Language games meet multi-agent reinforcement learning: A case study for the naming game

Van Eecke, Paul; Beuls, Katrien; Botoko Ekila, Jérôme; Radulescu, Roxana

Published in:
Journal of Language Evolution

DOI:
https://doi.org/10.1093/jole/lzad001

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Journal of Language Evolution, 2023, **XX**, 1–11
https://doi.org/10.1093/jole/lzad001
Advance access publication 18 April 2023
Methodology

Language games meet multi-agent reinforcement learning: A case study for the naming game

Paul Van Eecke[1,2,3,*,†], Katrien Beuls[4,†], Jérôme Botoko Ekila[1], and Roxana Rădulescu[1]

[1]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2]KU Leuven, Faculty of Arts, Blijde Inkomststraat 21, 3000 Leuven, Belgium
[3]KU Leuven, imec research group itec, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium
[4]Faculté d'informatique, Université de Namur, rue Grandgagnage 21, 5000 Namur, Belgium

[*]Corresponding author: paul@ai.vub.ac.be
†Joint first authors

Today, computational models of emergent communication in populations of autonomous agents are studied through two main methodological paradigms: multi-agent reinforcement learning (MARL) and the language game paradigm. While both paradigms share their main objectives and employ strikingly similar methods, the interaction between both communities has so far been surprisingly limited. This can to a large extent be ascribed to the use of different terminologies and experimental designs, which sometimes hinder the detection and interpretation of one another's results and progress. Through this paper, we aim to remedy this situation by (1) formulating the challenge of re-conceptualising the language game experimental paradigm in the framework of MARL, and by (2) providing both an alignment between their terminologies and an MARL–based reformulation of the canonical naming game experiment. Tackling this challenge will enable future language game experiments to benefit from the rapid and promising methodological advances in the MARL community, while it will enable future MARL experiments on learning emergent communication to benefit from the insights and results gained through language game experiments. We strongly believe that this cross-pollination has the potential to lead to major breakthroughs in the modelling of how human-like languages can emerge and evolve in multi-agent systems.

Keywords: language games; multi-agent reinforcement learning; computational modelling; naming game; emergent communication.

1. Introduction

The computational modelling of emergent communication in multi-agent systems is a topic of great interest to the artificial intelligence (AI) community, as achieving robust, flexible, and adaptive agent–agent and human–agent communication forms a key precondition for building truly intelligent systems (Mikolov *et al.*, 2016). Multi-agent reinforcement learning (MARL) forms a natural framework for learning emergent communication, given its adequacy to model, to a large extent, the conditions under which human languages emerge and evolve. The MARL framework has as a consequence been adopted in a number of influential papers on emergent communication, tackling a wide variety of tasks, including visual question answering (Das *et al.*, 2017), solving puzzles (Foerster *et al.*, 2016), negotiation (Cao *et al.*, 2018), reference (Lazaridou *et al.*, 2017), navigation (Sukhbaatar *et al.*, 2016; Bogin *et al.*, 2018; Mordatch and Abbeel, 2018), and coordination in self-driving cars (Resnick *et al.*, 2018). The focus of these experiments is typically on learning emergent languages that are effective at solving the task at hand, which explains that the experimental conditions widely vary and sometimes seem far removed from how human languages have emerged and continue to evolve. While this is not a problem in itself, it has important repercussions on the linguistic systems that emerge, and on the operational deployability of the models. For example, populations are almost always divided into speaker agents and listener agents, with speaker agents not being able to understand the language they learn to speak and listener agents not being able to speak the language they learn to understand.

Outside the MARL community, emergent communication is most prominently studied using the language game experimental paradigm (Steels, 1995, 2012a; Beuls and Steels, 2013).[1] One of the key defining properties of this paradigm is that the circumstances under which emergent communication is modelled resemble as much as possible those under which human languages emerge. Yet, the focus often lies on the emergence and evolution of particular linguistic phenomena and the tasks that are tackled are mostly limited to reference to objects in a scene (Beuls and Höfer, 2011; Spranger and Steels, 2015). The circumstances under which emergent communication is modelled within the language game paradigm, and of which we would argue that many are in line with the MARL framework and none are fundamentally incompatible, include the following:

- Languages emerge and evolve in a multi-agent setting, namely in a population of agents that participate in situated communicative interactions.
- Agents are autonomous[2] and communicate through language. They possess no mind-reading capabilities[3].
- Communicative interactions are local and learning is decentralised. Only those agents that participate in an interaction can exploit its outcome for learning.
- Communicative interactions are goal oriented. They serve a communicative purpose and can as such succeed or fail.
- The emerged languages are shaped by past successes and failures in communication.

While it is clear that the language game and MARL communities work on similar problems and share many of their objectives and conceptual foundations, the interaction between both communities has so far remained limited. This is to a large extent ascribable to the use of different terminologies and experimental designs, which often hinder the search for and interpretation of results achieved by the other community. This paper aims to remedy this situation by, on the one hand, formulating the challenge of re-conceptualising the language game paradigm in the framework of MARL, and, on the other hand, aligning the terms and concepts used by both communities. This terminological and conceptual alignment is put into practice through a case study in which the well-known naming game experiment (Steels, 1995) is reformulated in the framework of MARL. Apart from this terminological and conceptual alignment, the case study also introduces bidirectional dynamic Q-tables as a methodological innovation in the MARL framework. A full didactic implementation of the experiment accompanies this paper and is accessible at https://gitlab.ai.vub.ac.be/ehai/marl_language_games.

We strongly believe that cross-pollination between the language game paradigm and the MARL framework has the potential to lead to major breakthroughs in the modelling of how human-like languages can emerge and evolve in multi-agent systems. On the one hand, future language game experiments could benefit from the rapid and promising methodological advances in the MARL community, especially when it comes to dealing with complex, high-dimensional input data. Indeed, being able to handle input data that takes the form of raw images, videos or complex scenery would considerably enhance the application potential of the language game paradigm. On the other hand, future MARL experiments on emergent communication could benefit from the insights and results gained through language game experiments, thereby giving rise to emergent languages that exhibit the robustness, flexibility and adaptivity of human languages.

The remainder of this paper is structured as follows. Section 2 introduces the conceptual and methodological foundations underlying the language game paradigm. Section 3 provides a high-level discussion of the challenges involved in mapping the language game paradigm to the framework of MARL. Section 4 operationalises this mapping using a concrete case study in which the naming game experiment is implemented using the MARL framework. A concluding discussion is provided in Section 5.

2. The language game paradigm

The language game paradigm (see Steels, 2012b, for a brief introduction) embraces the view that human languages are evolutionary systems that emerge through the communicative interactions of language users, and are shaped by processes of variation and selection (Schleicher, 1863/1869; Darwin, 1871; J. M. Smith and Szathmáry, 2000; Oudeyer and Kaplan, 2007; Steels and Szathmáry, 2018). These processes take place within the linguistic system itself, on the level of concepts, words, grammar, and discourse, rather than in the genome of the language users (Steels, 2011; A. D. Smith, 2014). Sources of variation mainly stem from the creativity and problem-solving capabilities of language users, while the main selective pressures constitute communicative success and a reduction of processing effort (Grice, 1967; Echterhoff, 2013).

In terms of methodology, the language game paradigm employs multi-agent simulations for modelling the emergence and evolution of human-like languages. Such a simulation takes the form of a series of communicative interactions between autonomous agents in a population. A typical experiment proceeds as follows. At the beginning of each interaction, two agents

are selected from the population and are randomly assigned the role of either speaker or listener. The agents are placed in a particular scene and need to successfully communicate to solve a given task, which often consists in referring to objects or events that they observe in the scene. The agents are equipped with mechanisms for inventing and adopting linguistic means (e.g. words, concepts, or grammatical structures) that can be needed to achieve communicative success. After each interaction, the speaker provides feedback to the listener about the outcome of the task. This allows the listener to learn in the case that the agents did not reach communicative success. Additionally, both agents reward the linguistic means that were used in the case of a successful interaction, and punish these in the case of a failed interaction. As more and more games are played, the agents in the population gradually converge on a shared language. The language of each individual agent has been shaped by the communicative interactions it participated in and is, therefore, well adapted to the task and the environment.

During a communicative interaction, the speaker and listener go through the different processes depicted in Fig. 1. Both agents are situated in the same physical or simulated world, which they perceive through their sensori-motor system. The speaker maps its sensori-motor experiences to meaningful concepts and conceptual structures, abstracting away from the raw sensor values (*grounding and conceptualisation*). These conceptual structures are then expressed in the form of linguistic utterances (*production*). The listener perceives the utterances and uses its own linguistic system to reconstruct the conceptual structures underlying them (*comprehension*), which it then interprets with respect to the world (*grounding and interpretation*). Operationalising language game experiments requires the implementation of processing, invention, adoption, and alignment mechanisms for each of these processes, although one or more levels can be scaffolded in individual experiments.

The overall objective of language game experiments is to find adequate invention, adoption, and alignment mechanisms that allow a population of agents to self-organise a conceptual and linguistic system that allows them to communicate for successfully solving an open-ended set of tasks in an ever-changing environment.

3. Language games and MARL

Many of the central ideas that underlie the language game paradigm are also characteristic of the MARL framework as applied to experiments in emergent communication. First and foremost, both methodologies make use of multi-agent simulations to investigate how a population of agents can learn to communicate through task-based communicative interactions. Second, the main forces driving the dynamics of the simulation are the rewarding of the agent's language use in the case of a communicatively successful interaction and the punishing of its language use in the case of a failed interaction. Finally, the languages that emerge can be human languages that were learnt in a tutor–learner scenario or artificial languages that do not exist outside the simulation.

There are other aspects of the language game paradigm that are either more challenging to operationalise within the MARL framework, or that are often not addressed in MARL-based experiments on emergent communication. However, these aspects concern desirable properties of emergent communication experiments and have the potential to lead to the emergence of more human-like languages. These aspects include the following:

- Agents should be fully autonomous, in the sense that they make their own decisions and are not subject to any form of central control.

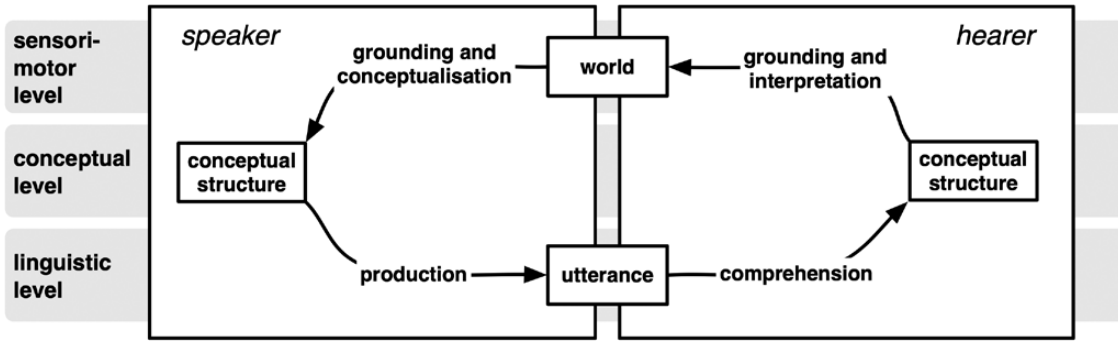

Figure 1 The semiotic cycle depicts the sensori-motor, conceptual, and linguistic processes that a language game involves for a speaker (left) and a listener (right).

They should not have mind-reading capabilities and interact only with the world and each other through their own sensors and actuators. This is necessary to ensure that the languages can emerge in populations of heterogeneous agents, which might not share the same hardware or software architectures (de Greeff and Belpaeme, 2011).
- The communicative interactions should be local and only accessible to the agents that participate. Consequently, this means that learning should be decentralised, so that the languages emerge through self-organisation (i.e. a global system arising from purely local interactions). Such decentralised, self-organising systems are known to be able to self-repair substantial perturbations, a form of robustness that is necessary for modelling the emergence and evolution of truly human-like languages (Heylighen *et al.*, 2001; Pfeifer *et al.*, 2007).
- The agents in the population should be able to take up the roles of both speaker and listener and their comprehension and production processes should be integrated. The agents should be able to express the concepts, words, and grammatical structures that they have learned in the listener role, and be able to understand the utterances that they have produced in the speaker role (Pickering and Garrod, 2013; Van Eecke, 2015).
- The emergent languages should be flexible and adaptive to changes in the tasks and environment of the agents. It should be avoided at all costs that a substantially different language needs to emerge when minor changes in the tasks and environment occur (Steels, 2000; Cornudella Gaya *et al.*, 2016).
- The linguistic inventories that contain representations of concepts, words, and grammatical structures should be dynamically expandable, so that new words, concepts, and grammatical structures can be introduced should the need arise (de Boer, 2001; Goldberg, 2019; Hoffmann, 2019).

In sum, it is clear that on a high level, the main conceptual and methodological foundations of the language game paradigm and the framework of MARL are very much in line with each other. A precise terminological and conceptual alignment between individual aspects of the two paradigms, including the more challenging aspects listed above, is put into practice in the next section through a case study in which the most canonical language game experiment is implemented in the framework of MARL.

4. Case study

This case study re-conceptualises the canonical naming game experiment in the framework of MARL. The naming game experiment, as initially introduced by Steels (1995) and Steels and Kaplan (1998), was a foundational experiment on the emergence and evolution of language. The naming game experiment consists of a population of autonomous agents participating in pairwise communicative interactions, called games, and gradually establishing a naming convention for referring to the objects in their environment. The naming game has attracted a great deal of attention in different areas of research, including AI (see e.g. Lenaerts *et al.*, 2005; de Vylder and Tuyls, 2006), linguistics (see e.g. van Trijp, 2013; Rădulescu and Beuls, 2016; Lipowska and Lipowski, 2022), semiotics (see e.g. Vogt, 2003), and statistical physics (see e.g. Loreto *et al.*, 2011). It still serves today as the basis of the language game paradigm and is therefore the logical starting point for taking on the challenge of bridging the gap between the language game paradigm and the MARL framework. A didactic Python implementation of the MARL-based naming game described below accompanies this paper and is accessible at https://gitlab.ai.vub.ac.be/ehai/marl_language_games.

4.1 The canonical naming game

A canonical naming game is defined by the following properties:

WorldThere exists a world $W = \{o_1,..., o_i\}$ that is a set of i objects.

PopulationThere exists a population $P = \{a_1,..., a_j\}$ that is a set of j agents. Each agent $a \in P$ is initialised with an empty vocabulary V_a.

VocabularyA vocabulary V of an agent $a \in P$, notated as V_a, is a potentially empty set of words, with each word of the vocabulary $w \in V$ being a coupling $w = (o, f, s)$ between an object $o \in W$, a form $f \in F$ and a score s. F is an infinite set of forms, typically enumerated through a regular expression.

ExperimentA naming game experiment $E = (W, P, G)$ is defined as a coupling between a world W, a population P and a sequence $G = (g_l{}^k{}_{l=1})$ of k games.

GameEach game $g \in G$ proceeds as follows:

1. **Context selection**A context $C = \{o_1,..., o_m\} \subseteq W$ consisting of a subset of m objects from the world is randomly selected.
2. **Agent and role selection**Two agents $a_1, a_2 \in P$ are randomly selected from the population. a_1 is assigned the role of speaker $S = a_1$, while a_2 is assigned the role of listener $L = a_2$.
3. **Topic selection**A topic object $T \in C$ is randomly selected from the context and is only disclosed to the speaker S. It is the task of S to draw the attention of the listener L to T using a word from the speaker's vocabulary $w \in V_S$.

4. **Production**If the speaker S knows a word $w = (T, f, s) \in V_S$, that is, for which the object matches the topic object T, S utters the form f as the utterance U, that is, $U = f$. If multiple such w's exist, the f of the w with the highest score s is uttered as U. If no such w exists:
   4a. **Invention**A new word $w = (T, f, s)$ is added to the speaker's vocabulary V_S, with f being randomly selected from the infinite set of forms F and s being assigned an initial value. Then, $U = f$.
5. **Comprehension**Then, the listener L searches for a word in its vocabulary $w = (o, U, s) \in V_L$, that is, where the form matches the utterance U. If such a w is found, L points to object o in context C. Otherwise, no pointing happens. As there is no noise in the canonical naming game and as new word forms are guaranteed to be unique (see Vocabulary above), it never occurs that L points to a wrong object.
6. **Feedback**If the listener L pointed indeed to the topic object T, the speaker S signals success. Otherwise, S signals failure and provides feedback by pointing to T.
7. **Alignment**If the game g was successful, the speaker S will increase the score s of the used word from its vocabulary $w = (T, U, s) \in V_S$ and the listener L will increase the s of the used word from its vocabulary $w = (T, U, s) \in \mathrm{V_L}$. At the same time, S will decrease the s of other, competing words $w = (T, f, s) \in V_S$ that refer to the topic object using a different form than U and L will decrease the s of other, competing words $w = (T, f, s) \in V_L$. If the game failed, S will decrease the s of the used $w = (T, U, s) \in V_S$, and L will decrease the score s of the used $w = (o, U, s) \in V_L$ (if L knew indeed a w with form U) and:
   7a. **Adoption**A new word $w = (T, U, s)$ is added to V_L, with s being assigned a fixed initial value.

The exact way in which the scores of words are increased or decreased is defined by an update rule. Here, we make use of the standard interpolation rule described by De Beule and Bergen (2006). This rule increases the scores of words according to $s \leftarrow \alpha + (1 - \alpha)s$ and decreases the scores of words according to $s \leftarrow (1 - \alpha)s$. s stands here for the score of the word and α stands for the learning rate. Using this update rule, the scores of the words are bounded between 0 and 1. 0.5 is typically chosen as the initial score of a new word.

When a naming game experiment is implemented as defined above and run in simulation, we can observe the typical language emergence dynamics (Blythe and Croft, 2012) that are depicted in Fig. 2. The experiment is run here with a population consisting of ten agents, a world consisting of ten objects, a context consisting of five objects and a learning rate (α) of 0.5. The solid green line shows on the left y-axis the degree of communicative success in the function of the number of games played. Communicative success is a binary measure that indicates whether a communicative interaction succeeded or failed. The dashed yellow line shows on the right y-axis the average number of words known by an individual agent. Words that have attained a score of under 0.01 are not counted. The dotted red line shows on the left y-axis the degree of lexical coherence. This is calculated as a binary measure that indicates whether both the speaker and the listener agent would use the same form to describe the topic object. In other terms, they achieve lexical coherence if the highest scored word referring to the topic object has the same form in the vocabularies of both agents. All results are averaged over ten independent experimental runs with error bars indicating a single standard deviation. Communicative success and lexical coherence are drawn using a sliding window of 100 interactions.

We can see in the figure that communicative success starts at 0. This is expected, as all agents start with an empty vocabulary and can, therefore, not yet successfully communicate with each other. It then gradually rises to 1 over the course of 1,500 interactions. At this point, every communicative interaction between two agents is successful. Like communicative success, the average lexicon size starts at 0. It then rises quite rapidly to over 25 after 450 interactions as new words are invented to suit the communicative needs of the agents. Often, many different words for the same object are invented independently by different agents during different communicative interactions and propagate in the population. As a consequence of the alignment process in which the scores of the words are updated, the average number of words per agent then starts to decline until it stabilises at 10 after about 2,500 interactions. This is the optimal vocabulary size, as it allows the agents to uniquely identify the ten objects in the world. The degree of lexical coherence also starts at 0 and gradually rises to 1 after 2,500 interactions. At this point, the agents cannot only successfully communicate, but they also use the same words to refer to the same objects.

The dynamics of the competition between different forms for the same object in a single agent and a single experimental run are shown in Figure 3. At the beginning of the experiment, the agent strongly associates the object with the form 'zojose' and also encounters the less successful form 'sekohi'. Later, the form 'xiyaje' takes over until it is overtaken itself by the form 'petevi'. After 1,500 interactions, the form 'petevi' has reached a score of 1 while the associations of other forms with the same object have reached a score of 0.

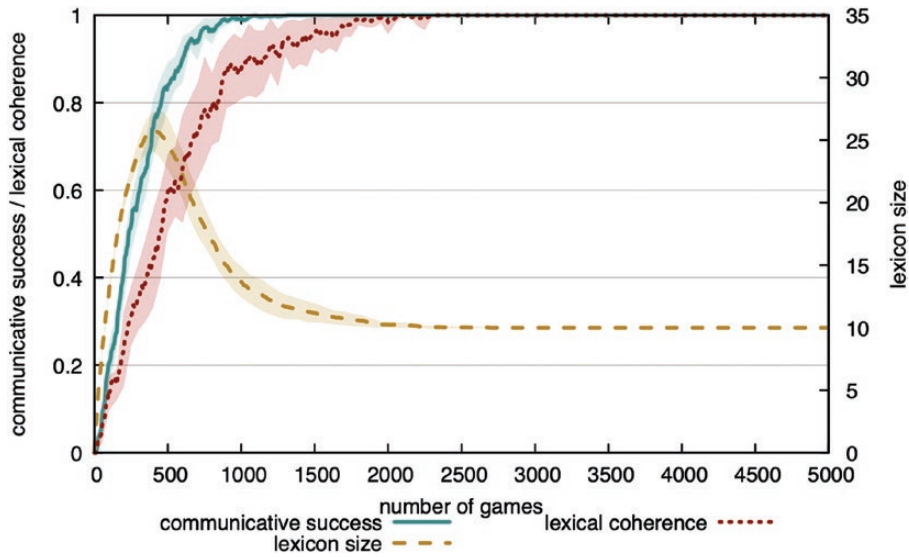

Figure 2 Dynamics of a canonical naming game experiment, with ten agents communicating about ten objects.

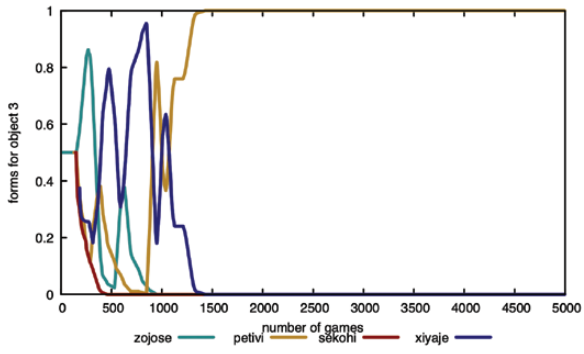

Figure 3 Competition of word forms over time for a single object in a single agent (canonical naming game).

From then on, this situation remains stable for the rest of the experiment.

Concerning the properties of emergent communication experiments listed in Section 2, we can conclude that the naming game models the emergence of a flexible and adaptive naming convention in a population of agents that are fully autonomous, in which interactions are local and decentralised, in which agents can serve the roles of both speaker and listener, and in which the vocabularies of the agents are dynamically expandable.

4.2 A MARL-based naming game

In the MARL framework, the naming game experiment can most straightforwardly be conceptualised as an independent Q-learning (IQL) problem (Tan, 1993) involving cooperative agents. In such a problem, every agent in the population treats the other agents as part of the environment and thereby autonomously learns its own action-selection policy.

Many aspects of the naming game experiment can straightforwardly be mapped to an IQL problem. Each game corresponds to an *episode*, in which two agents communicate with each other. These agents are randomly selected from the population and are randomly assigned the roles of speaker and listener. The *environment* consists of a number of objects, the discourse roles of the agents, the topic object, and the utterance. Additionally, for each agent, the other agents are also part of the environment. The environment is only *partially observable*. From the perspective of the speaker agent, its discourse role, the objects in the context, the listener agent, and the topic object are observable. From the perspective of the listener agent, its discourse role, the objects in the context, the speaker agent, and the utterance are observable. Importantly, only the actions and external appearance of the interlocutor are observable, not its knowledge or reasoning processes. Moreover, the other agents in the population do not have access to the interacting agents, the utterance, or the topic object. Each communicative interaction is thus local to the two interacting agents.

The *observation space* of an agent in the speaker role consists of all possible objects in the world while the observation space of an agent in the listener role consists of all possible utterances. Conversely, the *action space* of an agent in the speaker role consists of (uttering) all possible utterances while the action space of an agent in the listener role consists of (pointing to)

all possible objects. The observation space and action space of an agent can be open ended, as the set of possible objects and the set of possible utterances can be infinite. The environment of an agent is non-stationary, as it includes the other agents, which are also learning, and because of the open-ended nature of the set of possible objects in the world and the set of possible forms in the language. The *action-selection process* of the speaker and listener agent, which corresponds to language production and language comprehension, respectively, consists of mapping from an *observation*, that is, a topic object for the speaker and an utterance for the listener, to an *action*, that is, (uttering) an utterance for the speaker and (pointing to) an object for the listener. During each episode, the two agents that participate in the interaction perform thus a single action.

At the end of each episode, the two interacting agents receive a positive reward if the communicative interaction succeeded (as defined in step 6 of the formal definition of the naming game provided in Section 4.1) and a negative reward if the interaction failed. The reward is not shared with the other agents in the population.

4.3 Bidirectional dynamic Q-learning

Intuitively, the vocabularies of the agents can be represented using *Q-tables*. A Q-table represents a mapping between observations and actions, with the value of each observation–action pair corresponding to the expected reward if this mapping is used. Q-learning algorithms (Watkins, 1989) can be used to update the values of the mappings in the table.

The use of standard Q-tables to represent linguistic knowledge in emergent communication experiments comes with two important limitations. First of all, standard Q-tables fail to capture the bidirectional nature of linguistic knowledge. Second, the action space and observation space are fixed and known beforehand while the possible objects and utterances in human-like languages (and the naming game) are open-ended. In order to overcome this problem, we introduce the use of *bidirectional dynamic Q-tables*.

The action space and observation space of an agent depend on its discourse role. As a speaker, the observation space consists of all possible objects and the action space consists of all possible utterances. As a listener, the observation space consists of all possible utterances and the action space consists of all possible objects. Importantly, the mappings between utterances and objects are independent of the discourse role. In other terms, the association between an object and its name is bidirectional in the sense that the object and the utterance are respectively an observation and an action for the speaker, and an action and an observation for the listener. This can be accounted for in Q-learning by considering the Q-table as a bidirectional lookup table. If the rows represent the objects and the columns represent the utterances, the rows represent the observations of the agent in the speaker role and the actions of the same agent in the listener role. The columns then represent the actions of the agent in the speaker role and the observations of the agent in the listener role. Using a bidirectional Q-table, an agent can readily use the linguistic knowledge it has learned in the listener role through its language comprehension process for language production in the speaker role and vice versa. The criterion that agents should be able to take up the roles of both speaker and listener and that their language comprehension and production processes should be integrated is thus satisfied.

The set of possible objects and the set of possible utterances are not known by an agent beforehand. New objects and utterances can be encountered at any moment and might stem from an infinite set (see e.g. the definition of the vocabulary of an agent in Section 4.1). The Q-table of an agent can, therefore, not be initialised with all possible objects and utterances at the beginning of an experiment. Each agent is, therefore, initialised with an empty Q-table, to which new rows, representing objects, and new columns, representing utterances, can be dynamically added. An agent in the speaker role will add a row and a column when it does not find a mapping for a given topic object, corresponding to the invention phase in the naming game. An agent in the listener role will add a column and potentially a row when it does not find a mapping for an observed utterance, corresponding to the adoption phase in the naming game.

The values of the Q-table can be updated using the update rule defined in Equation 1, in which $Q(o_t, a_t)$ stands for the value of an observation–action pair in the table, α stands for the learning rate, and r_t stands for the reward obtained:

$$Q_{\text{new}}(o_t, a_t) \leftarrow Q_{\text{old}}(o_t, a_t) \; + \alpha \times \; (r_t - Q_{\text{old}}(o_t, a_t)) \tag{1}$$

This update rule corresponds both to the standard Q-learning update rule for single-action episodes, and to the interpolation update rule described in Section 4.1. When we use the same learning rate as above (α = 0.5), define the reward r_t to be 0 in case of a failed interaction and, respectively, 1 and 0 for used words and their competitors in case of a successful interaction, and use a greedy action-selection strategy (exploitation without exploration, i.e. always selecting the word with the highest score), our MARL-based implementation of the naming game is equivalent to its more traditional language game implementation. Consequently, the experimental results of the MARL-based naming

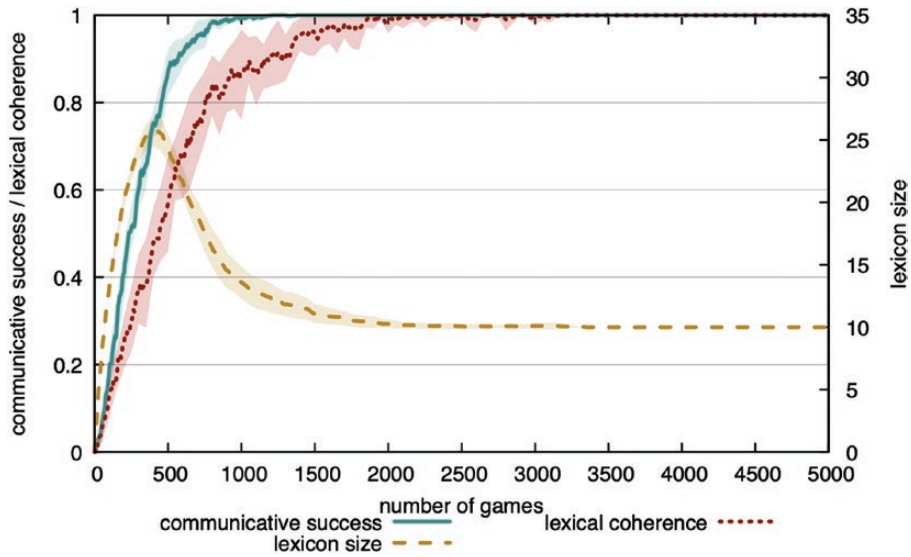

Figure 4 Dynamics of the MARL-based naming game experiment, with ten agents communicating about ten objects.

game, shown in Figs. 4 and 5, exhibit the same global dynamics as those of the canonical naming game presented in Figs. 2 and 3.

Figure 4 presents the communicative success, lexical coherence and lexicon size over time, averaged over ten experimental runs of the MARL-based naming game experiment. Again, all three measures start at 0. Over the course of the first 1,500 interactions, communicative success gradually grows to 1. The lexicon size grows to over 25 after 450 interactions and then starts to shrink until it stabilises at 10 after about 2,500 interactions. At the same time, the lexical coherence stabilises at 1.

Figure 5 shows the dynamics of competition between different forms for the same object in a single agent in a single experimental run. Again, we can see that the agent acquires at first different competing word forms for referring to a single object, in this case, six word forms for referring to object 3. After about 2,500 interactions, the word form 'suyiru' wins the competition and stays the preferred word form for referring to this object for the rest of the experiment.

Tables 1 and 2 show snapshots of the bidirectional dynamic Q-table of a single agent during the same experimental run. The columns and rows, respectively, correspond to the word forms and objects known by the agent. The numbers in the tables represent the scores of the form-meaning mappings. Only the forms that are associated with an object with a score of >0.01 are included. The snapshot shown in Table 1 was taken after forty-three episodes. At this point, the agent has learned seven word forms for six objects. The word forms 'mociqu' and 'wodaze' compete for object 3,

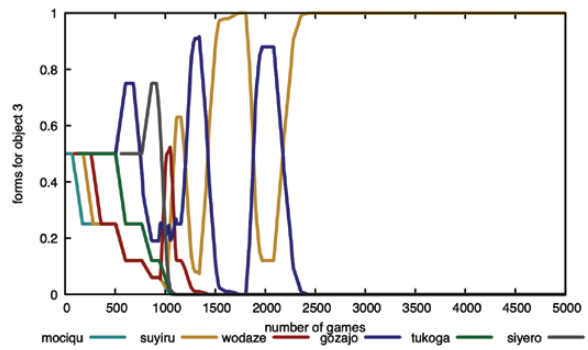

Figure 5 Competition of word forms over time for a single object in a single agent (MARL-based naming game).

with 'wodaze' being the highest scored word form. The snapshot shown in Table 2 was taken after 4,996 episodes. At this point, the population has fully converged on a shared lexicon, as can be read from Fig. 4. This is reflected in the Q-table by the fact that every object is associated with a single, unique word form. The word 'suyiru' has now become the preferred word form for referring to object 3. Indeed, this information is in line with the competition graph shown in Fig. 5.

5. Discussion and conclusion

This paper started from the observation that the emergence and evolution of human-like languages in populations of artificial agents is today studied by two largely distinct communities, respectively, adopting MARL and the language game paradigm as their underlying

Table 1 Bidirectional dynamic Q-table for agent 4 after 43 episodes, showing competition between the word forms `mociqu' and `wodaze' for referring to object 3.

m/f	lejiro	leyovo	mawexi	mociqu	netoso	susavi	wodaze
o-1						0.25	
o-2	0.5						
o-3				0.125			0.75
o-6		0.5					
o-7			0.25				
o-9					0.25		

Table 2 Bidirectional dynamic Q-table for agent 4 after 4,996 episodes. The population has now reached full convergence. Only lexical items with a score >0.01 are shown. The word `suyiru' has become the preferred word to refer to object 3.

m/f	buxowo	ditiye	dolujo	pifije	sohene	suyiru	tofoku	tuqeqo	vegopo	zihuvo
o-1		1.0								
o-2					1.0					
o-3						1.0				
o-4				1.0						
o-5							1.0			
o-6								1.0		
o-7									1.0	
o-8			1.0							
o-9										1.0
o-10	1.0									

methodological framework. While it is clear that the language game and MARL communities share their main objectives and conceptual foundations, the interaction between both communities is often hindered by the use of different terminologies and experimental set-ups. In order to remedy this situation, we have formulated the challenge of re-conceptualising the language game paradigm in the framework of MARL, and have put part of this re-conceptualisation into practice through a case study involving the canonical naming game experiment.

We have conceptualised the naming game as an independent Q-learning problem, in which the environment is only partially observable to the individual agents. We observed that the observation space of an agent in the listener role corresponds to the action space of the same agent in the speaker role and vice versa, while the mapping between both spaces is independent of the discourse role of the agent. This has led us to the development of bidirectional Q-tables that represent the linguistic inventories of the agents. Depending on the discourse role of an agent, one of the dimensions of the table corresponds to its action space and the other to its observation space. The mappings that are learned in one discourse role can then readily be used in the other role. Moreover, we have introduced the use of dynamic Q-tables, which support the incremental extension of the action–observation space of an agent during an experiment and can consequently be used in situations where the possible actions and observations are not known beforehand. The bidirectional dynamic Q-tables can be updated using the standard Q-learning update rule and lead in combination with a greedy action-selection strategy to the same results as the traditional naming game.

The reinforcement learning set-up that we have introduced in this paper provides a way of incorporating the desirable properties of language game experiments into MARL-based emergent communication experiments. It effectively models the emergence and evolution of a flexible and adaptive language in a population of autonomous agents, in which interactions are local and decentralised, in which agents can serve the roles of both speaker and listener, and in which the linguistic inventories of the agents are dynamically expandable.

Our case study on the canonical naming game constitutes only a first step towards the larger goal of aligning the language game and MARL paradigms in

all their aspects. Further research in the same direction will be needed to extend this re-conceptualisation to more advanced language game experiments, including experiments on the emergence and evolution of conceptual and grammatical structures.

We sincerely hope that this paper will open a fruitful discussion between the MARL and language game communities, which can in turn lead to valuable collaborations that will push forward the state of the art in the modelling of the emergence and evolution of human-like languages.

Funding

The research reported on in this paper was financed by the Research Foundation Flanders (FWO - Vlaanderen) through postdoctoral grants awarded to Paul Van Eecke (75929) and Roxana Rădulescu (1286223N), and by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 951846 (MUHAI - https://www.muhai.org).

Notes

1. We focus here solely on computational models of emergent communication, that is, models in which languages emerge to support the communicative needs of agents in task-oriented interactions. Outside this scope, the iterated learning paradigm (see e.g. K. Smith *et al*., 2003; Griffiths and Kalish, 2007; Kirby *et al*., 2014; A. D. Smith, 2014; Kirby, 2017) has been a popular framework for modelling the cultural evolution of language from the perspective of subsequent generations of agents.
2. By autonomous agents, we mean agents that sense and act through their own sensors and actuators, make their own decisions, and are not under any form of central control.
3. We use the term mind-reading capability to refer the capacity of an agent to access another agent's thoughts. We do not claim that agents should not build a theory of mind of other agents, that is, reason about what other agents might or might not know.

References

Beuls, K., and S. Höfer. (2011). 'Simulating the Emergence of Grammatical Agreement in Multi-agent Language Games'. In: T. Walsh (ed), *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, pp. 61–66. Palo Alto, CA: AAAI Press.

Beuls, K., and L. Steels. (2013) 'Agent-Based Models of Strategies for the Emergence and Evolution of Grammatical Agreement', *PLoS One*, 8: e58960.

Blythe, R. A., and W. Croft. (2012) 'S-Curves and the Mechanisms of Propagation in Language Change', *Language*, 88: 269–304.

Bogin, B., M. Geva, and J. Berant. (2018). 'Emergence of Communication in an Interactive World with Consistent Speakers'. arXiv preprint arXiv:1809.00549.

Cao, K., A. Lazaridou, M. Lanctot, J. Z. Leibo, K. Tuyls, and S. Clark. (2018). 'Emergent Communication Through Negotiation'. In: I. Murray, M. Ranzato, & O. Vinyals (eds), *Proceedings of the 6th International Conference on Learning Representations*, pp. 1–15.

Cornudella Gaya, M., T. Poibeau, and R. van Trijp. (2016). 'The Role of intrinsic Motivation in Artificial Language Emergence: A Case Study on Colour'. In: Y. Matsumoto, & R. Prasad (eds), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 1646–56. New York, NY: Association for Computational Linguistics.

Darwin, C. R. (1871). *The Descent of Man, and Selection in Relation to Sex*, 1st edn., Vol. 1. London: John Murray.

Das, A., S. Kottur, J. M. Moura, S. Lee, and D. Batra. (2017). 'Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning'. In: R. Cucchiara, Y. Matsushita, N. Sebe, & S. Soatto (eds), *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2951–60. New York, NY: IEEE.

De Beule, J., and B. K. Bergen. (2006, April). 'On the emergence of compositionality'. In: A. Cangelosi, A. D. M. Smith, & K. Smith (eds), *The Evolution of Language. Proceedings of the 6th International Conference*, pp. 35–42). Singapore: World Scientific. doi:10.1142/9789812774262_0005

de Boer, B. (2001). *The Origins of Vowel Systems*. Oxford: Oxford University Press.

de Greeff, J., and T. Belpaeme. (2011). 'The Development of Shared Meaning Within Different Embodiments'. In: *Proceedings of 2011 IEEE International Conference on Development and Learning (ICDL)*, Vol. 2, pp. 1–6. New York, NY: IEEE.

De Vylder, B., and K. Tuyls. (2006) 'How to Reach Linguistic Consensus: A Proof of Convergence for the Naming Game', *Journal of Theoretical Biology*, 242: 818–31.

Echterhoff, G. (2013) 'The Role of Action in Verbal Communication and Shared Reality', *Behavioral and Brain Sciences*, 36: 354–5.

Foerster, J., I. A. Assael, N. de Freitas, and S. Whiteson. (2016). 'Learning to Communicate with Deep Multi-agent Reinforcement Learning'. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (eds), *Advances in Neural Information Processing Systems*, Vol. 29, pp. 2137–45. Red Hook, NY: Curran Associates, Inc.

Goldberg, A. (2019). *Explain Me This*. Princeton, NJ: Princeton University Press.

Grice, P. (1967). 'Logic and Conversation'. In: P. Grice (ed.), *Studies in the Way of Words*, pp. 41–58. Cambridge, MA: Harvard University Press.

Griffiths, T. L., and M. L. Kalish. (2007) 'Language Evolution by Iterated Learning with Bayesian Agents', *Cognitive Science*, 31: 441–80.

Heylighen, F., et al. (2001). 'The Science of Self-organization and Adaptivity'. In: L. Kiel (ed.), *Knowledge Management, Organizational Intelligence and Learning, and Complexity. The Encyclopedia of Life Support Systems*, pp. 253–80). Oxford: EOLSS Publishers.

Hoffmann, T. (2019) 'Language and Creativity: A Construction Grammar Approach to Linguistic Creativity', *Linguistics Vanguard*, 5: 20190019.

Kirby, S. (2017) 'Culture and Biology in the Origins of Linguistic Structure', *Psychonomic Bulletin & Review*, 24: 118–37.

Kirby, S., T. Griffiths, and K. Smith. (2014) 'Iterated Learning and the Evolution of Language', *Current Opinion in Neurobiology*, 28: 108–14.

Lazaridou, A., A. Peysakhovich, and M. Baroni. (2017). 'Multi-agent Cooperation and the Emergence of (Natural) Language'. arXiv preprint arXiv:1612.07182.

Lenaerts, T. et al. (2005) 'The Evolutionary Language Game: An Orthogonal Approach', *Journal of Theoretical Biology*, 235: 566–82.

Lipowska, D., and A. Lipowski. (2022) 'Emergence and Evolution of Language in Multi-agent Systems', *Lingua*, 272: 103331.

Loreto, V. et al. (2011) 'Statistical Physics of Language Dynamics', *Journal of Statistical Mechanics: Theory and Experiment*, 2011: P04006.

Mikolov, T., A. Joulin, and M. Baroni. (2018). 'A Roadmap Towards Machine Intelligence'. In: A. Gelbukh (ed), *Computational Linguistics and Intelligent Text Processing. CICLing 2016. Lecture Notes in Computer Science*, Vol. 9623. Berlin: Springer. https://link.springer.com/book/10.1007/978-3-319-75477-2

Mordatch, I., and P. Abbeel. (2018). 'Emergence of Grounded Compositional Language in Multi-agent Populations'. In: S. A. McIlraith, & K. Q. Weinberger (eds), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 1495–502. Palo Alto, CA: AAAI Press.

Oudeyer, P.-Y., and F. Kaplan. (2007) 'Language Evolution as a Darwinian Process: Computational Studies', *Cognitive Processing*, 8: 21–35.

Pfeifer, R., M. Lungarella, and F. Iida. (2007) 'Self-organization, Embodiment, and Biologically Inspired Robotics', *Science*, 318: 1088–93.

Pickering, M. J., and S. Garrod. (2013) 'An Integrated Theory of Language Production And comprehension', *Behavioral and Brain Sciences*, 36: 329–47.

Rădulescu, R., and K. Beuls. (2016) 'Modelling Pronominal Gender Agreement in Dutch: From a Syntactic to a Semantic Strategy', *Belgian Journal of Linguistics*, 30: 219–50.

Resnick, C., I. Kulikov, K. Cho, and J. Weston. (2018). 'Vehicle Communication Strategies for Simulated Highway Driving'. arXiv preprint arXiv:1804.07178.

Schleicher, A. (1863/1869). *Darwinism Tested by the Science of Language*. English translation of Schleicher 1863, translated by A. V. W. Bikkers. London: John Camden Hotten.

Smith, A. D. (2014) 'Models of Language Evolution and Change', *Wiley Interdisciplinary Reviews: Cognitive Science*, 5: 281–93.

Smith, J. M., and E. Szathmáry. (2000). *The Origins of Life: From the Birth of Life to the Origin of Language*. Oxford: Oxford University Press.

Smith, K., S. Kirby, and H. Brighton. (2003) 'Iterated Learning: A Framework for the Emergence of Language', *Artificial Life*, 9: 371–86.

Spranger, M., and L. Steels. (2015, July). 'Co-acquisition of Syntax and Semantics: An Investigation in Spatial Language'. In: Q. Yang and M. Wooldridge (eds.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 1909–15). Palo Alto, CA: AAAI Press.

Steels, L. (1995) 'A Self-organizing Spatial Vocabulary', *Artificial Life*, 2: 319–32.

Steels, L. (2000). 'The Emergence of Grammar in Communicating Autonomous Robotic Agents'. In: W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence*, pp. 764–9. Amsterdam, Netherlands: IOS Press.

Steels, L. (2011) 'Modeling the Cultural Evolution of Language', *Physics of Life Reviews*, 8: 339–56.

Steels, L. (ed.). (2012a). *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Steels, L. (2012b). 'Self-organization and Selection in Cultural Language Evolution'. In: L. Steels (ed.), *Experiments in Cultural Language Evolution*, Vol. 3, pp. 1–37. Amsterdam: John Benjamins.

Steels, L., and F. Kaplan. (1998). 'Spontaneous Lexicon Change'. In: C. Boitet & P. Whitelock (ed), *Proceedings of COLING 1998, the 17th International Conference on Computational Linguistics*, pp. 1243–50. Stroudsburg, PA: Association for Computational Linguistics.

Steels, L., and E. Szathmáry. (2018) 'The Evolutionary Dynamics of Language', *Biosystems*, 164: 128–37.

Sukhbaatar, S., A. Szlam, and R. Fergus. (2016). 'Learning Multiagent Communication with Backpropagation'. In: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds), *Advances in Neural Information Processing Systems*, Vol. 29, pp. 2244–52. Red Hook, NY: Curran Associates, Inc.

Tan, M. (1993). 'Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents'. In P.E. Utgoff (ed), *Proceedings of the Tenth International Conference on Machine Learning*, pp. 330–7. Amherst, MA: Morgan Kaufmann.

Van Eecke, P. (2015). 'Achieving Robustness Through the Integration of Language Production in Comprehension'. In G. Airenti, B.G. Bara, & G. Sandini (eds), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science*, pp. 187–92. CEUR-ws.org.

van Trijp, R. (2013) 'Linguistic Assessment Criteria for Explaining Language Change: A Case Study on Syncretism in German Definite Articles', *Language Dynamics and Change*, 3: 105–32.

Vogt, P. (2003) 'Anchoring of Semiotic Symbols', *Robotics and Autonomous Systems*, 43: 109–20.

Watkins, C. J. C. H. (1989). *Learning from Delayed Rewards*, Unpublished doctoral dissertation, University of Cambridge, Cambridge.