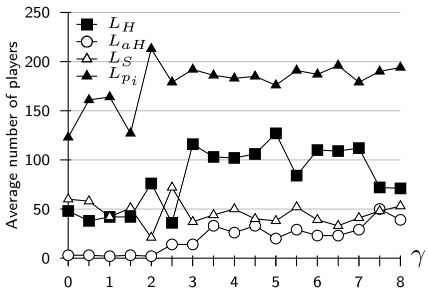


Figure 8. Left picture: Exemplary stable distribution of RL players learned a signaling language L_1, L_2 or none of them on a 20×20 toroid lattice playing the Lewis game with $\gamma = 7.0$. *Right picture:* The average number of signaling or pooling language players over 15 trials each for different γ-values.

Now let's take a look at the results for RL agents playing a DOPL game. The resulting society of agents playing the weak DOPL game simulated for different γ-values is depicted in the left picture of Figure 9. Each data point represents the average number over 15 trials of language player types for an appropriate γ-value. For $0 \leq \gamma < 2$ around 150 pooling players, 50 Horn players, 50 Smolensky players and only a few anti-Horn players emerged. For higher γ-values, those numbers increased. For $\gamma > 3$, stable groups of around 190 pooling players, 100 Horn players, 40 anti-Horn players emerged. Thus the number of Smolensky players stays at 50 players independently of the γ-value, whereas γ influences the number of learners of the other languages. By taking a look at the overall number of agents who learned one of the target languages of Table 3, we see that for $0 \leq \gamma \leq 2.5$ around 310 agents learned a target language whereas for $\gamma > 2.5$ around 380 agents learned one of them. But in each case the number of Horn and anti-Horn players in sum accounts for less than half of all agents.

Finally, the results for RL agents playing the strong DOPL game are depicted in the right picture of Figure 9. As you can see for $\gamma \leq 2$ no agent learned any target language. With increasing the γ-value, the number of Horn, anti-Horn, Smolensky and pooling players increased. For $\gamma \geq 2$, the number of Horn players stabilized around 40, anti-Horn less than 5, and Smolensky around 10. For $\gamma \geq 5$ the number of pooling players stabilized around 125. By summing this up, the overall number of agents that learned a target language is less than half of all agents even for high γ-values. Further, the emergence of anti-Horn players is probable, but those groups are minute.



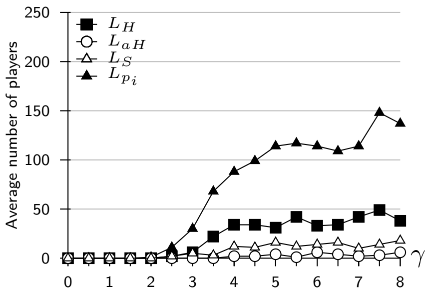


Figure 9. The average number of different player types over 10 trials each for different γ-values. *Left picture:* The result for the weak DOPL game. *Right picture:* The result for the strong DOPL game.

All in all, the difference between BL and RL agents is the following: RL agents don't all learn a signaling language; in fact they don't all learn any of the predefined languages. But nevertheless by increasing the γ-value, both the number of agents who learned any language and the number of agents who learned a signaling language increased. It is important to recognize that because of being less dynamic, even for high

γ-values, less then half of all RL agents learned a signaling language for the Lewis and weak DOPL game, and even less then 50 for the strong DOPL game. In contrast, except for the agents on the border between two signaling regions, all BL agents learned a signaling language. Here the γ-value influences the probability of the emergence of only one or both signaling languages. Ergo with respect to the DOPL game, the γ-value influences the probability for anti-Horn players to emerge. In the next subsection, we examine how a limited memory size changes the resulting structures.

5.2. The influence of bounded memory

In experiment 3, I examined the influence of bounded memory on the behaviour of RL agents. I started simulations for agents with a memory size of 100 for different γ-values. The overall result is that the bounded memory makes the agents less inert. The left picture of Figure 10 shows the results. Each data point displays the percentage of both signaling languages emerging for the Lewis game, the weak DOPL game, and the strong DOPL game. This result is comparable, not to say conspicuously similar to the result of experiment 1 for BL agents with unbounded memory. As you can see for the Lewis game for $\gamma < 2$ the whole society learned one of both signaling languages. For $2 \leq \gamma < 3$ the probability for the emergence of both signaling languages increased, and for $\gamma \geq 3$ the society split into stable groups of players who learned both signaling languages, with agents on the border learning no language.

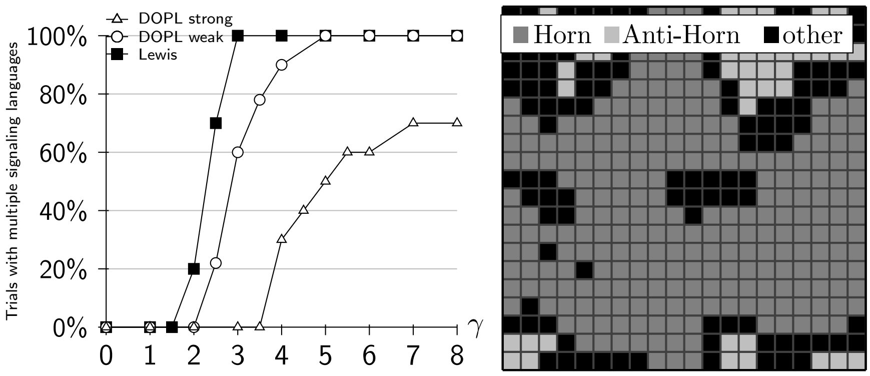


Figure 10. Left picture: The percentage of 15 trials ending up in a stable state with multiple signaling languages for RL agents with a bounded memory of 100. For high γ-values agents learn both signaling languages for the Lewis game, the weak and the strong DOPL game as well. *Right picture:* An exemplary distribution of different language players for a weak DOPL game and $\gamma = 7.0$. The whole society is distributed in Horn and anti-Horn players with boarder players in between.

For the weak DOPL game, $\gamma < 2.5$ induced all agents learning the Horn language, for $2.5 \leq \gamma \leq 4$ the probability of the emergence of stable groups of anti-Horn players increases with the γ-value, and for $\gamma > 4$ the society is split in Horn players and a smaller group of anti-Horn players, both separated by a border of agents that learned no language. For the strong DOPL game with $\gamma < 4$, all agents learned the Horn language, and for $4 \leq \gamma < 7$ the probability for the emergence of anti-Horn players increases with γ. For $\gamma \geq 7$ the probability for the emergence of stable groups of anti-Horn players levelled off around 70%.

These results show that by increasing the γ-value the probability for the emergence of multiple signaling languages increases. The right picture of Figure 10 shows a exemplary distribution of agents playing the weak DOPL game with a memory of 100 for $\gamma = 7.0$. This picture also looks remarkably similar to the resulting pattern of BL agents: every agent learned a signaling language, Horn or anti-Horn, and only the agents on the border learned no signaling language. This result leads me to the following proposition:

> *The systemic behaviour of RL agents with bounded memory is similar to the systemic behaviour of BL agents with unbounded memory.*

In experiment 4A I examined BL agents with bounded memory. With this setting each trial resulted in a society with only one signaling language independent of the γ-value. In the DOPL games all agents learned the Horn language at the end of a simulation run. But by taking a closer look at the simulation runs you can see that in a lot of trials substantial islands of anti-Horn players emerged *during* a simulation run. The left picture of Figure 11 shows an example for such a trial with BR agents with a bounded memory of size 400 for $\gamma = 5.0$. Here a group of more than 60 anti-Horn players emerged and even if it takes more then 2000 steps, this group was driven to extinction.

The same happened in case of the Lewis game. During a trial, islands of both signaling languages emerged. One of both language groups constituted i) the majority within the trial and ii) the one and only signaling language at the end of the trial after the language group constituting the minority was driven to extinction. Thus the systemic behaviour of BL agents with bounded memory is different to that of BL agents with unbounded memory, for whom islands of minorities stay stable, if they are big enough. BL agents with bounded memory always end up in a society with one unique signaling language.

In experiment 4B, I wanted to determine the influence of BL agents' memory size on the number of simulation steps required until a signaling language has captured the whole society. Thus I started simulations with BL agents with a memory size of 50, 100, 200 and 400 for the weak

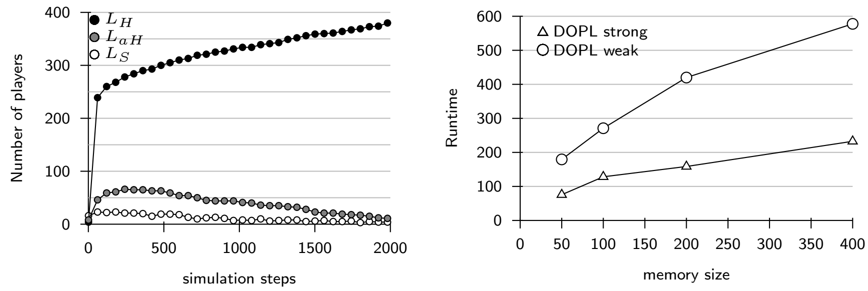


Figure 11. Left picture: A data plot for a strong DOPL game. Anti-Horn players emerge temporary and in a large group but are driven to extinction by surrounding Horn players. *Right picture:* The average runtime over 10 trials for a society-wide signaling language (here: Horn language) to emerge; for different memory sizes with $\gamma = 5.0$. Result: the smaller the memory size, the shorter the runtime.

and the strong DOPL game and measured the average runtime over ten trials each. As you can see in the right picture of Figure 11, the runtime for the whole society to learn the unique signaling language increases linearly with the memory size.

Finally Table II pictures a summary of all experiments' results.

Table II. Summary: Results of the different experiments

	Part A	**Part B**
Exp. 1: BL agents & unbounded memory	- all agents learn a signaling language (exc. on borders) - increasing $\gamma \to$ probability for multiple signaling languages increases	- all agents learn a signaling language - increasing population size $\to$ probability for multiple signaling languages increases
Exp. 2: RL agents & unbounded memory	- less than half of all agents learn a signaling language - increasing $\gamma \to$ number of signaling players increases	
Exp. 3: RL agents & bounded memory	- all agents learn a signaling language (exc. on borders) - increasing $\gamma \to$ probability for multiple signaling languages increases	
Exp. 4: BL agents & bounded memory	- all agents learn the same unique signaling language (L_H in a DOPL game)	- increasing memory size $\to$ trials' runtime increases

6. Conclusion

The *Part A* experiments compared four different types of agents, arising out of the combinations of learning dynamics (RL and BL) and memory settings (unbounded and bounded). By comparing these results I am inclined to introduce a property that signifies the distinction of the systemic behaviour of each agent type. I call this property *flexibility.* The simulation results suggest a classification of three different levels of flexibility. Level 0 is extremely inert behaviour like those of RL agents with unbounded memory: most of the agents learn no signaling language because their behaviour is not flexible enough for a successful language to spread society-wide. Level 1 is much more flexible. Here one global signaling language or local groups of signaling languages emerge and stay stable, where only the border players between local groups fail to learn a language. This behaviour is observable for BL agents with unbounded memory as well as RL agents with bounded memory. Level 2 is the most flexible case, so that convex regions cannot stay stable and are driven to extinction. That means that in general only one global signaling language will emerge for the whole society. This holds for BL agents with bounded memory. Table III depicts the three different levels of flexibility.

Table III. Three different levels of flexibility

	Agent type	**Resulting society**
Flexibility Level 0	- RL & unbounded memory	No emergence of society-wide signaling language(s)
Flexibility Level 1	- RL & bounded memory - BL & unbounded memory	*Local communication:* emergence of several local signaling languages *Global communication:* emergence of one society-wide unique and efficient signaling language
Flexibility Level 2	- BL & bounded memory	emergence of one society-wide unique and efficient signaling language

It was also shown in experiment 4B that, with respect to a lower limit, the shorter the memory size, the more flexible the agents and the faster a stable state emerged. This result covers the previous results, that agents with unbounded memory are much less flexible. The three general conclusions drawn from the simulation results are the following:

1. BL agents are in general much more flexible than RL agents.
2. The degree of locality γ influences the agents' behaviour in the following way: the higher the γ-value, the stronger the tendency for agents to form local signaling languages.
3. The memory size influences the agents behaviour in the following way: the shorter the memory, the more flexible the behaviour of the agents.

Thus we can answer the first question: *What circumstances could be responsible for an evolutionary process that doesn't lead to the expected DOPL convention?* The answer is that a local communication structure (high γ-value) and a flexibility level < 2 leads to resulting structures including more than the expected Horn players. *To which extend is rationality necessary for the emergence of a society-wide convention à la DOPL?* The results showed that rationality plays a role in that sense that it influences flexibility. They also reveal that the higher the rationality, the higher the flexibility, all else being equal. But agent flexibility can also be influenced by interior conditions like memory size or by external circumstances like the degree of locality. Thus rationality is not the bottom line factor for such an emergence: e.g. for the more rational BL agents there is no emergence of one unique signaling language if their memory is unbounded and the communication structure is local, whereas for the less rational RL agents such a unique one would emerge, if their memory is bounded and the communication structure is global.

References

Barrett, J., Zollman, K.J.S.: 2009, The role of forgetting in the evolution and learning of language, *Journal of Experimental and Theoretical Artificial Intelligence 21(4)*, pp. 293–309.

Benz, A., Jäger, G., van Rooy, R.: 2005, An introduction to game theory for linguists, *Game Theory and Pragmatics*, pp. 1–82.

Blutner, R., Zeevat, H.: to appear, Optimality-Theoretic Pragmatics, *Semantics: An International Handbook of Natural Language Meaning*.

Brown, G.W.: 1951, Iterative Solutions of Games by Fictitious Play, *Activity Analysis of Production and Allocation*.

Foster, D., Young, H.P.: 1990, Stochastic Evolutionary Game Dynamics, *Theoretical Population Biology 38*, pp. 219–232.

Franke, M., Jäger, G.: to appear, Bidirectional Optimization from Reasoning and Learning in Games, *Journal of Logic, Language and Information*.

Horn, L.: 1984, Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature, *Form, and use in Context: Linguistic Applications*, pp. 11–42, Washington: Georgetown University Press.

Huttegger, S.M., Zollman, K.J.S.: 2011, Signaling games: dynamics of evolution and learning, *Language, Games, and Evolution*, Berlin: Springer.

Jäger, G.: 2008, Evolutionary stability conditions for signaling games with costly signals, *Journal of Theoretical Biology 253* **Vol. 1**, pp. 131–141.

Jäger, G., Ebert, C.: 2009, Pragmatic Rationalizibility, *Proceedings of Sinn and Bedeutung 13, SinSpeC. Working Papers of the SFB 732* **Vol. 5**, pp. 1–15

Levinson, S.C.: 2000, *Presumptive meanings: The theory of generalized conversational implicature*, MIT Press.

Lewis, D.: 1969, *Convention*, Cambridge: Harvard University Press.

Roth, A.E., Erev, I.: 1995, Learning in extensive form games: Experimental data and simple dynamical models in the intermediate term, *Games and Economic Behavior* **Vol. 8(1)**, pp. 164-212

Parikh, P.: 1991, Communication and strategic inference, *Linguistics and Philosophy* **Vol. 14**, pp. 473–513

Schaden, G.: 2008, Say Hello to Markedness, *Proceedings of DEAL II*, LiP 28, pp. 73–97.

Skyrms, B.: 2009, Evolution of signalling systems with multiple senders and receivers, *Philosophical Transactions of the Royal Society of London B* **Vol. 364**, pp. 771–779.

Skyrms, B.: 2010, *Signals: Evolution, Learning and Information*, Oxford University Press.

Taylor, P.D., Jonker, L.: 1978, Evolutionarily stable strategies and game dynamics, *Mathematical Biosciences* **Vol. 40**, pp. 145–156.

Tesar, B., Smolensky, P.: 1998, Learnability in Optimality Theory, *Linguistic Inquiry* **Vol. 29**, pp. 229–268.

van Rooij, R.: 2008, Evolutionary motivations for semantic universals, *Variation, Selection, Development – Probing the evolutionary model of language change*, pp. 103–142.

van Rooy, R.: 2004a, Signaling games select Horn strategies, *Linguistics and Philosophy* **Vol. 27(4)**, pp. 493–527.

van Rooy, R.: 2004b, Evolution of conventional meaning and conversational principles, *Synthese* **Vol. 139(2)**, pp. 331–366.

Wagner, E.: 2009, Communication and Structured Correlation, *Erkenntnis* **Vol. 71**, pp. 377–393.

Zollman, K.J.S.: 2005, Talking to neighbors: The evolution of regional meaning, *Philosophy of Science* **Vol. 72**, pp. 69–85.