THE UNIVERSITY
of EDINBURGH

An Autopoietic Approach to Cultural Transmission

Alex Papadopoulos-Korfiatis

Doctor of Philosophy

Institute for Language, Cognition and Computation
School of Informatics
University of Edinburgh

2017

Abstract

Non-representational cognitive science is a promising research field that provides an alternative to the view of the brain as a "computer" filled with symbolic representations of the world and cognition as "calculations" performed on those symbols. Autopoiesis is a biological, bottom-up, non-representational theory of cognition, in which representations and meaning are framed as explanatory concepts that are constituted in an observer's description of a cognitive system, not operational concepts in the system itself. One of the problems of autopoiesis, and all non-representational theories, is that they struggle with scaling up to high-level cognitive behaviour such as language.

The Iterated Learning Model is a theory of language evolution that shows that certain features of language are explained not because of something happening in the linguistic agent's brain, but as the product of the evolution of the linguistic system itself under the pressures of learnability and expressivity. Our goal in this work is to combine an autopoietic approach with the cultural transmission chains that the ILM uses, in order to provide the first step in an autopoietic explanation of the evolution of language.

In order to do that, we introduce a simple, joint action physical task in which agents are rewarded for dancing around each other in either of two directions, left or right. The agents are simulated e-pucks, with continuous-time recurrent neural networks as nervous systems. First, we adapt a biologically plausible reinforcement learning algorithm based on spike-timing dependent plasticity tagging and dopamine reward signals. We show that, using this algorithm, our agents can successfully learn the left/right dancing task and examine how learning time influences the agents' task success rates.

Following that, we link individual learning episodes in cultural transmission chains and show that an expert agent's initial behaviour is successfully transmitted in long chains. We investigate the conditions under which these transmission chains break down, as well as the emergence of behaviour in the absence of expert agents. By using long transmission chains, we look at the boundary conditions for the re-establishment of transmitted behaviour after chain breakdowns.

Bringing all the above experiments together, we discuss their significance for non-representational cognitive science and draw some interesting parallels to existing Iterated Learning research; finally, we close by putting forward a number of ideas for additions and future research directions.

Acknowledgements

Writing these acknowledgements is proving to be the hardest part of this thesis; I am grateful to so many people for their support throughout the years of my PhD that I am overwhelmed. The space here is not enough to fit everyone I want, and either way a short phrase about each of them will never be enough to convey just how much they mean to me; I hope they know.

Let me start by thanking my supervisors: Jon Oberlander, Simon Kirby and Barbara Webb. I could not hope for better supervision and support; it goes without saying that without them I would be completely lost. I owe this thesis to them. The weekly meetings with Jon in the first year of my studies were incredibly stimulating and always a highlight in my week; I will remember them fondly.

Dan, Rimvydas and Tom were not only supportive friends but also helped shape this thesis through countless discussions about philosophy, cognitive science and beer brewing. (Tom, thanks for all the climbing!) Alessia, Chiara, Ed and all the members of the Joint Action Reading Group: thank you for making me feel part of a community; the Innovative Learning week events were so much fun! (Alessia, also thanks for all the cooking and the music.)

Paris, thanks for coming with me to Edinburgh and for all the support throughout the years - keep building stuff! (Hopefully we'll build the next Pixelaras together; Erica, please put up with us.) Andreas, what can I say - my constant PhD buddy and fellow cooking enthusiast - and a rock star! Anastasis and Ime, my Edinburgh family. Edinburgh would be completely different if not for you two. The family times in the flat were some of my nicest moments in Edinburgh. (Anastasis... Tsk tsk tsk.)

On the topic of Edinburgh family: Johanna, Cedric, Victor; thanks for everything all those years in Arden Street, what would I have done without you? (Not forgetting everyone else from Arden: Gastone, Marina, Michele, Ruth - I will miss you all!) Ieva, thanks for all the good times (so many trips!) and good luck with your own PhD. Seb and Veni, your support in the last few hard months meant more than you realise; same of course for Giorgos who had to put up with me in those last months - thanks for everything!

Speaking about putting up with me, I can't thank Jasmine enough for all of her support and for being all around wonderful, I hope I will be able to return the favour in the future - shie shie Liang-Yu!

My “old” friends from Greece: Pavlos, Dimos, Ilias, Giorgos, Giannis. You supported me during my PhD as you’ve been supporting me in everything for way too long now; while Edinburgh was an amazing city, it could never have been my home because all of you were not here. Pavlos, I appreciate the fact that you tried to make it a bit more like home by following me for a couple of years - even though you left, Edinburgh still seems to be talking about you!

Of course, the Edinburgh University Tango Society cannot be absent from these acknowledgements. In what seems to be a recurring theme, they were like a family to me; Edinburgh and tango will be always connected in my mind. I won’t start mentioning names or I could fill pages upon pages; thanks everyone for all the nice dances but mostly thanks for being an amazing community. Toby, it’s all thanks to you as well. I will keep pushing into the floor!

Last but definitely not least, I would like to thank my wonderful parents, Haris and Mary, and my sister Ioanna: I cannot express how much they mean to me here so I will leave it at this.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Alex Papadopoulos-Korfiatis*)

Table of Contents

Chapter 1
Introduction

In this thesis, we will describe a first step towards the combination of a non representational theory of cognition, *autopoiesis*, with a functionalist approach to evolutionary linguistics, *Iterated Learning*. Our approach to this combination is practical: we will make use of simulated robots, neural networks and joint action experiments. Nonetheless, it is informed and motivated by philosophical enquiries about the nature of human cognition; what we aspire to be doing is "philosophy of mind using a screwdriver" (Harvey, 2000). It seems fitting, then, to start by explaining what kind of "philosophical enquiries" led us to believe that the combination we mentioned is worthwhile in the first place; this is what we will do in this first chapter.

1.1 The computational approach to cognition

A commonly held hypothesis in the fields of Artificial Intelligence and Cognitive Science has been that cognitive agents use symbols to *represent* their environment, and that the nature of cognition is the formal manipulation of these representations inside their brain. This assumption, usually referred to as the "computational hypothesis" (Van Gelder, 1998) was perhaps most famously formulated by Newell and Simon (1976) in their claim that *"a physical symbol system has the necessary and sufficient means for general intelligent action"*. We can roughly translate this claim into the following propositions:

(i) All known facts about the world are represented as symbols in an agent's brain; through some mechanism, sensory input updates these representations.

(ii) "Thinking", or cognising, happens through computations on these symbolic representations.

(iii) Finally, the agent acts on the world through its motor facilities.

The symbolic approach is attractively intuitive; however, while it initially was highly successful in "solving" high-level reasoning problems that could be captured in a reasonably-sized set of symbolic rules (chess being the most obvious example), it has proven particularly inept at facing seemingly easier cognitive problems such as sensorimotor coordination, real world planning or common–sense reasoning (Van Gelder, 1998). "General intelligence" was predicted to be solved "within 20 years" in the 1970s (Russell and Norvig, 1995, p. 21), but we are still a long way from even building systems with the cognitive capacities of human infants.

This is perhaps most visible in the field of robotics; much like humans and other animals, robotic agents need to control a complex body in a real environment — an environment not simple enough to be described using the propositional logics of symbolic AI, in principle and in practice (Dreyfus, 2007). Indeed, motion control in the real world has proven to be a notoriously difficult task; despite significant theoretical and technological advances, "much work remains to be done" (Belta et al., 2007). One reason for this inability to deal with "simple" sensorimotor coordination and cognitive tasks is a manifestation of an issue that has plagued the field of Artificial Intelligence since its very beginning. The issue (known as the *frame problem*) is the following: how do we, as agents, decide what is relevant in a given situation and what is not?

Minsky (1977) has famously tried to answer this question by introducing the concept of *frames*, or structures that represent and encode relevancy for a particular situation. The problem, however, goes deeper than that: to decide what situation we are in (or what frame we are to use), we must have knowledge of what aspects of our environment are relevant. To decide what aspects are relevant, in turn, requires us to know what frame to use. This circularity poses a deep threat to Minsky's frame system and similar approaches (Dreyfus, 2007; Vervaeke et al., 2012).

The frame problem, in addition to posing a practical issue for the fields of robotics and AI, has been appropriated by philosophers (Chow, 2013) as an *epistemological* argument against symbolic cognitive science approaches, adding to the arsenal of critics that describe such approaches as a dead end in the goal of understanding human cognition (Dreyfus and Dreyfus, 1988; Dreyfus, 2007; Van Gelder and Port, 1995). In fact, this philosophical debate pre-dates the field of cognitive science altogether; it can be traced back to a long history of 20^{th} century continental philosophers (notably Martin Heidegger and Maurice Merleau-Ponty) pointing out that rationalist views of cognition (Hobbes' proposal of "reasoning as computation" or Descartes' "mental representations") have fundamental issues with "significance and relevance" (Dreyfus, 2007).

1.2 The embodied approach to cognition

In fact, it is from continental philosophy's focus on "acting in the world" that we are provided with an alternative to the symbolic approach to cognition. This approach, *embodied cognition*, is a popular research paradigm (Chemero, 2009) that recognises that any attempt to explain cognition has to consider not only an agent's brain but, equally, its body and its situatedness in an environment. In other words, cognition is the product of the interaction between brain, body and environment; however, because these systems are so tightly interconnected, they can only be explained as a whole and not independently from each other (Beer, 2000). This, then, constitutes a *systemic* approach to cognition as opposed to the *internalist* computational approach.

A useful methodology to research cognition as a brain, body and environment interaction comes from adopting what Chemero (2009) calls the "dynamical stance": the use of tools borrowed from dynamical systems theory to research cognitive phenomena, analysing an agent's brain, body and environment as interacting non-linear dynamical systems (see also the "dynamical hypothesis" of Van Gelder, 1998). We will use two experiments by Randall Beer to illustrate what such an approach looks like.

The first experiment (Beer, 1995a) was a simulation of broadly insect-like agents. Each of the agents had six independent legs that could "swing" in space, propelling the agent forward if they were in contact with the ground; each leg was controlled by a 5-node continuous-time recurrent neural network.[1] The agents had the task of moving the furthest possible distance; using a genetic algorithm that selected agents for their fitness in this task, Beer evolved agents that could reliably synchronise the swinging of their legs into a stable walking gait.[2]

The second experiment (Beer, 2003) was a simulation of agents that exhibited a behaviour more likely to be recognised as "cognitive": categorical perception. This time, the simulated agents had a sensory system (a basic visual system of seven "rays" giving feedback on object collision) as well as a motor system (two motors, whose combined motor forces allowed the agents to move horizontally); the task they were asked to accomplish was to distinguish between "falling" objects of various shapes ranging from perfect squares to circles.

[1] Continuous-time recurrent neural networks are approximators of dynamical systems, and are often used in dynamical cognition research; we will come back to them in Chapter 4.

[2] Note that agents' legs also had a sensor reporting the angle between the leg and the body; stable (though less resilient to perturbation) walking gaits evolved with these sensors disabled as well.

Once more, genetic algorithms selected agents for their fitness in this task and produced evolved agents that could successfully categorise the objects using *active perception*: while the input to the visual system was identical for all shapes when stationary, by actively moving the agents were able to associate different shapes to different sensory-motor patterns.

The most interesting part of Beer's experiments, however, was that he performed a detailed dynamical analysis of the evolved agents' behaviour. In the case of the "walking insects", he showed that while all the agents exhibited very similar behaviour, the networks were in some cases vastly different; furthermore, no discrete function of the system could be mapped to a specific network node. Similarly, in the case of the "categorisers", the differentiation between shapes could not be attributed to specific points in the network; it only arose from the interaction between the evolved networks and the environment. While only the weights of the nodes are changed by the process of evolution, Beer claims, what is *selected for* is "a property of the dynamics of the entire coupled system" (Beer, 2003, p. 236).

Beer's experiments establish that autonomous, arguably (minimally) cognitive behaviour[3] does not equate necessarily to the manipulation of symbolic representations; by doing so, they provide a practical, workable alternative to the computational view of cognition. (For an in-depth discussion of more experiments that adopt similar dynamical approaches see Chemero, 2009.)

1.3 A note on data-driven approaches

In the last years, a new and altogether different approach to Artificial Intelligence based on using machine learning and vast amounts of example or input data has been attracting publicity through important milestones; for example, beating human competitors in the games of Jeopardy (Lally and Fodor, 2011) and Go (Silver et al., 2016).

At the same time, similar approaches have been achieving significant advances in problems that had, until now, proved very challenging for computers to solve; good examples of such problems would be speech recognition (Amodei et al., 2015), machine translation (Sutskever et al., 2014) and image classification tasks (Krizhevsky et al., 2012); for a comprehensive overview of deep learning, one of the most successful data-driven approaches, see LeCun et al. (2015).

[3]This is a point of contention; Edelman (2003), for instance, describes Beer's experiment as a "toy problem".

However, these data-driven approaches make no claims regarding the nature of cognition, nor do they try to give any insight into cognitive phenomena; instead, the problem domains mentioned above are treated as purely engineering challenges. For this reason, we will not elaborate further on such approaches or try to compare them to the computational or embodied approaches to cognition; while surpassing the results of either of these approaches in the fields mentioned, and while definitely relevant for AI as "computer science", they are not relevant to cognitive science per se or to the goals that we have set out to for this thesis.

1.4 Representations and "scaling up"

Returning to the embodied approach to cognition, even within the embodied cognitive science community, there is a long standing debate on the importance and need of mental representations. One side completely rejects representations as a concept that is required to explain human cognition; the philosophical critique of Dreyfus and Dreyfus (1988) and the modelling experiments of Beer (2000) belong to this side, as do the "dynamical hypothesis" research programme of Van Gelder and Port (1995) and the coordination dynamics of Kelso (1995). Chemero (2009), firmly in the anti-representational camp himself, groups all these approaches under the title "Radical Embodied Cognitive Science".

The other side, which would be the "non-radical" embodied cognitive science, recognises the importance of embodiment in the study of cognitive behaviour but joins the critics of non-representational approaches (Edelman, 2003) in their claim that any explanation that makes no use of representations as a concept might work for low-level behaviour but cannot scale up to what Clark and Toribio (1994, p. 28) term "representation-heavy" domains, two major examples of which are high-level planning and language. Even a dynamical systems approach, the less radical camp claims, does not theoretically preclude representations (Bullock, 2004), which can take the form of "trajectories or attractors of various kinds, or even such exotica as transformations of attractor arrangements as a system's control parameters change" (Van Gelder and Port, 1995).

An alternative approach to combine representations and embodied cognition comes from *action–oriented* or *minimal* representations. These representations, unlike the traditional representations of the computational approach to cognition, would in the words of Wheeler (2005, p.197) mirror "how the world [...] is itself encoded in terms

of possibilities for action". Whether these minimal representations should still count as representations is, however, also a much debated topic (Gallagher, 2008; Routledge et al., 2008); as is the matter of whether there really is a need for this kind of "contamination" of the radical dynamical approach (Garzón, 2008).

1.5 Language

At the same time, a different but related debate is taking place amongst scientists investigating the origins of language. On one side, *nativist* accounts of language evolution posit that language is encoded in a dedicated, biologically evolved "language faculty" (Chomsky, 1995; Pinker and Bloom, 1990). The claim usually involves the evolution of symbolic thought capacity in one of our close biological ancestors; this "language of thought" then gets *externalised* into human language (see for example Chomsky, 2007, pp. 24). This rings very close to the computational view of cognition as symbolic manipulation; indeed, Chomsky's conclusions have been very influential in establishing the computational hypothesis in cognitive science (see Chemero, 2009, p. 6).

On the other side, a growing number of researchers have been questioning this nativist view and adopting more systemic approaches. One such approach, *Iterated Learning*, stands out in our view by establishing that certain features of language can be explained by its nature as a culturally transmitted system (Kirby, 2002a); instead of biological evolution encoding language structure in human brains, it is language itself that evolves structure through its transmission in iterated learning chains.

Taking a step back from both debates, the fact remains that non-representational accounts of cognition are faced with a genuine problem: language is representational by definition, as linguistic statements *refer* to objects or events (Ikegami and Zlatev, 2007, p. 242). Any non-representational theory, then, needs to be able to explain the representational character of language in order to be complete.

However, while the representational nature of language is a major challenge to non-representational accounts of cognition, it also presents the potential for a very powerful explanatory mechanism. Working out how the representational structure of language emerges from a dynamical substrate is only a short step away from a systemic, bottom-up explanation of the emergence of representations and symbolic thought *from* language. Such an explanation has interesting parallels to Vygotskian views of thought as *internalised* language (as opposed to language as *externalised* thought; see Donald, 2000, p. 33).

Arguably, the best candidate that can extend a bridge from dynamical low-level behaviour to language is the theory of *Autopoiesis* (Maturana and Varela, 1980); a complete and rigorous theory of cognition as a biological phenomenon. Autopoiesis is a very influential theory within the non-representational cognitive science camp; we conveniently omitted any mention of it so far with the goal of discussing it (along with Iterated Learning) in detail in Chapter 2.

We believe that the intersection between dynamical cognitive science as a systemic approach to human cognition and non-nativist approaches to language evolution has fascinating potential; it offers a glimpse at the connection between pre-symbolic and human-level cognition, as well as a bottom-up explanation of the emergence and grounding of representations. This is the motivation that drives the work we will present in this thesis. Our goal, of course, is not to provide this explanation in its entirety, but to take a first step towards it.

1.6 Contributions

In this thesis, we take the first step towards an autopoietic account of language structure by establishing that cultural transmission chains, an essential component of the Iterated Learning account of the evolution of language, can be instantiated from an autopoietic, non-representational substrate. Our contributions are the following:

1. We propose the combination of Autopoiesis and Iterated Learning as a promising practical approach to an autopoietic account of language structure, and a joint action task ("L/R dancing") that allows us to bridge the theoretical incompatibilities between the two theories.

2. We design and build a system that allows us to run experiments of cultural transmission chains with simulated robots performing the L/R dancing task; in order to do so, we adapt a biologically plausible learning algorithm that allows the simulated robots to learn the dancing task.

3. We demonstrate successful cultural transmission chains and the spontaneous emergence of dancing between interacting agents; we also connect the behaviour of the system to issues of current interest to Iterated Learning research.

1.7 Thesis outline

The rest of this thesis is organised in the following chapters:

In **Chapter 2** we introduce *Autopoiesis* as a theory of cognition and *Iterated Learning* as a model of language evolution; we consider the problems that arise from a potential combination of these two fields and propose ways around these problems.

In **Chapter 3** we detail the design of a system and a task ("L/R dancing") that will allow us to build transmission chains, needed for Iterated Learning, while following the constraints introduced by an autopoietic approach.

In **Chapter 4** we describe the implementation of a biologically plausible reinforcement learning algorithm and establish that agents using this algorithm are able to learn the dancing task; we examine how learning time influences the agents' success rate at the task.

In **Chapter 5** we connect individual dancing episodes into cultural transmission chains and establish that stable chains are possible; we examine the stability of the chains as a function of agent learning time.

In **Chapter 6** we look at the spontaneous emergence of dancing behaviour in the absence of pre-trained agents, in both isolated interacting agent pairs and in cultural transmission chains.

Finally, in **Chapter 7** we bring everything together and discuss the results of all experiments as a whole, connecting them to Iterated Learning and Autopoiesis research; we conclude by presenting a list of potential future research directions.

Chapter 2
Background & related work

In this chapter we will examine two leading fields in the areas of interest for this thesis: the cognitive theory of *Autopoiesis*, which aims to provide a complete account of how life and cognition are based on biology; and the research paradigm of *Iterated Learning*, that shows how the transmission of language from generation to generation leads to the evolution of linguistic structure. We will consider the problems that arise in trying to combine these two fields and put forward a potential way to accomplish a combination.

2.1 Autopoiesis

2.1.1 Theory

One of the most influential contributions in non-representational embodied cognitive science has come from two Chilean biologists, Humberto Maturana and Francisco Varela, who in 1972 introduced the concept of "autopoiesis". Around this concept, which we will explain in this section, Maturana and Varela built a complete biological theory of cognition; introduced in their book *Autopoiesis and Cognition: The Realization of the Living* (Maturana and Varela, 1980), this theory was enriched in numerous subsequent books and publications (most notably Maturana and Varela, 1992; Varela et al., 2000) and has been a major influence in a plethora of fields. In the cognitive sciences, these include modern philosophy of mind (with the most prominent example being Alva Noë's work— Noë, 2004; O'Regan and Noë, 2001; Noe, 2009), Artificial Intelligence and robotics (Morse and Ziemke, 2007; Froese and Ziemke, 2009; Froese and Di Paolo, 2011). The power of autopoiesis as a theory, however, is possibly best attested by its effect on more distant fields such as sociology (influencing scholars

like Luhmann and Habermas, as discussed in Leydesdorff, 2000; Luhmann, 1986) and architecture (Schumacher, 2011).

Maturana and Varela's theoretical biology starts from the recognition of the inherent circularity present in any definition of living systems as objects observed and described (Maturana and Varela, 1980, p. v) and goes on to define life and cognition as the property of *autopoietic organisms*: autonomous, operationally closed systems that both create and specify themselves. As a complete theory, it provides the explanatory power to study all aspects of cognition, including joint action, communication, social constructs and language; however, as is often the case with complete theories, it is described in an idiosyncratic language and comes bundled with a number of definitions. While we could not hope to explain all aspects of the theory in detail, in this chapter we will attempt to introduce some of the basic concepts of autopoietic theory that are relevant for the work we will present in this thesis.

2.1.1.1 Autopoietic systems

Autopoiesis is "a theory of the organisation of living organisms as autonomous entities" (Maturana, 1975); an autopoietic system is defined as a system of molecular networks and interactions that:

1. Produce themselves. The organisation of this set of networks and interactions is such that they generate themselves.

2. Specify their own limits. Some of the components that are part of the system form a *boundary*, a limit to this network of interactions.

In other words, an autopoietic system's organisation and boundary are maintained as products of its own operation. While this process of self-production and self-definition is maintained, the system is question is an autopoietic organism and is alive: in this way, the definition of autopoiesis is at the same time a definition of life.

The boundary created by an autopoietic system allows us, as an outside observer, to refer to it as a *unity*. This distinction is always made in contrast to an environment or background which Maturana names the autopoietic system's *medium*. The medium and the autopoietic organism are mutually specified; neither of them can be defined independently, and by defining one we are also defining the other.

Since an autopoietic system both defines and produces itself, it is classified as an *operationally closed* system. This means that the system's behavioural space is generated from within and not dictated from outside; in other words, it only depends on its

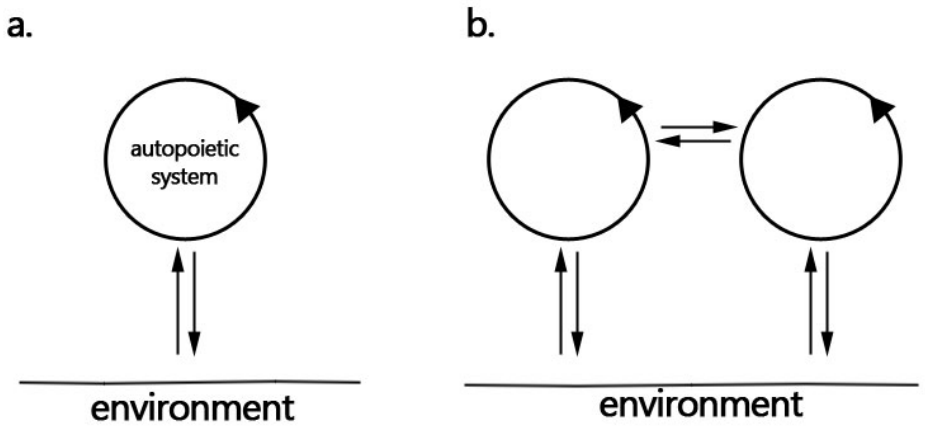


Figure 2.1: Sketch of an autopoietic system interacting with its environment (a) and of two autopoietic systems in structural coupling with each other (b).

current structure and previous states and not on anything that is part of the environment (or "medium").

Even though the domain of possible states the system can find itself in is generated by the autopoietic system and not by its medium, the environment can interact with the system through perturbations that trigger state transitions (Fig. 2.1a). The history of structural changes that an autopoietic system goes through as a result of such perturbations constitute the *ontogeny* of the system.

2.1.1.2 Structural coupling

Just as an autopoietic system can interact with its environment, it can also interact with other (possibly autopoietic) systems. It is important to note that from the viewpoint of a given organism (autopoietic system) there can be no distinction between its environment and another organism; we, however, as outside observers, can make that distinction. Interactions between two organisms, when recurrent, can lead to plastic structural changes in both systems, thus changing the possible future state space of each system. Whenever there is a history of such reciprocal plastic interactions, we can speak of *structural coupling* between the two systems (Fig. 2.1b). A special case of structural coupling would be a series of autopoietic systems connected through a reproductive chain while maintaining adaptation; this is the *phylogeny* of a species.

2.1.1.3 Behaviour and the nervous system

Even the simplest autopoietic systems interact in some way with their environment. We can define this organism-environment interaction as the organism's *behaviour*. Once more, this definition is not of something intrinsic to the organism but a distinction made by us as external observer. In other words, behaviour is not something that the organism *does* but rather a way for the observer to *refer* to the interaction taking place between an autopoietic organism and its environment.

Through structural changes that propagate from generation to generation of autopoietic organisms ("structural drift") and natural selection, together forming the process of evolution, even simple autopoietic organisms have multiple ways of interacting with their environment. Depending on the nature of that interaction, we can refer to the areas of the system taking part in it as *sensors* (surfaces sensitive to external perturbation) or *motors* (areas capable of producing movement). Internal correlations between the sensory and motor areas of an organism lead to more complex forms of behaviour and sensorimotor coordination.

A long history of evolution means that some, more complex, autopoietic systems have extensive recurrent connectivity networks between their sensory and motor areas ("nervous systems"). By enabling a huge number of sensory-motor correlations, nervous systems open up increasingly complex behavioural domains. However, as should be apparent from the definitions given in this section, we repeat that behaviour is not generated by the nervous system; it simply is expanded in scope by it.

2.1.1.4 Learning

In Section 2.1.1.1, we mentioned that the *ontogeny* of an autopoietic system is the history of structural changes it has been through as a result of its interactions with its environment. These structural changes also apply to an organism's nervous system, often leading to changes in the organism's behaviour. This ontogeny of the nervous system and subsequent change in behaviour is *learning*, defined by Maturana and Varela (1980, pp. 35-38) as:

1. "A phenomenon of transformation of the nervous system associated to a behavioural change that takes place under maintained autopoiesis."

2. "The change in the domain of possible states the nervous system can adopt, taking place along the ontogeny of the organism as a result of its interactions."

2.1.1.5 Social and communicative behaviour

As we saw, the structure of the nervous system of autopoietic organisms opens up more complex behavioural domain spaces. One of these behavioural domains is the *social* domain. Social phenomena are constituted in the structural coupling of organisms to form social systems. Interestingly, some kinds of social systems also adhere to the autopoietic principles of producing themselves and specifying their own boundaries; this is what opens up sociological phenomena to autopoietic interpretations (Leydesdorff, 2000).

More relevant for this work, however, is the fact that the social domain enables another kind of joint behaviour, *communication*. Communication refers to coordinated behaviour mutually triggered between autopoietic organisms in a social context; thus, communicative behaviour only arises in a structural coupling of organisms that are constitutive parts of a social system. We can distinguish between two forms of communicative behaviour: phylogenic and ontogenic.

1. Phylogenic (or *innate*) communicative behaviour is dependent on structures that arise during an organism's developmental process, independently of its particular ontogeny.

2. Ontogenic (or *acquired*) communicative behaviour is dependent on the particular ontogeny of the organism and its history of social interactions.

2.1.1.6 Language

This is where things get more opaque: ontogenic communicative behaviour is defined as behaviour in the *linguistic* domain (Fig. 2.2). The linguistic domain forms the basis for *language*: the behavioural domain of operations in a linguistic domain that result in coordination of actions that pertain to the linguistic domain itself.

In other words, language is the *linguistic coordination* of *linguistic coordination* of behaviour, or second-order linguistic coordination. It is important to note that language is itself a form of behaviour; an action. It is easy to forget this while using the term "language" as a noun. In order to focus on the character of language as an action, Maturana and Varela often use the verb term *languaging*: when we are "languaging", we are acting in the behavioural domain of second-order linguistic coordination.

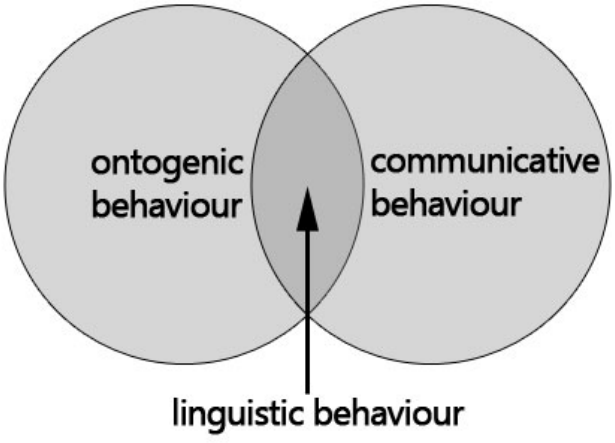


Figure 2.2: Behaviour in the linguistic domain is defined as *ontogenic*, *communicative* behaviour.

2.1.1.7 Beyond language

Maturana and Varela's framework does not stop at this point; humans (as autopoietic systems) are constituted in the domain of language and this domain is used to give an account of human mental life, including human self-consciousness and experience (Maturana and Varela, 1980, p. 50, (iv)) . Such an analysis, however, falls outside the scope of this project so we will not attempt to elaborate on it.

2.1.2 Discussion

In this section, we will attempt to consider the implications that the adoption of an autopoietic framework has for experimental design, look at some of the most common criticisms that the theory of autopoiesis faces as a cognitive research project and expand on the role of language in autopoiesis.

2.1.2.1 Building blocks and explanatory concepts

At various points in the previous section, we pointed out several terms as not intrinsic to the autopoietic organism, but distinctions that we can only make as observers. This clear differentiation between intrinsic concepts and ones that depend on our roles as observers is central to Maturana's discourse (Maturana and Varela, 1980, p. 8).

This is not just a theoretical concern; it is a design decision that is essential in building actual experiments and systems that are explicitly non-representational. It allows

us, as cognitive scientists designing an experiment, to distinguish between *building blocks* that we can use to build our systems and *explanatory concepts* that are useful in describing the system's behaviour afterwards, as observers. It is crucial, then, in the design and building of a system that is consistent with autopoietic principles, not to conflate our system's operational domain (where our "building blocks" are found) with concepts only existing in the domain of descriptions that an observer brings forth.

There are three explanatory concepts that are highly relevant for the work described in this thesis: *representations*, *information* and *objects*.

Representations: The main explanatory concept that we are trying to target is the concept of representations. As discussed in Chapter 1, this is a divisive topic in the cognitive sciences; the explanation and grounding of representations as a mental construct that interacts with physical matter has proven so far out of reach.

Maturana writes in the *Biology of Language*:

> Representation, meaning, and description are notions that apply only and exclusively to the operation of living systems in a consensual domain [...] for this reason, these notions have no explanatory value for the characterization of the actual operation of living systems as autopoietic systems. (Maturana, 1978)

The "consensual domain" of linguistic observers describing a system is where representations are constituted in the autopoietic view; by actively distinguishing between "fundamental" building blocks of our system and representations as a post-linguistic explanatory concept, we avoid the need to provide an account of representations as a basis of all other cognitive phenomena. At the same time, we are not outright discarding representations, but only moving them in the space of post-linguistic cognition as explanatory instead of foundational concepts.

This shift in perspective provides us with a clear path towards the explanation of symbolic representations as a descriptive concept grounded in language and the observer; of course, this means that we now have to explain the symbolic nature of language in a non-representational fashion instead. This switch from representations to language as the base of symbolic cognition is one of the main motivations behind the work we will describe in this thesis; we will discuss it more explicitly later in this chapter.

Information: In addition to representations, another concept that is often misused in the study of communicative behaviour in cognitive science is "information". Often, information is presented as being a "thing", an actual material substance that gets transmitted from agent to agent. However, as Oyama puts it, "Information is not out there [...], it is a way of talking about certain interactions rather than their cause or a prescription for them" (Oyama, 2000, p. 197).

Information is always dependent on context and observer; it always corresponds to a meaningful event, action or utterance and arises within an act of interpretation (Di Paolo, 1997). As such, it is a descriptive, explanatory concept we use as observers to make sense of communicative behaviour; not a building block that plays an operational role in it (Di Paolo, 1997, p. 5). For the purposes of an autopoietic account at least, communicative behaviour is grounded in the biological domain, not in the posterior domain of descriptions information is constituted in.

Objects: More controversial than representation and information, something else that only appears in the domain of descriptions and language are object concepts:

"As we language, objects arise as aspects of our languaging with others, they do not exist by themselves." (Maturana, 2002, p. 28)

Object perception and categorisation might seem like a low-level behaviour, but as Di Paolo (2009) suggests, object and property distinctions are not basic forms of perception but rather a "higher form of cognition", already symbolic and representational in nature. It involves regarding an object in a detached matter and being aware that the same object can have different meanings for other agents. This view is supported by psychophysical experiments as detailed in Gallagher (2009).

2.1.2.2 Autopoietic design and experiments

Part of the power of autopoiesis as a theory of cognition comes from its holistic character: the same principles that are used to explain the very basic aspects of life, grounded in biology, are also used to explain behaviour, learning and language. This is appealing because some of the problems that cognitive science still faces today are circumvented, but at the same time it is extremely limiting. Designing an experiment that is relevant and interesting as a cognitive scientist without having access to descriptive concepts like "information" and "representations" is already challenging; having to

use autopoietic systems that "create themselves" and "specify their own boundaries" seems a rather daunting, if not impossible task.

There are some computer experiments that do actually simulate autopoietic organisms from the ground up, by building systems that are indeed comprised of a network of processes that build themselves and specify their own boundaries. In fact, a computational model of autopoiesis was provided by Maturana and Varela at the time their theory was first introduced (Varela et al., 1974), and this low-level research has been continuing up to the present (Zeleny and Pierre, 1976; Suzuki and Ikegami, 2009; for a comprehensive review, see McMullin, 2004). These experiments are more relevant in synthetic biology, however, rather than cognitive science; it seems incredibly difficult to scale such a system up to research of higher level, cognitively relevant behaviour. As we saw in Chapter 1 and will repeat in the next section, this is a common issue of all non-representational research, taken to the extreme.

Autopoietic experiment design however doesn't necessarily need to start from the deepest layer of biological processes. We can describe an autopoietic system's behaviour in a certain domain without having to provide a description of their constitution in the basic molecular, biological domain. According to Maturana:

> So, living systems exist in two non-intersecting domains, the domain of their components as molecular autopoietic systems, and in the domain in which they operate as organisms (totalities) in a medium that makes them possible. These two domains do not intersect, the processes that take place in one cannot be reduced to the processes that take place in the other. (Maturana, 2002, p. 15)

As long as we follow through the implications and constraints that arise from the fact that any agents we use in our experiments are constituted in the first domain, and thus are autopoietic in nature, we can treat them as "black boxes" or "totalities" when examining their operation in the second domain. What we need to keep in mind then, when designing a cognitive science experiment that is still faithful to the theory of autopoiesis, are the following three points:

Agent design: When designing agents, we do not necessarily need to model them as autopoietic systems. We do however need to treat them as if they were and follow any implications that this has on the interaction of the agent with its environment. More specifically, the agents need to be operationally closed: the states they can be in are determined intrinsically and not dictated by the environment. In other words, the agent has no "first hand" awareness of its environment but is structurally coupled to it through motors and sensors.

Behaviour design: When designing the agents' behaviour, we need to design from a systemic perspective, taking into account the agent - environment - task dynamical system. "Behaviour" is not something that is generated by the agent but an explanatory concept that only arises when we describe, as observers, the whole system of an agent and its environment: "[...] behaviour as a relational dynamics involves both the organism and the medium in which it exists as a totality" (Maturana, 2002, p. 13).

System design: When designing the whole system, we need to carefully avoid conflating explanatory and operational concepts. When using explanatory concepts, especially in the design of an agent, we need to be aware of that and be able to justify their use or explain it using an autopoietic framework.

2.1.2.3 Criticism and challenges

A major problem that autopoietic theory faces as a theory of cognition is the "life-mind continuity gap" as identified by Froese and Di Paolo (2009). This problem is not specific to autopoiesis; it is a problem that we identified in Chapter 1 as present in all non-representational accounts of cognition, usually referred to as the "cognitive gap" issue. The "gap" here is referring to the non-linear progression from minimal behaviour (basic biological processes like bacterial locomotion) and intuitively physical tasks (such as obstacle avoidance, locomotion and sensory-motor coordination tasks) to *representation-hungry* (Clark and Toribio, 1994) cognitive behaviour such as high-level planning, symbolic thought, abstract reasoning or language.

A widespread view amongst cognitive science researchers is that it simply is not possible to scale up from the first type of skills to the second one without resorting to the use of representations as an explanatory mechanism (Clark, 1997; Edelman, 2003). Furthermore, outright discarding representations as a concept is problematic as phenomenology (and direct conscious experience) show us that we, as humans, do use representations and therefore a theory of cognition needs to account for this. Interestingly, however, one of the strengths of autopoiesis is that, as a complete theory, it provides a framework that can be used to explain representations and higher-level cognitive phenomena. This framework is the domain of language; we will discuss this in more detail in a moment.

Another criticism, this time specific to autopoiesis, is that contrary to the theory's claim, the theoretical tools it provides are not sufficient to explain a central concept in

the cognitive sciences, *meaning-making* (Cuffari et al., 2015, p. 8). The concept, however, is extensively covered by Maturana (Maturana, 1978; Maturana and Varela, 1980; Maturana, 2002), who argues against teleological terms in theories of cognition: there is no meaning or purpose intrinsic to an autopoietic organism. There is only a network of processes that either produce themselves, operationally closed from the organism's environment — and then the organism is alive — or fail to produce themselves or their own boundary and the organism disperses, not alive or a unity any more. "Meaning" and "purpose" are distinctions made by an observer *about* an organism interacting with its environment, and conflating the operational domain of the organism with the domain of descriptions where these distinctions appear would be a methodological error. As we will see shortly, as humans we can be observers of our own selves and thus "make" meaning.

Varela, however, in his later work (Varela et al., 2000; Weber, 2002) distanced himself from this rejection of teleology. Trying to combine the theory of autopoiesis with phenomenology, he proposed an account of meaning that is intrinsic to living organisms, who generate it from within in a process called *sense-making*. This alternative approach, not shared by Maturana, has flourished into a popular research paradigm in modern cognitive science, *enactivism* (Froese, 2011). While intriguing, for the work of this thesis we will not consider the expanded ideas on meaning-making and adaptivity (Di Paolo, 2005) that the framework of enactivism provides. As Kravchenko (2011) notes, by introducing teleology in the mechanistic framework of autopoiesis, we are risking a step back towards computational theories of the mind and all of the methodological and theoretical issues those entail. A purely mechanistic account like Maturana's autopoiesis is more restrictive but more principled in its explanatory power.

2.1.2.4 On the role of language

The powerful influence of language on human cognition is well-recognised; it is often put forward as a defining aspect of the human species that sets us apart from animals (Becker, 1991) and as a very useful cognitive tool, used not only for communicative purposes, but also —in both verbal and gestural form— to help the speaker in cognitive tasks (Goldin-Meadow, 1999). Clark (1998) eloquently describes it as an external artefact that "augments human computation". In autopoietic theory, however, language plays an even more essential role in explaining human cognition. In the *Tree of Knowledge*, Maturana writes: "We work out our lives in a mutual linguistic coupling because we are constituted in language in a continuous becoming that we bring forth with oth-

ers." (Maturana and Varela, 1992, ch. 9)

What this means is that in addition to existing as molecular autopoietic systems, constituted in the biological domain, humans are in parallel constituted in the domain of language. We cannot explain human cognition without taking that fact into account: language changes everything by allowing us to become observers, including observing our own behaviour and thus recursively interacting with it. Meaning, representations, objects and categories, information: all of these concepts are constituted in the domain of descriptions that language generates.

The constitution of meaning and symbolic representation in the linguistic domain has distinct similarities to the Vygotskian "Outside-Inside" principle, according to which language is not externalised thought, but rather symbolic thought is internalised language. According to Vygotsky's observations, developing children do not initially have access to inner speech or language of thought. Only after access to language do they start "playing out" symbolic thought, first only in action and subsequently internalised. Evidence from deaf signers points out that naming and symbol use are not inherent human behaviour; names, labels and symbols always come from the outside (Donald, 2000, p. 14).

In this way, language becomes an entity of its own: a bootstrapping system, an external representational framework. It is important, however, to keep in mind that in the first place, the act of language is an action, a social behaviour that living agents take part in. If we want to be able to use the explanatory power of language in a cognitive theory, we must be able to explain, in *biological* terms, where the behaviour of *languaging* came from in the first place. This is an area where Maturana's autopoietic theory is left wanting: languaging is only described as an external observation—linguistic coordination *of linguistic coordination*—but very little is said about how languaging as a behaviour historically evolved in autopoietic agents (Cuffari et al., 2015).

There are a few attempts at autopoietically informed approaches to linguistics (Kravchenko, 2004, 2011; Bottineau, 2008), but again none of them look at the problem of how a linguistic system evolved in human history. This is exactly the question, however, that the field of *evolutionary linguistics* is trying to answer. If we want to begin providing an answer of what an autopoietic view of language evolution would look like, this is where we need to look next.

2.2 Experimental Semiotics

The investigation of the origins of language and of the evolution of symbols and linguistic structure is a very important part of language research. The respective field, *evolutionary linguistics*, has been dominated for a long time by nativist accounts that attribute most aspects of language, including its structure and grammar, to biological evolution and the organisation of the human brain (Pinker and Bloom, 1990). A major argument in favour of this view, originally formalised by Gold (1967) and popularised in linguistics by (Chomsky, 1992) is the *poverty of the stimulus* argument (Berwick et al., 2011) according to which the amount and type of linguistic data developing infants are subject to is not enough for them to infer the inherent grammar of language. Nevertheless, not only do children still learn grammar, but they do so in a surprisingly quick and robust fashion. This fact, according to nativists, means that universal features of grammar (or language structure) must be innate and as such a product of biological evolution. Some mutation in one of our ancestors must have introduced the capability for symbolic thought; this mutation, being of significant evolutionary advantage, was selected for and spread through our species. According to this view, then, language is the externalisation of this internal symbolic "language of thought"; the basic structure common to all natural languages (referred to as *Universal Grammar*) stems from the innate rules of symbolic thought.

This explanation of language is obviously not compatible with non-representational accounts of cognition. Even worse, it is in direct opposition with the view of language that we put forward in Section 2.1.2.4 according to which language is an external symbolic system that gets internalised (as opposed to an internal system that gets externalised). Furthermore, it seems that it is also not compatible with more recent experimental evidence about language acquisition (Elman, 1998; Pullum and Scholz, 2002; Zuidema, 2003). Accordingly, the field of evolutionary linguistics, once dominated by nativist accounts, is steadily moving towards explaining the evolution of language using processes beyond biological evolution alone. Such processes include joint action in a social context (Di Paolo, 1997; Steels, 2012), cultural transmission (Smith et al., 2003; Oudeyer, 2005) and self-organisation (Oudeyer, 2013).

Many of the approaches mentioned above are collectively referred to as forming the field of *experimental semiotics*: an investigation of "the emergence of novel forms of communication in the laboratory" (Galantucci, 2009; Galantucci and Garrod, 2011). The field's focus on experiments leads to an appealing ability to flesh out and test spe-

cific hypotheses; the focus on novel forms of communications means that experimental semiotics approaches are easily applicable to AI and robotics research. Furthermore, the study of communication in the context of its use —joint action— is especially relevant for the aim of this project: the investigation of the evolution of language using an autopoietic framework.

In this section we are going to examine two of the most influential models in the experimental semiotics field: the Iterated Learning Model (Kirby, 2002b) and the Talking Heads experiments (Steels, 2003). The Iterated Learning Model ("ILM") uses the fact that language is culturally transmitted from generation to generation in order to explain how important structural properties of language can evolve without the need for a "language organ" in the speaker's brain; the Talking Heads experiments use a number of agents taking part in language games to study how meaning is created, coordinated and assigned to symbols in a social context.

2.2.1 Iterated Learning

The main idea behind the Iterated Learning Model (Kirby, 2002b) is that the features of the natural languages that we use cannot be fully determined only from the biological cognitive basis of language in humans. Language is in itself a system that evolves, and this evolution shapes its structure (grammar) and its features. Instead of evolution by biological reproduction, languages evolve by being *culturally transmitted* from generation to generation of linguistic learners and teachers. Instead of natural selection, languages are selected for their *learnability* to young, developing minds. Brighton and Kirby (2005) sum up the ideas behind Iterated Learning in the following three principles:

Principle 1 (Innateness hypothesis): This principle states that the cognitive basis for language in humans is biological. Biological evolution has equipped humans with a number of skills, not necessarily specific to language, that allowed the emergence of language.

Principle 2 (Situatedness): This principle states that the cognitive basis mentioned in Principle 1 *underdetermines* the structure of language. In order to have a complete explanation of language structure, we must also take into account the historical process of language evolution through cultural transmission and selection.

Principle 3 (Function independence): This principle states that we do not need to take into account the *function* or *use* of language in order to explain at least some of its structural features.

Another important aspect of the ILM methodology is the enforcement of a *bottleneck* on language transmission. The presence of this bottleneck mirrors one of the aspects of the poverty of the stimulus argument mentioned above: the fact that when learning language, learners are exposed to only a part of all possible grammatically correct sentences. The bottleneck is however doubly important: it also puts a selective pressure on languages to be learnable. According to the ILM, this pressure for learnability is exactly what leads to the evolution of the structure of language, leading to the eloquent statement that "the poverty of the stimulus solves the poverty of the stimulus" (Zuidema, 2003).

Finally, ILM experiments do not provide language learners with feedback on their use of language. Developing language learners have no negative feedback, so no indication of when they uttered a non-grammatical sentence (Zuidema, 2003, pp. 1-2); by not including any linguistic feedback, the ILM makes sure it conforms to the constraints of the "poverty of the stimulus" argument.

2.2.1.1 Simulation experiments

The initial ILM experiments were performed using computer modelling and simulation. An overview of early simulation-based ILM is given in Kirby (2002b); a simulation model that produces compositional languages is presented in Smith et al. (2003). Smith et al.'s model has the following components:

1. A meaning space: This is a collection of meanings, each meaning being represented as a point in a discrete multi-dimensional space. A single meaning can have multiple components; this allows the model to encode structure in the meaning space.

2. A signal space: This is a collection of signals, strings of characters drawn from a specified alphabet. Each of the meanings is initially assigned a random signal.

3. A number of agents. Each agent is modelled as a network with two sets of nodes; one set represents meanings, either complete (with all their components) or incomplete (where some components are missing). The other set represents signals. All meaning-nodes are fully connected to all signal-nodes and vice versa.

4. Learning happens by changing the weights of the connections between the nodes that correspond to the meaning-signal pair observed.

5. Production of a symbol, when prompted with a certain meaning, happens by picking the meaning-signal pair with the highest sum of weights from the agent's network.

A similar production mechanism called the *obverter method* was also used in previous ILM computational models (for example Kirby, 2002b). Kirby is more explicit about the assumptions of the production mechanism, stating that "the idea behind *obverter* is for a communicative agent to produce signals that maximize the chance of the hearer understanding the correct meaning" (Kirby, 2002b, p.124) and that "the speaker's own mapping approximates that of the hearer" (Kirby, 2002b, p.125).

Returning to Smith et al.'s model, the simulation world includes a number of learning agents and a number of teaching agents. Teachers produce a specified percentage of the whole language for learners to learn; by changing that percentage, a bottleneck of varying width can be introduced in the system. After a specified number of learning events, learners become teachers themselves and new learning agents are introduced in the simulation world. The population number is kept stable by removing an "old" teacher for each new learner introduced. By running the simulation for a number of agent generations, the initial (random) language evolves; the resulting language can then be checked for elements of structure. The specific structure of interest in this experiment was compositionality. (A complex structure is *compositional* when its meaning is a function of the meanings of its parts; Smith et al., 2003, p. 372. Natural languages are compositional, as for example the meaning of the utterance "come here" is a function of the meanings of its parts, "come" and "here".) The degree of compositionality was measured by the correlation of two distances: the distance between two meanings and the distance between their paired signals.

In Fig. 2.3, we can see the mean results of 1000 simulation repetitions for two different experimental cases. In the first case, there is no bottleneck: each teaching event includes all of the language pairs. The result (Fig. 2.3a) is that most of the resulting languages are holistic: meaning-signal pairs with similar "signals" do not necessarily have similar "meanings", so the meaning and signal distances are not correlated, resulting in low compositionality scores. In the second case, there is a narrow bottleneck: only 40% of the total language pairs are produced by the teaching agent. The result (Fig. 2.3b) is very different; a significant percentage of the resulting languages are

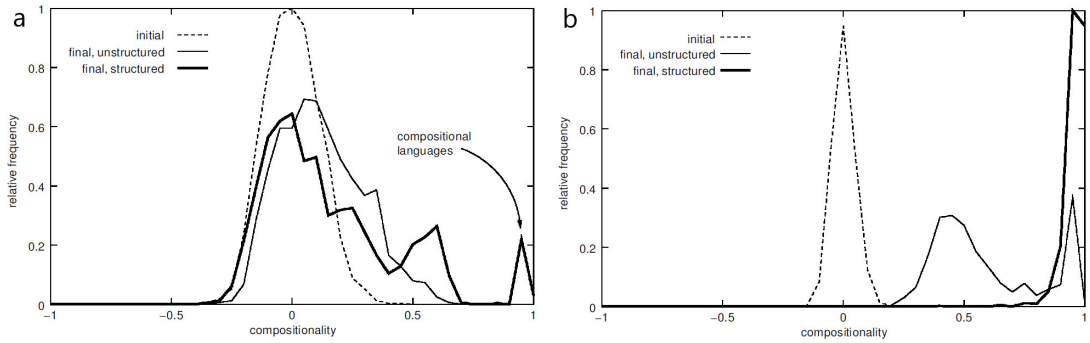


Figure 2.3: Results for the ILM simulation experiment in Smith et al. (2003). A simulation without a bottleneck (a) results in almost all languages being holistic; a simulation with a bottleneck (b) results in most languages being compositional, especially when the meaning space is structured.

compositional. Furthermore, when the meaning space is structured (bold continuous line) this is the case for almost all languages.

A number of simulation ILM experiments (Griffiths and Kalish, 2005, 2007; Kirby et al., 2015) also use mathematical (Bayesian) models of transmission chains instead of simulated agents.

2.2.1.2 Human experiments

In addition to computer modelling experiments, the ILM has been applied to human experiments in the lab (Kirby et al., 2008). The participants were asked to learn an alien language; as in the simulation experiments, the language is as a set of associations between a meaning space and a signal space. The experimental design was the following:

- Each meaning was a picture of an object with a certain shape, colour and movement, creating a three-dimensional meaning space.
- Each signal was a label (a text string). The initial values for these labels were random.
- The language, a set of pictures of objects (meanings) and labels (signals), was initially randomly divided in two sets: SEEN and UNSEEN. Participants were first trained on the language using the SEEN set only; after this training period,

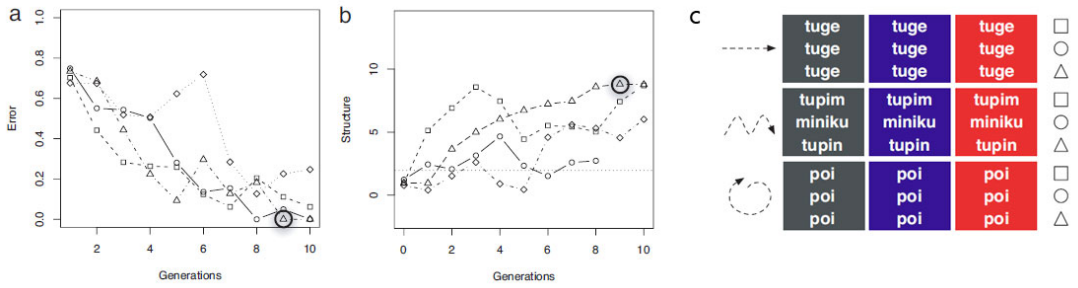


Figure 2.4: First ILM human experiment. Error rate decreases (a) and measure of structure increases (b) as the language is transmitted from generation to generation. However, the resulting language is not expressive; the same signals are used for multiple meanings (c) (adapted from Kirby et al., 2008).

they were tested on all sets, both SEEN and UNSEEN. This models the bottleneck aspect of language learning.

- Finally, the language *produced* by each participant was randomly divided once more into SEEN and UNSEEN sets and used to *train* the next participant. The participants were not aware of this; this both models the cultural transmission of language from generation to generation and adheres to the "no feedback" requirement that we mentioned earlier.

An analysis of the error rates of the participants as well as a measure of the structure of the resulting language (based on the correlation between the similarity of signals and the similarity of meanings, compared to the equivalent correlation of a random language) can be seen in Fig. 2.4a and 2.4b. We can see that with each successive generation, the error rate tends to decrease and the measure of structure tends to increase, indicating that the language becomes progressively more structured and at the same time more learnable. On a closer analysis of the resulting language, however, it turns out that the structure that has evolved is "*systematic under-specification*"; the language expressivity has collapsed and the same signals are used for multiple related meanings (Fig. 2.4c). This allowed learning to successfully generalise the language in the presence of a bottleneck.

Kirby et al. consequently designed a second experiment identical to the first except for the addition of an extra (hidden) constraint: after each participant produced a language and before that language was used to train the next participant, any change that led to under-specifying languages was discarded; in that way, no two meanings

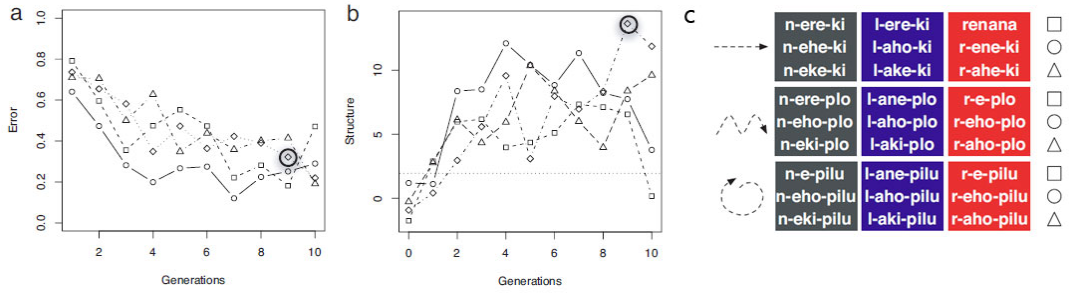


Figure 2.5: Second ILM human experiment; this time, language expressivity is retained using an extra hidden constraint. Once more, error rate decreases (a) and measure of structure increases (b) as the language is transmitted from generation to generation. This time, the resulting language is compositional (c) (adapted from Kirby et al., 2008).

with the same label were presented to the participants, ensuring that language expressivity was always maintained. The results of this second experiment were similar to the first one: with each successive generation, the language becomes more learnable and more structured (Fig. 2.5a, 2.5b). This time, however, the resulting language is highly compositional (Fig. 2.5c).

2.2.1.3 Further experiments

Thanks to the power of transmission chains, the Iterated Learning approach has led to a very fruitful research programme. Since the experiments detailed in the previous sections, there have been a plethora of others (for a recent review, see Tamariz and Kirby, 2016). We will mention some below.

Graphical signals: A series of studies (Garrod et al., 2007; Fay et al., 2010) use graphical instead of symbolic communication systems. In contrast to the previous Iterated Learning experiments we presented, the transmission chains in this case are "horizontal" (*intragenerational*) instead of "vertical" (*intergenerational*). This means that the chains do not involve learning episodes between trained and naive individuals, but repeated interactions amongst a population of pairs of participants. These pairs were asked to communicate a certain meaning using drawings; while initially the variation in the drawings used was very high, after a few episodes of transmission the drawings converged to those that were simpler and more abstract, an indication of a move from iconic to symbolic signs (Garrod et al., 2007, p. 983).

In studies with smaller populations, Fay et al. (2010) reported similar findings: the drawings the pairs used to communicate became progressively more *symbolic* (harder to understand for observers not belonging to the population) as opposed to *iconic* (closely reflecting the meaning to be communicated). A characteristic example of this switch from iconic to symbolic for the concept of "Brad Pitt" is shown in Fig. 2.6.

Theisen-White et al. (2011) expanded on this paradigm by using pairs that were trying to communicate a meaning with drawings and then passing the resulting communication systems from generation to generation of pairs. With this integration of vertical and horizontal transmission, the drawings used to communicate started exhibiting structure.

Continuous signals: Instead of using discrete signals for meaning communication, Verhoef et al. (2011, 2012, 2014) used continuous signals produced by a slide whistle without any meaning association. The use of whistling instead of human speech allowed for a continuous medium while also minimising the impact of speech and language experience. Each participant listened to a set of 12 whistle sounds and had to memorise and play them back; this output was then used as the input set for the next participant (after being filtered for "no duplicate whistles" as an artificially enforced measure of expressivity). After a chain of 10 generations, the initially structureless whistling space became structured. As 12 different structureless continuous sounds are very difficult to remember and reproduce, participants make use of basic "building blocks" which they combine in different ways (Verhoef et al., 2014). Once more, evolutionary pressures for learnability and expressivity combined with a transmission chain lead to structure from an initially structureless space.

Animal experiments: Finally, a few studies focus on cultural transmission chains in non-human animals. Horner et al. (2006) set up an experiment with a population of chimpanzees. In the experiment, an initial model from the population was trained to use one of two methods ("lifting" or "sliding") to retrieve a reward from a box. A second chimpanzee observed them retrieving the reward from the box a number of times, then was given access to the box to try and retrieve the reward themselves; if they were successful, they served as a model for the next chimpanzee. Repeating this procedure leads to a chain of learning by observation from generation to generation. Horner et al. found that the initial behaviour was

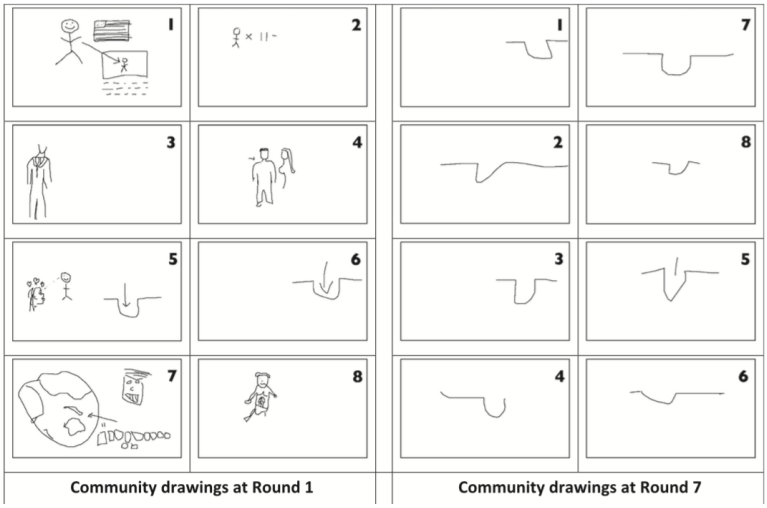


Figure 2.6: The drawing representing "Brad Pitt" in a pictorial communication game, initially iconic, becomes symbolic after some generations of horizontal transmission (adapted from Fay et al., 2010).

transmitted along this chain with high fidelity, showcasing a transmission chain in a population of chimpanzees that is very similar to the chains used in the ILM experiments (Fig. 2.7).

Claidière et al. (2014) used a population of baboons in a more traditional "ILM-style" chain. An initial model was shown a set of 50 random patterns that they had to reproduce on a grid of buttons (with a reward for successful or close to successful reproductions). Each animal's output patterns were then used as inputs for the next animal, with the process repeated for 12 generations. As was the case in the experiments we detailed in Section 2.2.1.2, both the rate of successful pattern reproduction and the measure of structure in the transmitted patterns increased by the end of the chain. This provides support to the claim that it is not only human-specific biological traits that determine the structure of language: the process of Iterated Learning, even in non-human animals, leads to structure in initial unstructured stimuli.

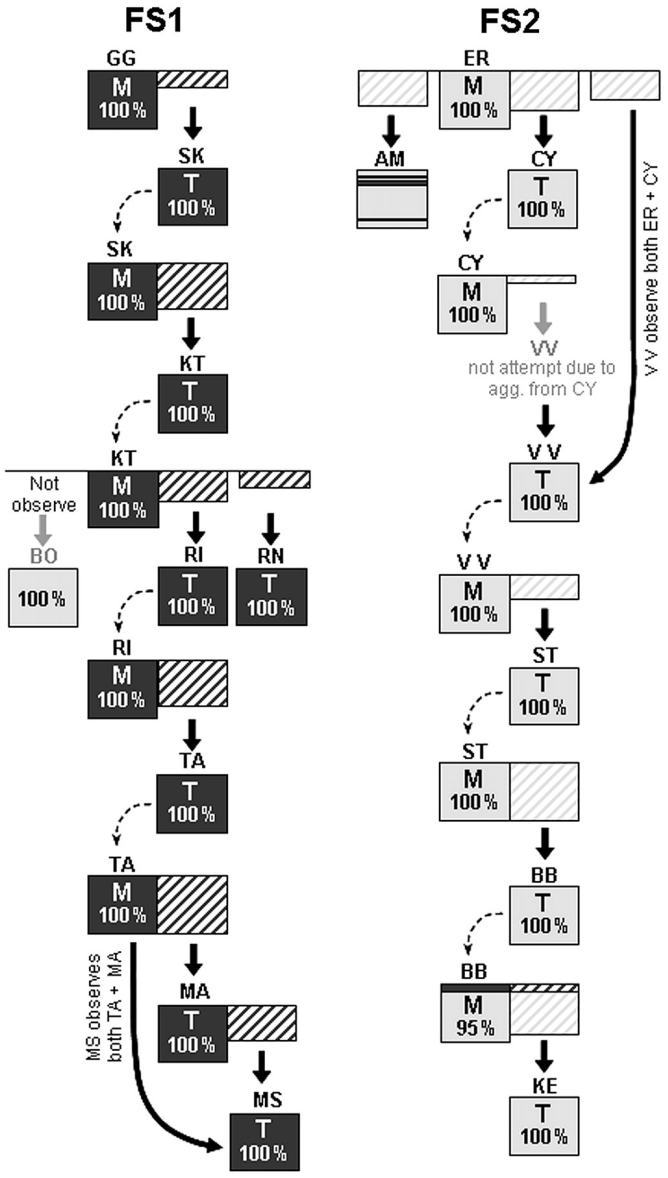


Figure 2.7: Two cultural transmission chains in chimpanzees (Horner et al., 2006). An initial model (GG, ER) was trained to use one of two methods ("lifting" in chain FS1, "sliding" in chain FS2) to retrieve a reward from a box. A second chimpanzee observed them retrieving the reward a number of times, then allowed to try and retrieve the reward themselves. Repetition of this procedure led to the maintenance of the initial model's behaviour for 6 generations in cultural transmission chains.

2.2.2 Talking Heads

The approach the ILM takes to studying the evolution of language is often called a *vertical* approach, referring to its focus on the transmission of language from generation to generation. The respective *horizontal* approach would then refer to the transmission and use of language for joint action within the same generation of a society (Theisen-White et al., 2011). One of the research projects that focuses on this approach is Luc Steels' Talking Heads (Steels, 2003).

Luc Steels' project, instead of studying the evolution of language structure, models the grounding, self-organisation and evolution of *meaning* in a society of agents taking part in "language games". In these games, populations of agents (robotic or simulated) interact with their environment and each other by making use of one or more sensorimotor systems. In their mutual interactions, the agents take turns producing or receiving speech signals in order to take part in joint action with specific scripts. One example of a language game is the *guessing game* (Steels, 2001). In this game, the speaker tries to draw the listener's attention to an object in the environment by, for example, pointing at it and "naming" it (producing the symbol they have associated with it).

Language in the Talking Heads experiments is, once more, an association between meanings (called here "representations") and signals (called here "symbols"). In addition to these, there must be a way to rate the associations; language is then a set of triplets ($< r, s, k >$: representation, symbol, past score). After each iteration of the game, a feedback loop updates the rating of associations. The steps to play the game are the following (taken from Steels, 2001):

1. Shared attention: The speaker somehow draws the listener's attention to a topic.

2. Speaker behaviour: The speaker conceptualises the topic (for example, if the environment only has one blue item, item colour is a good way to refer to the topic) and creates a representation. The symbol s that corresponds to the representation formed with the highest score is the best word to communicate; it is transformed into a speech signal.

3. Listener behaviour: The listener receives the signal (and symbol) and looks up all associations; if there is no association, a new word is created. Otherwise, the listener applies all representations to see if any of them yield a unique topic; if found, this is the topic the listener selects.

4. Feedback: the listener's selected topic and the speaker's original topic are compared; if they match, association scores for the triplets selected are increased, otherwise decreased. If there is no match, the process is repeated (with the decreased scores).

5. If either of the agents fail to conceptualise the scene, a concept-acquisition algorithm is triggered.

The result of the guessing game and similar experiments is the replication of a number of features of natural languages, like homonymy (Wellens et al., 2008), synonymy and ambiguity of meaning (Steels and Kaplan, 1999). The Talking Heads experiments are quite adjacent in spirit to embodied cognition models, as they adopt a "whole systems" approach to the study of the evolution of language. As the focus is on the evolution of meaning, the syntax and structure of symbols are arbitrarily instantiated and do not evolve; for this reason they are a good candidate for combination studies with ILM models (Vogt, 2005).

While we have only outlined one of the language games (the "Guessing Game"), the Talking Heads experiments were not limited to that. The "Discrimination Game", for example, focused on researching mechanisms of the emergence of categorisation, while the "Naming Game" studied how a certain meaning is grounded in an external object. Furthermore, these original experiments have since branched out to multiple research directions, making use of more complex grammar constructions, more powerful physical robots or studying the evolution of language syntax and structure in addition to the evolution of meaning. For a comprehensive picture of this very rich research program, see Steels (2015).

Despite the focus on joint action and a more systemic approach compared to the Iterated Learning models, the Talking Heads experiments are less suited for an autopoietic exploration of the evolution of language. The reason for that is that the main concern of these experiments is to explain the evolution of *meaning*; as we saw in Section 2.1.2.4, this is a sensitive area for both Maturana's original autopoietic theory and the more recent enactivist approaches. The account of meaning used here is not compatible with either approach. For this reason, in the rest of this chapter we will focus on discussing a combination of autopoietic principles with Iterated Learning.

2.2.3 Discussion

What we can see from both the simulation experiments and the human experiments is that the ILM presents a convincing argument against the claim that language structure is innate. The poverty of the stimulus argument, one of the main arguments in favour of the nativist view, is turned on its head as the experiments show that compositionality evolves in language *because* of the poverty of the stimulus and not *in spite* of it.

This result, in addition to being very significant for evolutionary linguistics, seems to suggest a tentative first step in the attempt to bridge the "cognitive gap" to language that we discussed previously. In the rest of this section, we will try to detail why this is the case by drawing attention to the differing approaches to "language" in the theory of autopoiesis and the Iterated Learning model; and expand on the challenges that occur from attempting to combine the two approaches.

2.2.3.1 Language in Autopoiesis and the ILM

Language is at the same time,

(L1) A behaviour that agents take part in ("languaging" in autopoietic terminology),

(L2) An external system of symbols and meanings that agents use in their languaging behaviour, transmitted from generation to generation of linguistic agents.

Autopoiesis places most of its explanatory focus on language as an action (L1): "linguistic coordination of linguistic coordination". An agent *languages* when they use learned communicative behaviour (see section 2.1.1.5) to coordinate *their communicative behaviour itself*. This self reference [1] is what leads to the rise of the observer and post-linguistic phenomena; it also implicitly points to the nature of language as an external system (L2).

What autopoiesis does not attempt to explain, however, is the transition from *first-order* communicative behaviour to *second-order*, self-referential languaging (with the features of language as we know them). In other words, it takes no part in the debate about the evolution of language despite the fact that, as we saw, language is a pivotal component of autopoietic theory. This is where Iterated Learning, covering exactly this explanatory gap, is of particular importance for the theory of autopoiesis; the goal

[1] This is a form of recursion, albeit not in the way that this term is used in linguistics — see for example Chomsky (2014).

being that with a proper selection of task and an initially random, non-self-referential language, we can show that a language with self-referential features evolves.

Iterated Learning, however, approaches language from the viewpoint of an external observer (L2), without placing too much focus on language as a social action or on what the action of languaging is used for ("function independence" principle, see section 2.2.1). In order to even take a first step towards the goal we stated, then, we need to re-frame the experimental setup that Iterated Learning experiments use to one that is compatible with the constraints that we detailed in Sections 2.1.2.1 and 2.1.2.2. This presents a number of challenges that we will try to detail below.

2.2.3.2 Choice of working domain

The most obvious problematic point in this re-framing is that in the ILM, "language" is defined as a set of symbol to meaning associations. This definition of language, while useful for the purpose of most ILM experiments, is provided in a domain of descriptions, not in an operational domain. It originates from an analysis of language from the point of view of an external observer, and since we want to explain how language evolved in societies of agents in the first place, we need to work in an operational domain. In such a domain, language is not a collection of words and meanings but a specific behaviour that agents take part in, "languaging". A number of other perceived incompatibilities are mostly just extensions of this choice of domain to work in.

2.2.3.3 Function independence and systemic approach

According to the "function independence" principle, we can explain some of the structural features of language without taking into account what that language is used for. In the ILM experiments that we detailed, there is no definition of an explicit purpose or function of language. There are two problems with this approach, the first being that it is not compatible with the systemic perspective that we adopted in Section 2.1.2.2, that states that any explanation of behaviour must take into account the interplay of agent and environment, as that domain is where behaviour is defined as a term. If we want to examine any aspect of languaging as a behaviour, we need to take into account what function that behaviour accomplishes.

The second problem is that regardless of our autopoietic perspective, agents using language are *always* performing an action. That is true for any Iterated Learning experiments as well: taking part in an ILM experiment *is* a joint action between a subject

and an experimenter, and the use of language in this joint action context cannot lack a function. In most ILM experiments, this hidden function is "reference".

In the case of the human experiments, the language game played by the participants is clearly one of reference: they are asked to look at pictures and learn (then produce) a caption for them. Also, the extra constraint of preserving language expressivity used in the experiments we detailed is an inherently functional constraint: an under-specifying language is not useful in the sense that it cannot *function* for a linguistic game of reference. In the same way, the "obverter function" used in the simulation experiments is directly modelling the use of language for object reference. We see then that despite the "function independence" principle, language does have a function in the ILM experiments, even if that function is implicit rather than explicit. Additionally, the externally enforced constraint on expressivity in that experiment has its roots in the use of language as a tool for object reference.

The importance of expressivity as a functional constraint is also highlighted in Kirby et al. (2015), which explicitly adds dyadic communication (and thus, a functional expressivity pressure) to Iterated Learning experiments. Using either vertical transmission chains of agent pairs (leading to high learnability pressures on the language the pair uses to communicate) or closed groups of horizontal transmission (leading to low learnability pressures), Kirby et al. examine the effect of learnability and expressivity pressures on the structure of the resulting language.

The results from that study show that in both simulation and human models, the emergence of compositionality in language requires a pressure for compressibility or learnability (coming from the process of inter-generational chain transmission) *and* a pressure for expressivity (coming from the functional need of intra-generational communication). Only one or the other pressure is not enough, and leads to either unstructured *holistic* languages (one-to-one symbol to meaning correspondence) or functionally useless *degenerate* languages (one-to-all symbol to meaning correspondence). This makes one of the implicit assumptions of the initial ILM experiment we reviewed earlier (the functional need for expressivity) explicit, and in doing so establishes its importance for the emergence of compositionality.

2.2.3.4 Object reference

Object reference as the function of language is directly tied to the view of language as a collection of meaning and symbol pairs, as *meanings* are object *representations* in the agents' minds. Transferring these pairs from generation to generation is transferring *information* about the language, completing the list of "forbidden" explanatory concepts we gave in Section 2.1.2.1.

However, trying to replace reference as the function of language is problematic. As it is hard to define what a pre-referential language would look like, almost all of the Iterated Learning experiments involve some sort of reference. Verhoef et al. (2012) is an exception to this rule, as that experiment uses signals without any meaning association; as the initially continuous signals go through transmission chains, the signal space becomes structured and combinatorial. However, the resulting combinatorial structure is conceded to have more to do with human biases rather than any functional pressure, so an autopoietic account of this phenomenon needs to both precisely know what those biases are and explain them in operational terms.

A more realistic and theoretically compatible first step for an autopoietic explanation of language evolution would instead draw inspiration from the animal experiments we mentioned in Section 2.2.1.3. Indeed, a chain of behavioural transmission like the one shown in Horner et al. (2006) is not necessarily based on object reference and has no explanatory need of concepts of meaning or representation. We still need (minimally) two different behaviours and an account of learning; we will discuss these in Chapter 3.

Another potentially relevant experiment, this time from the field of evolutionary robotics, comes from Quinn (2001), who studied the evolution of communication in simple robotic agents that had no pre-set communication channels. For his experiment, Quinn used pairs of simulated Khepera robots, who have a circular body with a number of short-range proximity sensors and two wheels for differential movement. The robots were placed in a random orientation but in range of each other, and needed to move a certain distance in a limited time while staying in range. This means that they had to move in the same direction; however, accomplishing this is not trivial. If one of the agents takes initiative and moves first in a certain direction, the two agents end up out of range and fail the task. Furthermore, as the agents are circular, there is no way to determine the other agent's orientation.

Quinn used genetic algorithms to evolve a neural network controller for the robots

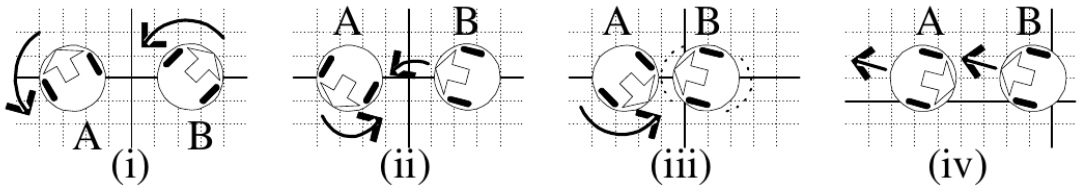


Figure 2.8: Simple physical task for the evolution of communication without a dedicated communication channel (Quinn, 2001).

that came up with a solution to this task, shown in Fig. 2.8. The solution involved the following steps:

1. The two agents start rotating counter-clockwise, until one of them "sees" the other agent in their frontal proximity sensor. In the scenario shown in Fig. 2.8, this is agent B.

2. Agent B approaches agent A and remains close, "jiggling" back and forth.

3. Agent A "sees" agent B jiggling in their front sensor.

4. Agent A reverses their direction and starts moving. Agent B follows agent A.

What is happening here is intriguing: a simple communication system has evolved between the two agents. A formerly non-communicative behaviour ("jiggling") is transformed into a signal that essentially means "I am waiting for you to lead and I will follow"; this allows the two agents, completely identical otherwise, to break the symmetry and adopt a specific role, thus escaping a deadlock situation and solving the task.

Quinn's experiments point us to an example of a task that, similar to the chimpanzee retrieval task of Horner et al. (2006), does not make explicit use of representations, meaning or object concepts. Again, we will come back to the discussion of a similar task that we will use for our experiments in Chapter 3. It bears clarifying that by using a simple, non-referential task we are not making an attempt to explain any of the structural features of language. Instead, we are only trying to show that it is possible to build a cultural transmission chain on a non-representational substrate. As we saw in Section 2.2.1, cultural transmission chains are of paramount importance for explaining universal features of language; with an autopoietic implementation of such a chain we would be opening the way for a more thorough non-representational exploration of the evolution of linguistic systems.

2.3 Summary & claim

So far, we can summarise our argument as it appears in Chapters 1 and 2 as follows:

1. As opposed to the view of cognition as computation, non-representational cognitive science sidesteps the "grounding problem" that the traditional approach faces and offers a bottom-up, systemic way of examining cognitive phenomena. Within the non-representational camp, *autopoiesis* is a biologically grounded theory of cognition particularly attractive because of its completeness and explanatory power.

2. Autopoiesis (and non representational cognitive science in general) struggle with scaling up to higher level cognitive behaviour such as planning and conscious thought.

3. In autopoietic theory in particular, language could help bridge the gap to representational cognition as the explanatory concepts of meaning, reference and representation are constituted in the domain of language. However, a major challenge remains because the emergence and evolution of language as a system still needs to be explained without having access to these explanatory concepts.

4. The Iterated Learning framework provides a powerful account of language evolution through cultural transmission while removing the focus from the agent and the brain in a way that holds promise for non-representational accounts of cognition.

5. Cultural transmission of behaviour is a fitting first step for an autopoietic exploration of language evolution through Iterated Learning. While being a prerequisite (and forming the basis) for Iterated Learning, it also does not require an account of meaning or reference.

Our claim then can be presented as follows:

It is possible to build a system that is both,

1. Consistent with the non-representational, autopoietic design principles that we outlined in Section 2.1.2.2 and,
2. Able to support a chain of cultural transmission of behaviour.

In the next chapter we will detail the design of a system that will allow us to examine this claim.

Chapter 3
Design

In Section 2.3, we made the claim that we can build a behaviour transmission chain on a system that is based on autopoietic principles. In this chapter, we will attempt to detail the design of such a system. In essence, three components are required for a transmission chain: a number of *agents* (the units between which transmission happens), an account of *learning* (the process of transmission itself) and a *task* (which determines the behaviour to be transmitted). These components will be covered in the first part of this chapter, while in the second part we will describe the overall design of component interconnection and the experimental setup.

3.1 Task design

3.1.1 Animal task

The starting point for a task that is simple, yet ecologically valid and proven to be transmissible in the wild comes from the chimpanzee cultural transmission experiment described in Horner et al. (2006). As we mentioned in section 2.2.1.3, each of the chimpanzees in that experiment had the task to retrieve a reward from a closed box which could be opened in two different ways. The fact that there were two possible, equally valid ways to solve the retrieval task is important because it means that it is possible to track the transmission and maintenance of each behaviour in cultural chains.

The problem with using a similar task for our system is that this is a single agent task: it relies on observational learning rather than any form of joint action. If we want our system to be a "first step" in an autopoietic account of language evolution, or more generally to be relevant for the study of the evolution of systems of communication,

we need to pick a joint action task that is potentially communicative.

3.1.2 Robot task

Such a task — both joint action based and potentially communicative — is used in the evolution of communication experiments (Quinn, 2001) that we detailed in section 2.2.3.4. The robots in that experiment had to solve the task of moving their combined center of mass in a maximum distance while not losing track of each other. Since each robot had no way to sense the other robot's orientation, the solution to this task (found by a genetic search algorithm) involved the evolution of signalling behaviour between the two robots that allowed them to adopt different leader and follower roles and successfully move together.

The problem with Quinn's specific task for our scenario is twofold. Firstly, it involves evolutionary search, not learning algorithms; the robot's signalling behaviour is completely innate and evolved. The time scales we are looking at for cultural transmission chains are not phylogenetic but ontogenetic (or rather, "glossogenetic": see Kirby, 2002b, p. 122): we want learning, not evolution. Secondly, it only involves a single behaviour and as we saw we need at least two different behaviours to examine their cultural transmission in chains.

3.1.3 Task criteria

Stepping back, we are looking for a task that meets the following criteria:

Joint action: The task needs to be based on joint action and coordination, so that it is potentially communicative and relevant to the study of linguistic systems.

Pre-communicative basis: The task needs to be based on pre-existing actions that are not necessarily communicative or even social in nature; this adds to the ecological plausibility of our system as it provides a potential explanation for the evolution of communication (Quinn, 2001, p. 358).

Transmissible: The task needs to provide some kind of feedback so that it is learnable (and teachable), so transmissible from generation to generation. Note that this does not contradict the avoidance of *linguistic feedback* that we mentioned in Section 2.2.1; *task feedback* is unavoidable and present in most learning, including human language acquisition.

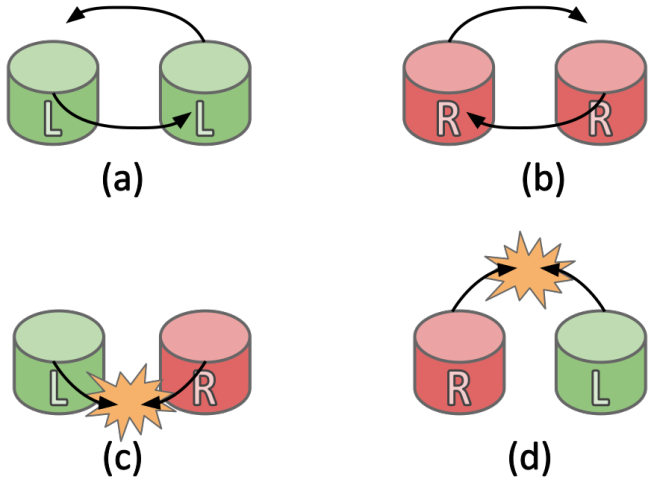


Figure 3.1: Two agents in the Left/Right dancing task. If both agents turn in the same direction (a, b) then they "dance" and successfully complete the task; if they turn in opposite directions (c, d) then they crash and fail the task.

Not binary: There must be more than one way or behaviour that allow our agents to succeed at the task; this is what makes it possible to examine a potential cultural transmission chain.

Potentially evolvable: Ideally, we would like the task to create the potential for an evolvable communication system; while this is not strictly necessary for a cultural transmission chain, it would mean that the task is complex enough to enable future investigation of the evolution of linguistic structure.

3.1.4 L/R dancing task

The **Left/Right dancing task** is a very simple joint action task that fulfils all of the above criteria. It is based on the non-communicative action of a single agent "turning around" an object, either counter-clockwise ("Left" direction) or clockwise ("Right" direction). This simple action, when put in a joint context by replacing the object with another agent, has a number of possible outcomes depending on whether each agent is dancing "Left" or "Right", as shown in Fig. 3.1.

If both agents, after they detect each other in proximity, try to turn around each other in the same direction (L-L or R-R), then they successfully "dance", either statically or in a moving spiral, depending on the *turning* behaviour implementation (Fig. 3.1a, b). If they try to turn around each other in opposite directions (L-R or R-L), then

they crash, failing the task (Fig. 3.1c, d). In essence, this is a physical, joint action implementation of an "XNOR" logical operation, only returning "success" if both of its inputs are the same.

The L/R dancing task fits the first four criteria we presented in the previous section (3.1.3): it is a *joint action* task between two agents, based on a behaviour ("turning around") that is *pre-communicative*; it is *transmissible*, as it can be learned by connecting the success or failure at performing the task with a reward; and it is *not binary* (completed or not completed), as it can be completed in more than one way (Left-Left or Right-Right turning both lead to successful dancing).

3.1.4.1 Relevance to language

As described, the L/R dancing task does not use signals or communication: it is a purely coordinative task. Our goal, as stated in section 2.3, involves taking a first step towards an autopoietic account of language evolution; is a non-communicative task of coordination even relevant as such a first step?

Going back to the theory of autopoiesis, coordinated behaviour in a social context is *communicative* behaviour; and ontogenic (in other words, learned as opposed to innate) communicative behaviour is behaviour in the linguistic domain, forming the basis for languaging behaviour (see Section 2.1.1.5). Coordinative behaviour, in this view, is definitely the first step towards an explanation of language.

What is missing from the L/R dancing task to take the next step into a more linguistic domain is some sort of signalling behaviour; while this might make learning it a lot more challenging, there are several ways in which the L/R task can be modified towards this direction. As an example, with the addition of agents whose turning direction depends on random or external factors, it is possible to introduce the need for signalling and a potentially evolvable communication system. We will come back to this with further details in section 7.7.1.

3.2 Agent design

Now that we have a task for our agents to transmit from generation to generation, we need to select what kind of agents we will use. The dancing task points to robot agents, either physical or simulated, rather than humans or animals. This is in line with the bottom-up approach that we have adopted: human subjects already have language and learning skills, and we are not trying to study those in humans but instead to "understand them by building" (as per the *synthetic methodology* of Pfeifer et al., 2005, p. 101). But should our agents be physical robots or simulated ones?

3.2.1 On robots versus simulation

Computational modelling experiments ("simulations") have the advantage of being much less demanding to work with. It is easier to set up a simulation, easier to control all aspects of both the environment and the simulated agents and easier to collect and evaluate any results. By not having a mandatory constant "connection" with the real world, it is easy to test different hypotheses or parameters and it becomes possible to analyse time scales relevant to biological evolution. In contrast, robotics experiments are hard to set up, hard to control and dependent on a number of functional or pragmatic constraints that are not present in computational modelling.

On the other hand, Loetzsch and Spranger (2010) give a convincing account of why actual robot experiments are preferable to computational modelling in the field of language evolution. Their four arguments can be summarised as follows:

Increased realism: Assumptions about how an agent's interaction with its environment happens are often unrealistic when tested out in the real world. By forcing researchers to actually handle this interaction and make it work, the resulting model is more realistic.

Robust models: A lot more can go wrong in robotics compared to computational models. Accounting for all the details that are abstracted away from in a simulation (like sensor noise or unexpected events) makes robotic models more robust.

Rich semantics: The real world is an extremely rich source of interaction for an agent. This richness of semantics is in part what drives the richness of natural language. Incorporating a subset of this richness in a simulated environment is exponentially difficult.

Figure 3.2: A physical e-puck robot (Mondada et al., 2006).

Evolutionary pressure: There is an increasing amount of evidence showing that language concepts are bodily grounded. In a simulation it is easy to abstract away from this bodily grounding, losing a potentially important force that drives language evolution.

In practice, however, many of these advantages are negated by the use of carefully constructed environments in order to keep the practical issues tractable; this was our experience as well in some initial experiments with physical robots. In the end, we settled for a compromise: all of the experiments described in this thesis were done using simulated robots, but our system architecture and simulated robots were chosen so that our model would be easy to extend to physical robot experiments (see section 7.7.4).

3.2.2 E-puck robots

The robots we chose to use were "e-pucks" (Fig. 3.2), a small but powerful robot designed for education and research (Mondada et al., 2006). Despite their small size and cost, e-pucks have a number of sensors (infrared proximity, accelerometer, microphones, camera) and actuators (motors, speaker, LEDs). Of those, the ones relevant for us are the motors that drive two individually controllable wheels, allowing for differential drive locomotion; and the array of 8 infrared proximity sensors, allowing for detection of nearby objects. The robots have a built in dsPIC microprocessor, so they can operate autonomously, but a Bluetooth connection also opens the possibility of remote control.

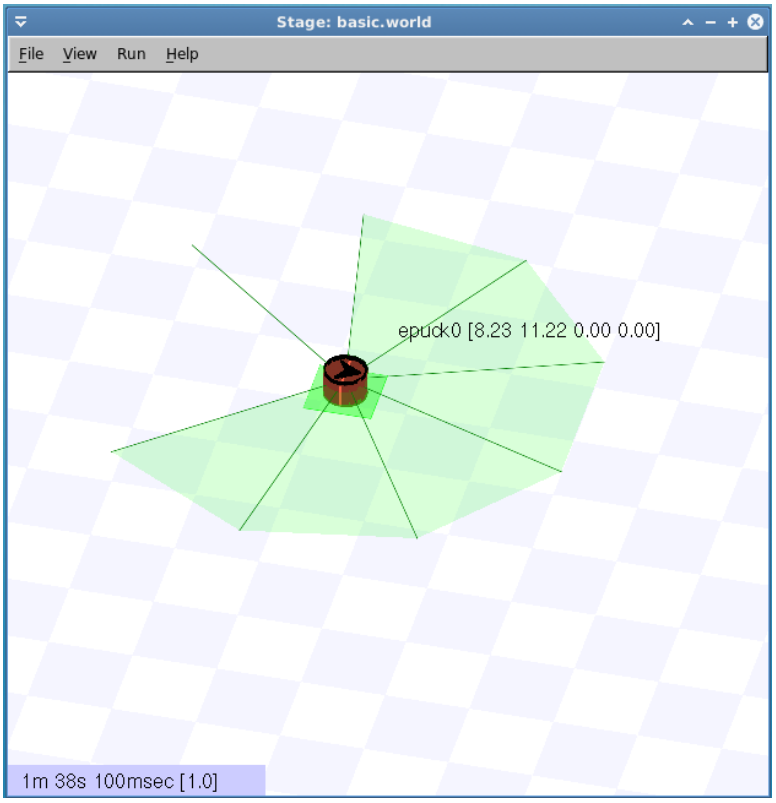


Figure 3.3: An e-puck robot simulation in Stage. Only the proximity sensors and the differential drive motors are simulated.

There also exist a number of software simulators that support e-puck robots, most notably Webots (Michel, 2004) and Enki (Magnenat et al., 2011). We used "Stage" (Vaughan, 2008) because of its built-in connectivity to ROS ("Robot Operating System", Quigley et al., 2009) which allows for seamless substitution of the e-puck computer simulation with a physical e-puck robot. We will come back to this point in section 3.6. Figure 3.3 shows a simulated e-puck in Stage; the simulation only includes the peripheral proximity sensors and the differential drive motors, as these are the systems we made use of for the experiments described in this thesis.

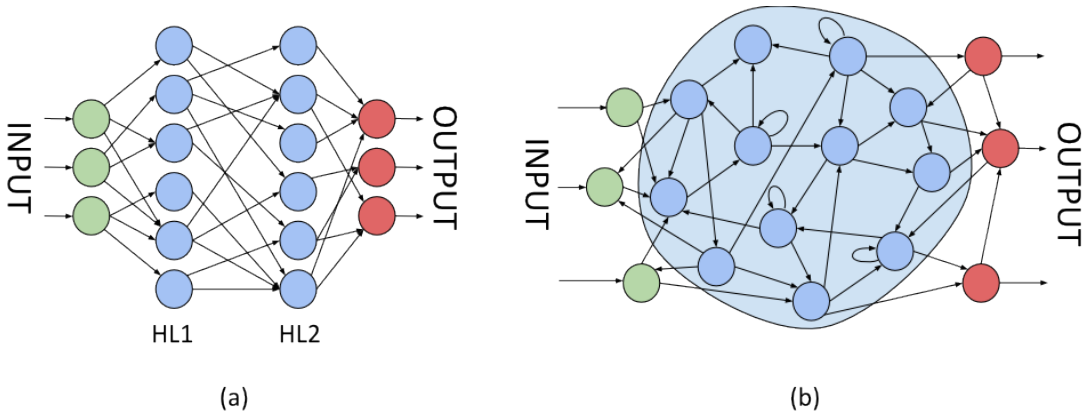


Figure 3.4: An example of a feed-forward neural network (a) and a recurrent neural network (b). Both networks have 18 nodes, 3 of which are used for external input and 3 for output. The recurrent neural network includes possibly circular connections or neuron self-connections.

3.3 Nervous system

As we saw in section 2.1.1.3, even very simple connections between sensory and motor areas can be described by an observer as "behaviour". Behaviour, then, is not generated in and does not even *need* a nervous system; however, a nervous system increases the number of ways the sensory and motor areas can be connected, so it greatly enhances an agent's behavioural space. A plastic nervous system that can be changed in a system's lifetime can also enable *learning*, which is a necessary element of transmission chains. In this section, we detail the choice of nervous system for our e-puck agents.

3.3.1 Recurrent neural networks

Our agents' nervous system will be modelled using *continuous time recurrent neural networks* (Beer, 1995b). In contrast to neural networks that only have feed-forward connectivity (Figure 3.4a: the input layer sends its output to a number of hidden layers, the last of which in turn sends its output to the output layer), the hidden layer of a *recurrent* neural network ("RNN") can also have circular connections to itself, including neuron self-connections (Figure 3.4b). Because of this connectivity pattern, the behaviour of RNNs is a lot more complex than feed-forward networks.

If the activation of each node of an RNN is not computed in discrete time but continuously, the network is described as a Continuous Time RNN ("CTRNN"). CTRNNs are dynamical systems that model (heavily abstracted) biological neural networks. Each node's activation is given by the solution of the following differential equations:

$$\dot{y_i} = \frac{dy_i}{dt} = \frac{1}{\tau_i}\left(-y_i + \sum_{j=1}^{N} w_{ji}\sigma(y_j + \theta_j) + I_i\right), i = 1,2,...,N \tag{3.1}$$

In Equation 3.1, y_i is the activation of node i, w_{ji} is the strength of the synaptic connection from node j to node i and I_i is an external input to the node. τ_i and θ_i are the node's time constant and bias accordingly, while σ is the logistic function:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3.2}$$

CTRNNs are simple, but powerful: Funahashi and Nakamura (1993) show that they can be universal approximators of any dynamical function. Their rich dynamics and biological basis make them strong candidates in dynamical and non-representational approaches to cognitive phenomena, so CTRNNs have been often used in simulations of locomotion and minimally cognitive phenomena (Beer and Gallagher, 1992; Beer, 1996, 2003), the evolution of coordinative behaviour and communication (Di Paolo, 1997, 2000) and for the control of physical robots (Di Paolo, 2004). They are also extensively used in the field of *evolutionary robotics*, where controllers for physical robots are evolved using genetic algorithms (Cliff et al., 1993; Harvey et al., 1997).

3.3.2 Spiking neural networks

An alternative to CTRNNs would be a spiking implementation of biological neural networks. These implementations range from very detailed, computationally heavy models that most closely capture the dynamics of real neurons to more abstract but "quicker" models (see Izhikevich, 2004). Even the simpler models are more biologically plausible as models of real brains than CTRNNs, as in CTRNNs all spiking has been abstracted away.[1]

[1] In a CTRNN, the activation y of a node represents the mean membrane potential and $\sigma(y)$ is indicative of the "short-term average firing frequency" (Beer, 1995b, p. 3). The actual spiking times are lost in the averaging process.

Spiking models are of high interest for dynamical explorations of cognitive phenomena (Di Paolo, 2003) and especially so for a biologically grounded framework like autopoiesis. Despite that fact, we decided not to use a spiking network implementation for our experiments in order to benefit from the (relative) simplicity of CTRNNs. We will discuss possible future work using a spiking implementation in section 7.7.3.

3.4 Learning

The Iterated Learning models are based on successive episodes of transmission from agent to agent, and in essence a transmission episode is an episode of learning. It follows, then, that the choice of implementation of learning is an important design decision. Autopoietic theory is not of much help here; as we saw in section 2.1.1.4, its definition of learning is very wide ("A phenomenon of transformation of the nervous system associated to a behavioural change that takes place under maintained autopoiesis"; Maturana and Varela 1980, pp. 35-38). This tells us that any plastic change of the nervous system that leads to a change in behaviour is learning, as long as it does not lead to a cease of the organism's process of autopoiesis (in other words, to its death).

3.4.1 Learning & CTRNNs

A more substantial constraint comes from our choice of nervous system: recurrent neural networks are powerful, but also very challenging to train (Jaeger, 2002). We will go over a short list of potential approaches.

"Engineering" approaches: Most training approaches for recurrent neural networks target engineering and signal processing applications; they also often use discrete time networks or supervised learning, both of which are incompatible with our scenario of two agents interacting and learning in continuous time. An extensive tutorial on some of those approaches (back-propagation through time, extended Kalman filtering, real-time recurrent learning, echo-state networks) is given by Jaeger (2002).

Genetic algorithms: Almost all research that employs CTRNNs uses genetic algorithms to train them (Beer and Gallagher, 1992; Di Paolo, 1997; Harvey et al., 1997; Quinn, 2001). Evolutionary computation methods have proven very useful

in finding sets of weights that enable CTRNNs to exhibit interesting behaviour. This is not an option in our case; as we repeated in section 3.1.2, transmission chains happen in ontogenetic and glossogenetic rather than phylogenetic (evolutionary) timescales.

Fixed weight learning: Yamauchi and Beer (1994) show that it is possible to evolve CTRNNs that learn based on reinforcement from the environment. Each evolved agent can exhibit a number of behaviours; in its ontogenesis, environmental reinforcement "selects" one of these possible behaviours. Furthermore, this happens purely based on network inner dynamics, without any weight changes. Blynel and Floreano (2003) expand this to learning how to solve a "T"-maze. This approach is potentially suitable for our experiment and intriguing from an autopoietic background, although it is hard to ground it in a biological explanation. We decided against adopting it, however, as the addition of an evolutionary stage would increase the scope of the project too far.

3.4.2 Hebbian learning & STDP

In addition to the approaches discussed above, there is a vast literature that studies learning from a biological perspective, as synaptic plasticity in the brain. The specific theories and mechanisms involved are numerous, and depend on the type of learning that is being explained. The type of learning most applicable to the task we described in 3.1.2 is *instrumental* or *operant conditioning*, in which reward (or punishment) is used to strengthen the association of a stimulus and a response. The most common theory of how instrumental conditioning works is *Hebb's rule*:

> Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. (Hebb, 1949)

In other words, when a node A fires and repeatedly causes node B to fire as well, the synaptic connection between them is strengthened. The biological process behind this is *spike-timing dependent plasticity* ("STDP"). According to STDP, if there is a spike in node A's activity *shortly before* a spike in node B's activity, the A-B synapse is strengthened (long-term potentiation, "LTP"); if the spike in node A's activity comes *shortly after*, the A-B synapse is weakened instead (long-term depression, "LTD").

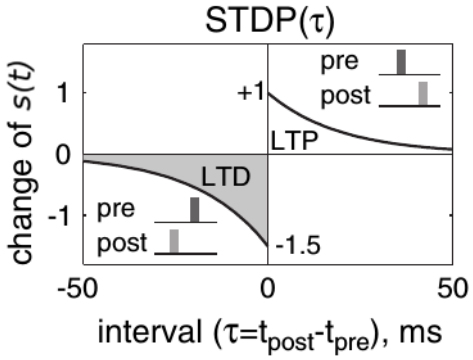


Figure 3.5: According to the STDP process, the synaptic strength $s(t)$ of the connection between two nodes is modulated by the interval between the pre-synaptic node firing (t_{pre}) and the post-synaptic node firing (t_{post}). When the pre-synaptic node fires shortly before the post-synaptic, $\tau = t_{post} - t_{pre} > 0$ and $s(t)$ increases; if the post-synaptic node fires first, $\tau < 0$ and $s(t)$ decreases (adapted from Izhikevich, 2007, p. 2444).

A graph of how STDP interacts with the time interval between node A firing (t_{pre}) and node B firing (t_{post}) is shown in Fig. 3.5. This temporal precedence $\tau = t_{post} - t_{pre}$ is a measure of causality in the activations of the nodes, also implicitly present in Hebb's definition of learning.

One issue that is problematic in most STDP accounts is known as the "distal reward" problem (Izhikevich, 2007). In most learning scenarios, any reward or punishment is not administered instantly but some seconds after the neural firing patterns that connect stimulus to behavioural output. Also, both before and after those firing patterns that lead to the behaviour that is to be reinforced, all nodes tend to fire randomly. How does the learning mechanism then know which connections to "reward"?

Izhikevich (2007) solves this problem by using a learning mechanism that combines STDP as a "tagging" mechanism with reward-based reinforcement. The way this works is detailed in Fig. 3.6; instead of STDP directly adjusting the synaptic strength s of the connection between two neurons, it only adjusts a "tagging" variable (or *eligibility trace*) c, increasing it if the pre-synaptic node fires shortly before the post-synaptic one; it then decays back to resting levels over some seconds.

When the system produces the wanted behaviour, it is rewarded by an increase in the concentration of dopamine, d, in the system. The synaptic strength is controlled both by the eligibility trace and by this concentration of dopamine:

$$\dot{s}(t) = c(t) \cdot d(t) \tag{3.3}$$

When the co-activation of the two nodes leads to a reward, the combination of $c > 0$ and $d > 0$ leads to the "tagged" connections being strengthened; since c decays slowly, this happens even if the reward comes a few seconds later than the initial co-activation.

A simplified implementation of this learning method, adapted for CTRNN instead of spiking networks, is what we will use for our experiments. We will detail and discuss our own implementation extensively in Chapter 4.

3.5 Experiment design

So far we have outlined what our agents will look like (simulated e-puck robots with continuous-time recurrent neural networks as nervous systems), what the task these agents will be asked to accomplish will be (a joint action left/right "dancing" task) and how the agents will learn how to perform that task (tag-based reinforcement learning). Here we will examine how these components tie together in a cohesive experiment.

3.5.1 Dancing task, input and output

The L/R dancing task that we described in section 3.1.4 can essentially be summarised by the following rule: "*If* there is another agent in proximity, *then* turn to the left / turn to the right". Which direction the robot needs to turn towards for successful dancing depends of course on the partner as well; regardless, this association between *proximity* and *turning left* or *turning right* is what the agent has to learn.

The input to the agent's nervous system comes directly from the e-puck infrared sensors and indicates proximity: another agent nearby stimulates the network's input node. (This stimulation happens regardless of which of the eight infrared sensors detected proximity.) The output is trickier, because while the response behaviour seems simple (turning around an object) it is actually quite complex, as it involves the use of sensors as well as motors (in order to keep the distance from the target object consistent). To simplify the task, we decided to "offload" the turning behaviour to a separate *action module* that takes a simple output from the agent's nervous system ("R" or "L") and translates it to the respective low-level behaviour (Fig. 3.7).

While not ideal, this design decision is justifiable, since a model of the acquisition of sensorimotor associations is beyond the scope of our project. Through either a developmental or evolutionary process, our agents have a pre-existing (and admittedly, quite poor) behavioural repertoire. Their task is then to learn to associate one of these

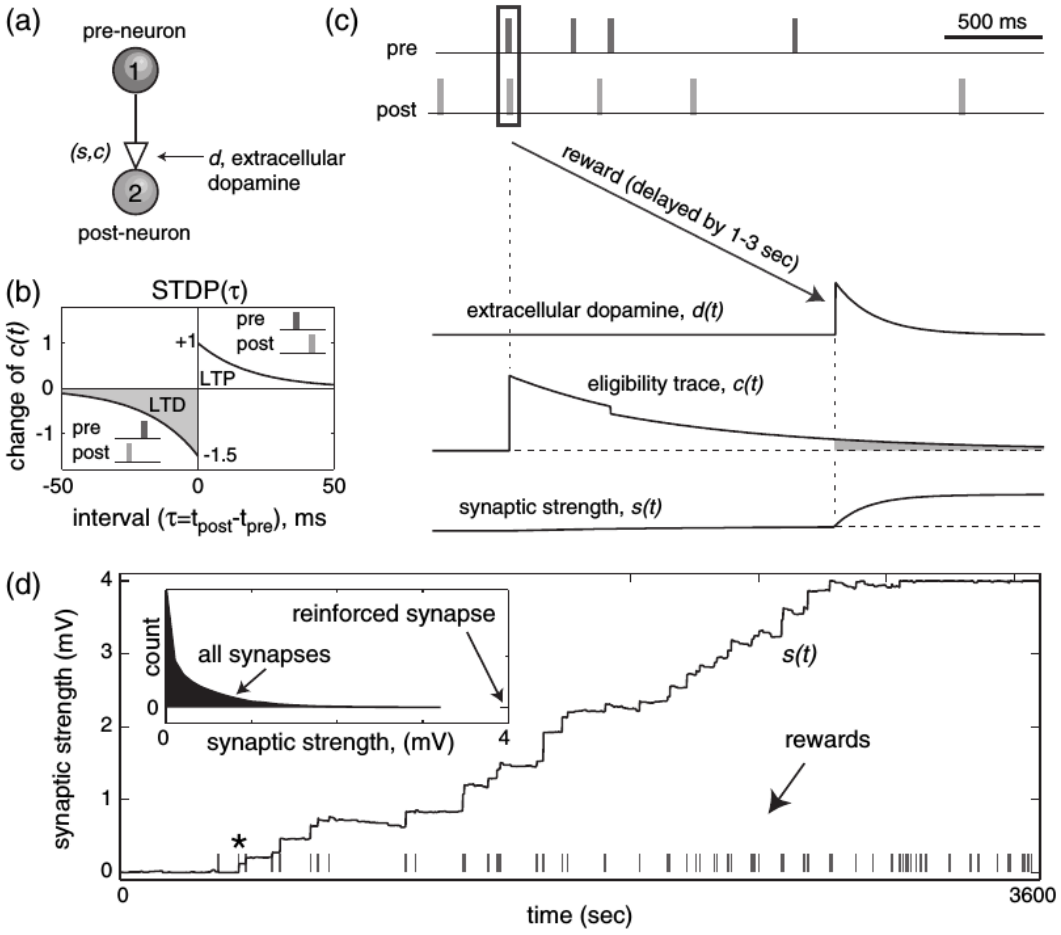


Figure 3.6: A method combining STDP and dopamine rewards to solve the "distal reward" problem (Izhikevich, 2007). (a): Given a neural connection between neuron 1 (pre-neuron) and neuron 2 (post-neuron), s is the synaptic strength of the connection and c is an "eligibility trace" that acts as a tagging variable. (b): The value of $c(t)$ is given by the STDP function; if neuron 1 fires shortly before neuron 2, $c(t) > 0$. (c): If a reward $d(t)$ is delivered, even if it is a few seconds after the co-incidental firing, the synaptic strength $s(t)$ of the connection is still increased, as the decay rate of $c(t)$ is longer than a few seconds. (d): By repeatedly rewarding co-incidental firings of neurons 1 and 2, the connection between them is gradually changed to be significantly stronger than other synapses in the system.

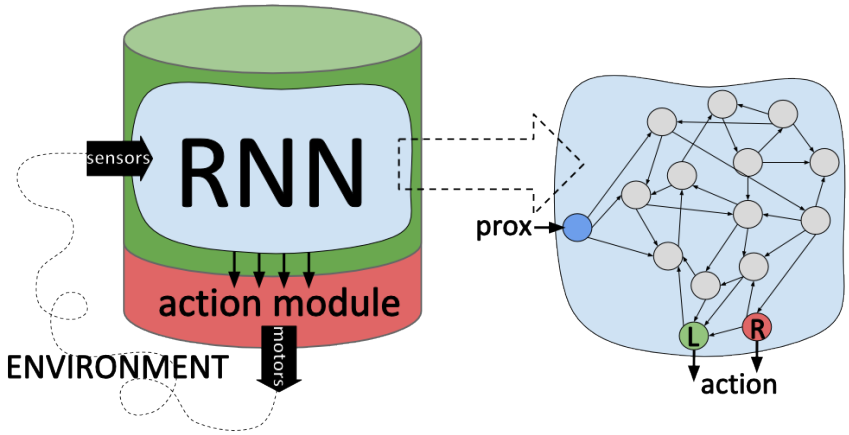


Figure 3.7: Schematic of e-puck robot, action module and RNN. Input from the robot's proximity sensors directly stimulates the neural network's input node; the values of the two "R" and "L" output nodes are redirected to the action module, which in turn translates them to low-level motor commands making the robot turn around in the respective direction.

behaviours with the appropriate stimulus (in this case, another agent's proximity).

3.5.2 Teaching and learning

We established what agents have to learn; the question that remains now is, how will they learn it? Since L/R dancing is a joint action task, learning must happen through reinforcement during an interaction of the learning agent (the "learner") with an expert agent already proficient at the task (the "teacher"). Turning in the same direction as the teacher leads to successful dancing; if this is rewarded enough times, the connection between stimulus and the correct response will be strengthened, leading to learning. On the other hand, turning in the wrong direction will lead to crashing; punishing any crashes will also help learning.

At this point we must note that while from now on we will often refer to the expert agent as the "teacher" for convenience, we are definitely not making the claim that this is *actual* teaching behaviour. While teaching is believed to occur in some animals such as bees and whales in addition to humans (Hoppitt et al., 2008), it is defined as "actively facilitating learning in others" (Hoppitt et al., 2008, p. 486) — something that is completely missing from our system, where the expert is "teaching" only by giving consistent left or right responses.

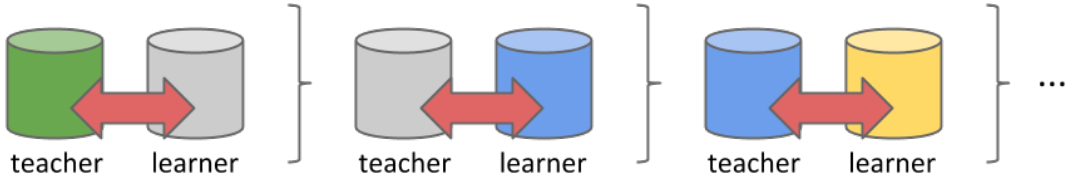


Figure 3.8: A L/R dancing transmission chain. A pair of teacher-learner agents interact, with the teacher removed in each successive generation and the previous learner interacting as a teacher with a new learner.

3.5.3 Transmission chains

Our goal is to link individual learning events into cultural transmission chains, maintaining a certain behaviour through successive generations of learning and teaching. In order to do this, we will start with two agents (an expert "teacher" and a "learner") interacting; for each generation, we will remove the teacher and add a new learner; the former learning agent will now play the teacher's role and interact with the new agent, hopefully transmitting whatever behaviour they learned from their own teacher (Fig. 3.8). We will get into more detail on transmission chains in Chapter 5.

3.6 System design

We have described a number of components: a *simulated robot* in an (also simulated) *environment*, a *neural network*, a joint action *task* that rewards or punishes the agents taking part in it, and an *experiment* that makes us of all other components. We will close this chapter by describing the design of the system that connects everything together.

The system was designed to be highly modular; in this way, we can easily test new modules and replace any of the existing ones without having to make changes to the rest of the system. A schematic of the whole system, including all modules and their connections, is sketched in Fig. 3.9. The individual modules are the following:

Robot & environment simulation: We implemented the e-puck robots (Fig. 3.9a) and their environment (Fig. 3.9b) in the *Stage* simulator (Vaughan, 2008). *Stage* provides a ROS (Quigley et al., 2009) interface through which each robot can be controlled and its sensory information accessed; it also provides information about the state of the environment, such as robot positions or simulation time.

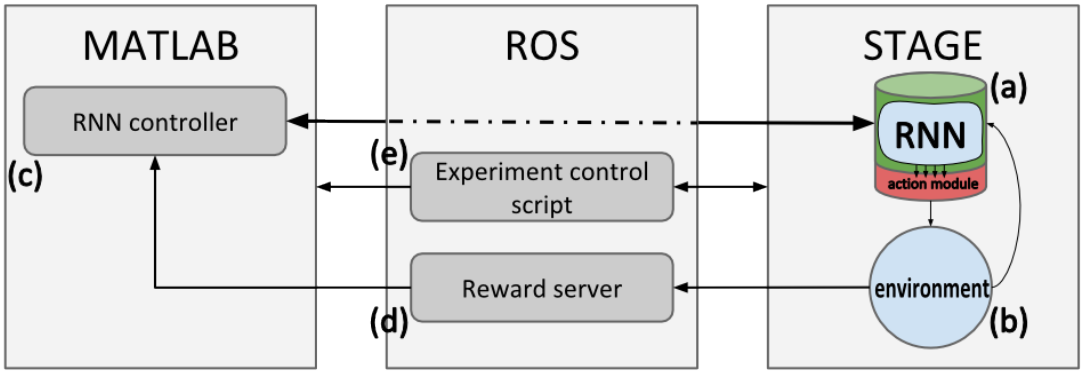


Figure 3.9: Schematic of all the system components and connections between them. The e-puck robots (a) and their environment (b) are implemented in a *Stage* simulator. Their neural network controllers (c) are implemented in MATLAB and interface with *Stage* through ROS. A reward server (d) collects information from the environment and, depending on the task, issues rewards or punishments to the neural network controllers for use in the learning algorithm. The experiment control scripts (e) coordinate all other components into coherent experiments, using simulation timing information from *Stage*.

Neural network: We implemented the RNN controllers (Fig. 3.9c), including the learning algorithm, in MATLAB (MATLAB, 2014). The controllers interface with the simulated robots through ROS, retrieving sensory input and issuing motor commands. (The robots' action module is also implemented in MATLAB.)

Task: The task is represented in a "reward server" (Fig. 3.9d) that we implemented in ROS; the server takes world information (agent positions, velocities and timing) from *Stage* and, depending on the task, decides whether to reward or punish each agent in the form of positive or negative dopamine administration. (Reward and punishment information is transmitted to the neural network controllers and used in the learning algorithm; on successful dancing or crashing, both agents are rewarded or punished accordingly.)

Experiment control: Finally, the specific experiments are coordinated using "experiment control scripts" (Fig. 3.9e) implemented in ROS; the scripts create, modify and replace agents as needed, using simulation timing information from *Stage* to make sure that all experimental steps are synchronised. The configuration of each experiment will be discussed in detail in the next chapters.

Chapter 4
Learning in isolated pairs

As mentioned in the previous chapter, at the base of cultural transmission chains are single episodes of learning. In this chapter we will go into detail on the implementation of a learning algorithm, based on the STDP "distal reward" learning proposed by Izhikevich (2007). We will first establish that the learning implementation works in both a disembodied "brain in a vat" test and in isolated pairs of learners and experts performing the joint action L/R dancing task. Having done that, we will examine how the length of interaction time in the learning pairs influences task success rates.

4.1 Learning implementation

In section 3.4.2 we mentioned a biologically plausible reinforcement learning model (Izhikevich, 2007) based on spike-timing dependent plasticity: when two neurons fire within a short time interval, indicating causality in their firing behaviour, their connection gets tagged. When reward comes in the form of dopamine (DA), only the connections that are tagged are strengthened. This synaptic tag takes a few seconds to decay; this means that even if the reward comes later than the neural activity which caused the rewarded behaviour, the right neural pathway can be strengthened.

In our case, there is a significant difference: we are using continuous time neural networks instead of spiking networks. CTRNNs don't simulate spiking incidents; instead, the average firing frequency of node i is represented by the node's activation value (y_i). Our learning algorithm is based on the Izhikevich model, but since we have no access to spiking times, we use a *co-activation detection* metric instead of STDP to detect causal activations. The basic idea, as shown in Fig. 4.1, is still the same: connections that activate together get tagged, and tagged connections get strengthened when a behaviour is rewarded (in the form of a supply of dopamine to the whole network).

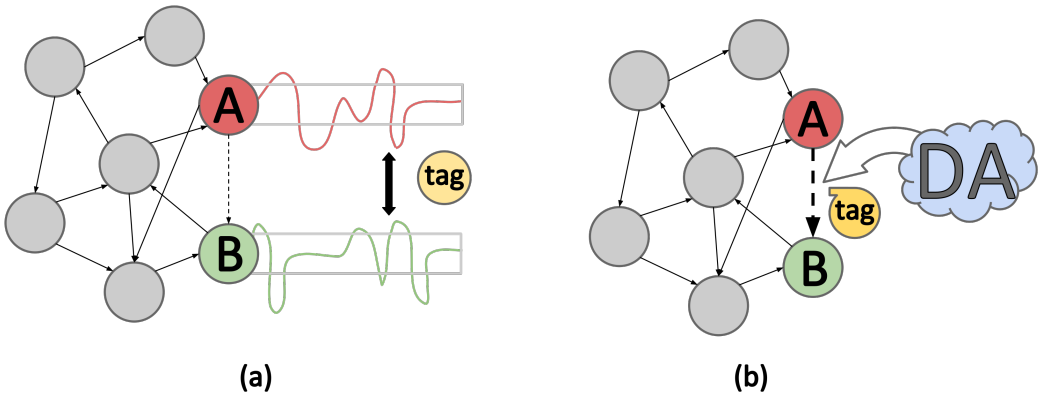


Figure 4.1: Basic idea of the "tag & reward" learning algorithm. An increase in node A's synaptic activity leads to a change in node B's activity; the correlation between the activation values y_A and y_B leads to the $A \rightarrow B$ connection being tagged. This persists for a few seconds (a). If the $A \rightarrow$ B activation pattern led to a rewarded behaviour, the dopamine released in the network strengthens the value of the tagged $A \rightarrow B$ connection. As the connection's tagged status persists, even late rewards can strengthen the right connections (b).

We will discuss some of the implementation details below and in the next section we will present a basic overview of the algorithm's flow.

4.1.1 Learning algorithm: main points

Neural network simulation: As we saw in section 3.4, CTRNNs are dynamical systems described by Equation 3.1. We approximate the solution of the dynamical system by using the Euler method; given a $\dot{y} = f(y)$ system and a time step size of h, the Euler method approximates the value of y as $y_{t+1} = y_t + h \cdot f(y_t)$. In our case, for a CTRNN, the Euler approximation is:

$$y_i(t+1) = y_i(t) + h \cdot \frac{1}{\tau_i}\left(-y_i(t) + \sum w_{ji} \cdot \sigma\left(y_j(t) + \theta_j\right) + I_i\right) \tag{4.1}$$

Once more, σ is the sigmoid function: $\sigma(x) = \frac{1}{1+e^{-x}}$. For a CTRNN simulation a value of $h = 0.01$, equivalent to 100 time steps for every simulated second, gives a sufficiently good approximation. This is the value we will be using for all experiments described from now on.

RNN parameters: There are two parameters in Equation 4.1 that control the behaviour of network nodes: τ (a node's time constant) and θ (a node's bias term). Lower τ leads to nodes that react more quickly to input, while θ controls the relative bias of each connection; positive θ_j values lead to amplified responses to stimulation from a specific connection between neurons i and j, while negative values lead to muted responses. Changing these parameters can lead to drastic changes in the behaviour of the network; by a process of trial and error, we settled on values of $\tau = 0.1$ and $\theta = -4$ for all nodes, which led to networks that respond quickly (but not instantaneously) to perturbation.[1] Further work could be done towards a more systematic approach to selecting these variables; they could also be potentially incorporated into the learning mechanism, allowing for different values of τ and θ for different nodes.

External stimulation: One or more nodes in the network (the "input" nodes) get stimulated by external perturbation; this is accomplished by manipulating the I_i parameter of input nodes. In our system, the external perturbation corresponds to the detection of a nearby object by the proximity sensors. For simplicity, we only use one input node that gets stimulated with a "current" of $I = 10$ if any of the eight proximity sensors detect a value lower than a threshold $prox_{\text{thr}} = 1.5$. In both the physical and simulated e-pucks, there can be momentary failures in sensory reporting that lead to erroneous behaviour in the system. This can be prevented by choosing higher values of τ for input nodes, providing a "buffer" against instantaneous changes; in our system, we instead kept a rolling window of the 10 latest values obtained from each sensor and used the average of those values to determine proximity, smoothing out any momentary sensor failure.

Noise: In order to be able to learn, our network depends on random co-activations of network pathways that happen to produce the correct response to a certain stimulus. These random co-activations would not be possible without some sort of noise. In actual nervous systems, this noise is electrical noise from surrounding neurons; in our neural network, we model this by additive white Gaussian noise with a signal to noise ratio of $SNR = 10dB$ via the external node input I_i. We made the choice of excluding the input and output nodes from this noise addition; sensor noise is already present in the input node as part of the e-Puck *Stage*

[1] Note that the effect of θ on a system must be balanced relatively to the intensity of any external stimulation perturbing it.

simulator, and we wanted the output nodes to only be driven by activations of pathways of the neural network.

Output control: As we mentioned in section 3.6, the output nodes of the neural networks controlling the agents are not directly connected to the agents' motors. Instead, two output nodes are each associated with a specific behaviour ("dance left" or "dance right") that a lower level action module then translates into motor commands. Along with these two action modules, another low-level component is a crash recovery module that takes over and forces the e-puck to back off if a crash is detected.

Barring that, behaviour selection depends on the relative activation levels of the output nodes. As an example, let us suppose a system with two behaviours (A and B) and two corresponding output nodes, with activation values of y_A and y_B. In order for behaviour A to be selected, two conditions need to be present:

$$y_A - ||y_A|| > \theta_{\text{out}}, \text{ and} \tag{4.2a}$$

$$\frac{y_A}{y_B} > \theta_{\text{diff}} \tag{4.2b}$$

Behaviour A can only be selected if the value of the corresponding node's activation y_A (after subtracting a measure of normalised activation $||y_A||$) is higher than a threshold θ_{out} (4.2a). $||y_i||$ refers to the *resting* activation of node i when no stimulus and no noise are present in the system. This ensures that without stimulus, the system's resting response is neither behaviour A nor behaviour B, which can only be produced in response to a stimulus. For the experiments we detail in this thesis the resting response is a *random walk*; without any object in proximity, agents wander around randomly. This facilitates encountering other agents to try and complete the L/R dancing task with.

In addition to the above constraint, in order for the system to respond with behaviour A, the relative activation of y_A compared to y_B must be higher than a certain threshold θ_{diff} (4.2b). In theory, lower values of θ_{diff} can lead to shorter learning times but less robust behaviour. In our experiments, we used threshold values of $\theta_{\text{out}} = 0.3$ and $\theta_{\text{diff}} = 1.1$.

One extra constraint added to make the agents' behaviour more stable was to lock the selected behaviour while the stimulus that caused it is present. This makes dancing easier to maintain; however, it also blocks agents from learning more complex combinations of behaviours and its biological plausibility is suspect.

Causal activation detection: Since, as we mentioned, spike timing information is abstracted away in CTRNNs, we need a way of detecting when two nodes y_i and y_j interact in a causal manner. Instead of STDP, we used the following "CA" co-activation measure, based on the multiplication of the rates of change of the activations of nodes i and j, passed through an arctan high-pass filter (Fig. 4.2):[2]

$$CA_{ij} = f(dy_i \cdot |dy_j| \cdot y_j), \text{ where} \tag{4.3a}$$

$$f(x) = \begin{cases} 0, & x < 0 \\ arctan(x), & x \geq 0 \end{cases} \tag{4.3b}$$

The value of $dy_i \cdot |dy_j|$ increases only when both nodes fire together (so that $dy_i > 0$ and $dy_j > 0$ at the same time). $|dy_j|$ is normalised to allow for the reinforcement of inhibitory connections; the multiplication by y_j and the *arctan* high-pass filter (4.3b) help with exaggerating the difference between two post-synaptic nodes that both get stimulated by the same pre-synaptic node. Note that this means that in our learning implementation, the co-activation variable *CA* is not binary: random input noise can mean that a single pre-synaptic node activation can lead to two post-synaptic nodes being marked as co-active in parallel with different intensity. This allows for the network pathways that lead to rewarded behaviour to be strengthened comparatively to all other stimulated pathways, as the *CA* measure is directly connected to the tagging variable c:

$$\dot{c} = -c/\tau_c + CA \tag{4.4}$$

This differential equation (4.4) means that when two nodes are detected as co-activating, their connection gets tagged; the tagging variable then decays at a rate of τ_c. In all the experiments described from here on, we used a value of $\tau_c = 2$ which leads to a decay time of around 5 seconds, consistent with biological results as reported by Izhikevich (2007, p. 2445).

As an example of the effect of noise in the tagging of specific connections, let us consider a basic network of three nodes, where node 1 is connected to nodes

[2] A better measure that we tested was based on the cross-correlation of the two signals; if the correlation between y_i and y_j is maximised when applying a negative time shift on y_i, we have an indication that a change in y_i led to a corresponding change in y_j. The measure we ended up using is much less computationally expensive, however, and it works well for small networks as long as there are no direct closed loops between two nodes.

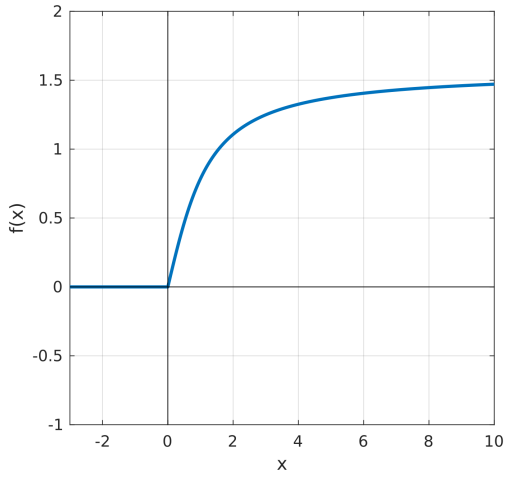


Figure 4.2: High-pass filter using $arctan$.

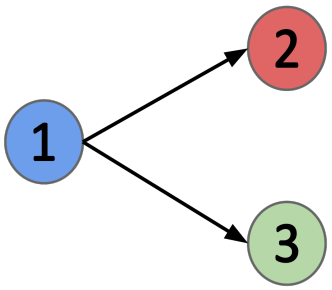


Figure 4.3: A simple network of 3 nodes; node 1 is connected to nodes 2 & 3 with equal weights $w = 2$.

2 & 3 and both connections have a weight of $w = 2$ (Fig. 4.3). Node 1 is externally stimulated, causing the value of its activation y_1 to change from 0 to 10 (Fig. 4.4a). This stimulation propagates to nodes 2 & 3 as well, changing the values of their activations y_2 and y_3. In the absence of noise, the values of y_2 and y_3 would be the same; however, higher noise in the input I_2 of node 2 makes it so that $y_2 > y_3$ (Fig. 4.4b), possibly leading to a different response for the rest of the network. Fig. 4.4c and 4.4d show the values of the co-activation detection variables (CA) and tagging variables (c) respectively, for each of the connections. While both connections are tagged, connection $1 \rightarrow 2$ has a higher tagging variable c compared to connection $1 \rightarrow 3$. In the event of a reward, w_{12} will be increased more than w_{13}. In Fig. 4.4d we can also see the gradual decay of the tagging variables c_{12}, c_{13} over approximately 5 seconds.

Weight changes: The final step left for our learning algorithm is the combination of the tagging variable c with an external influx of dopamine (DA) in order to strengthen the synaptic weights w of tagged connections. The externally administered DA sets the system's levels of dopamine, d:

$$\dot{d} = -d/\tau_d + DA \tag{4.5}$$

The time constant τ_d affects the rate of decay of the dopamine levels; we used a value of $\tau_d = 0.2$. Finally, high levels of dopamine in the system increase the

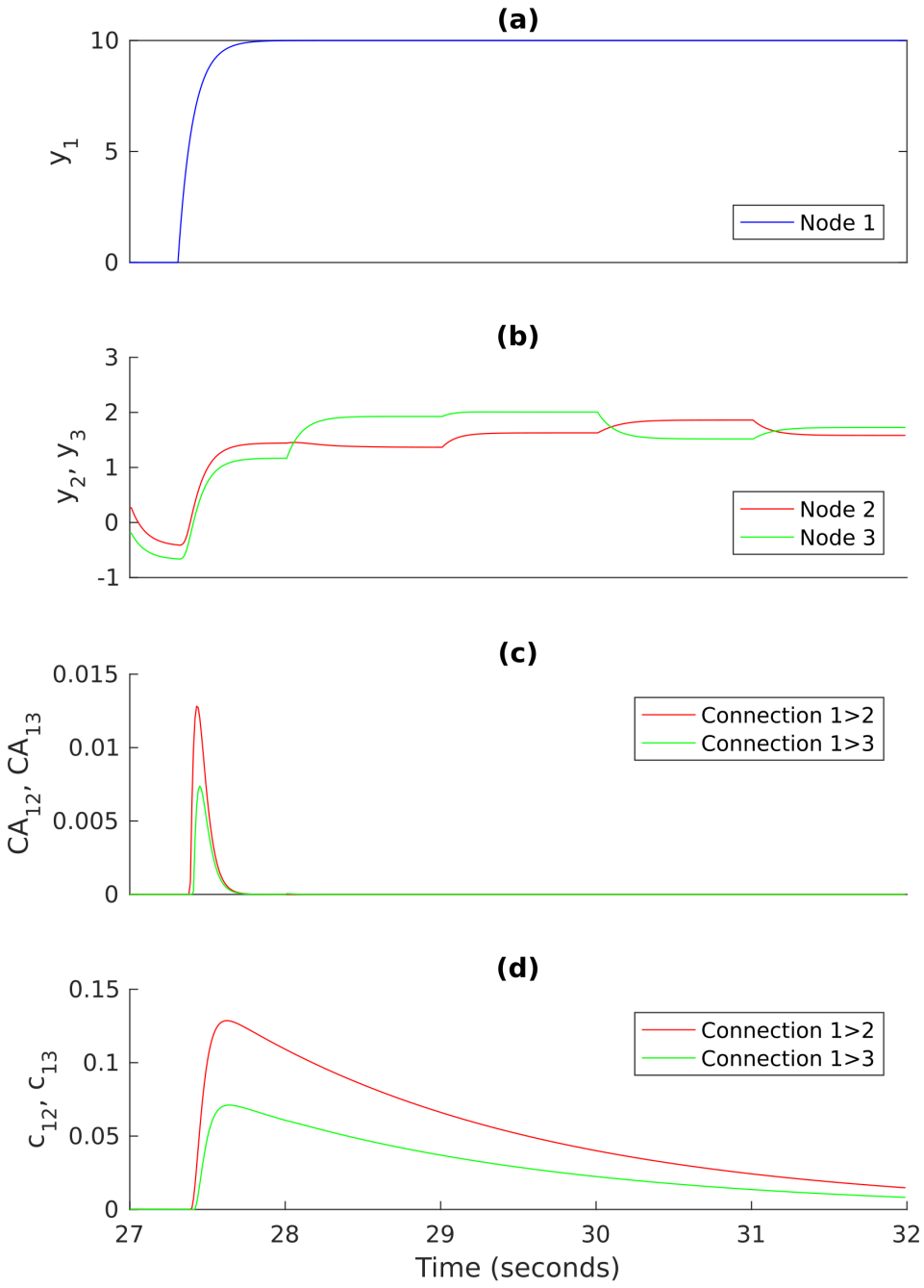


Figure 4.4: An example of the effect of noise in the tagging of the connections of a 3-node network. Stimulation of node 1 (a) leads to changes in the activation y of connected nodes 2 & 3 (b). White Gaussian noise leads to stronger initial activation in node 2; which in turn leads to increased co-activation detection (c) and, finally, tagging variable c_2 compared to c_3 (d). The tagging variable decays after approx. 5 seconds.

synaptic weights w of tagged connections, as long as they are within the bounds of the minimum and maximum values allowed, w_{min} and w_{max}:

$$\dot{w} = \begin{cases} c \cdot d, & w_{min} < w < w_{max} \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

The minimum value w_{min} is set to the initial weight values of each network, w_{init}. We need a minimum weight value in order to stop the networks from being able to degenerate into non-responsiveness.

Equivalently, without a maximum weight value constant positive reinforcement after the system reaches a state where it always outputs the reinforced behaviour leads to the weight values spiralling out of control. This is partly because of how we calculate the tagging variable for each synaptic connection. Higher weight values w_{ij} lead to higher activation y_j in the respective nodes, higher tagging variable c_{ij} (Eq. 4.3a, 4.4) and, finally, even higher weight values (Eq. 4.6).

The maximum weight limit could potentially be replaced (or reinforced) with the addition of a maximum value for the synaptic tag c; the spiking neuron implementation, however, also limits the weights of each connection to a specific range (Izhikevich, 2007, p. 2445). The maximum weight value we used in the experiments described in this thesis was $w_{max} = 10$.

While any of the weights hold this maximum value, all weights in the network stop increasing. This is a measure that is biologically suspect, but is needed in order to stop all weights in the network from eventually reaching w_{max}, undoing the comparative strengthening of the learning process; this issue is also recognised and discussed by Izhikevich (2007, p. 2447). It is important to note, however, that while the network weights are stopped from *increasing*, the learning process is not frozen at this point, as weights can still decrease as a result of punishment. Punishment is implemented through negative *DA* values; this, once more, is a biologically impossible "cheat" that is only used as a shortcut in place of a proper punishment mechanism (for a "proper" mechanism, see for example Cohen et al., 2012). We will return to the significance of the maximum weight values and the weight strengthening "freeze" for some of our model's interesting behaviour in Chapter 5.

4.1.2 Learning algorithm: flow overview

We detailed all the main points of our "tag & reward" learning algorithm in the previous section; here, we will give an overview of the algorithm's flow. Each time step corresponds to $h = 0.01 = 1/100$ of a simulated second. The correspondence of simulated seconds to real-time seconds depends on the simulation speed of our simulator, *Stage* (see section 3.6). The non-essential parts of the algorithm, such an initialisation, output, visualisations and graphing are not included below.

A: Pre-Euler actions (every 10 time steps)

1. **Update sensors:** Get the proximity sensor data from the simulator and the *DA* reward data from the reward server.
2. **Add noise:** Add Gaussian white noise to all nodes except for the input and output nodes.
3. **Synchronise timers:** Wait for the *Stage* simulator if the internal timer is ahead, notify if the internal timer is behind.

B: Euler simulation (every 1 time steps)

1. **CTRNN update:** Calculate the new values of y for the network (Eq. 4.1).
2. **Dopamine level update:** Update the level of dopamine d in the system based on the current value of *DA* (Eq. 4.5).
3. **Tagging update:** Tag co-activating connections (Eq. 4.3a, 4.3b).
4. **Weights update:** Update weights (Eq. 4.6).

C: Post-Euler actions (every 10 time steps)

1. **Crash avoidance:** If a crash is imminent, let the low-level crashing avoidance system take over.
2. **Output control:** Determine whether to produce any behaviour; if yes, call the respective low-level function (Eq. 4.2a, 4.2b).

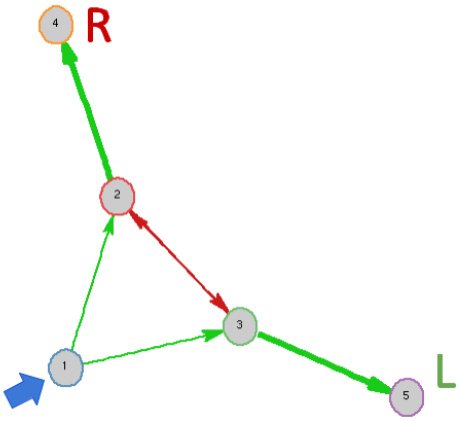


Figure 4.5: RNN used in the "brain in a vat" test. Connections shown in red are inhibitory, initialised with $w_{23} = w_{32} = -2$. All other connections are excitatory, initialised with $w_{12} = w_{13} = 2$ and $w_{24} = w_{35} = 5$. Node 4 is the "Right" behaviour output node, while node 5 is "Left". Node 1 is the input node.

4.1.3 Learning algorithm: "brain in a vat" test

In order to show that the "tag & reward" learning algorithm we detailed works as intended, in this section we describe a simple "brain in a vat" test case. For the test, we will use a 5-node RNN with 1 input and 2 output nodes (Fig. 4.5: connections shown in red are inhibitory, initialised with $w_{23} = w_{32} = -2$. All other connections are excitatory, initialised with $w_{12} = w_{13} = 2$ and $w_{24} = w_{35} = 5$.). Node 1 is the input node; node 4 is the "Right" behaviour output node, while node 5 is "Left".

Instead of connecting this network to the body of an agent, the input node is directly stimulated with $I_1 = 10$ for $t = 0.5s$, alternating with $I_1 = 0$ for $t = 0.5s$. *DA* diffusion is also directly controlled: a response of "Right" behaviour under stimulus is rewarded with $DA_r = 1$, while a response of "Left" behaviour under stimulus is punished with $DA_p = -0.5$. In both cases, the reward or punishment is artificially delayed by $t = 3s$ to represent the delayed reward of an actual task.

Fig. 4.6 shows an instance of delayed reward. Stimulation of the input node 1 leads to increased activation in both of its connected nodes, 2 & 3. However, input noise randomly causes node 2 to activate more (4.6a), leading to a higher tagging variable (4.6b) and, through the neural pathway $N1 \rightarrow N2 \rightarrow N4$, to the system responding with "Right" behaviour. Since this is the rewarded behaviour, 3 seconds later a *DA* influx increases the dopamine concentration in the system (4.6c). The combination of tagging ($c_{12} > 0$) and reward ($d > 0$) will lead to an increase in the connection's strength (w_{12}).

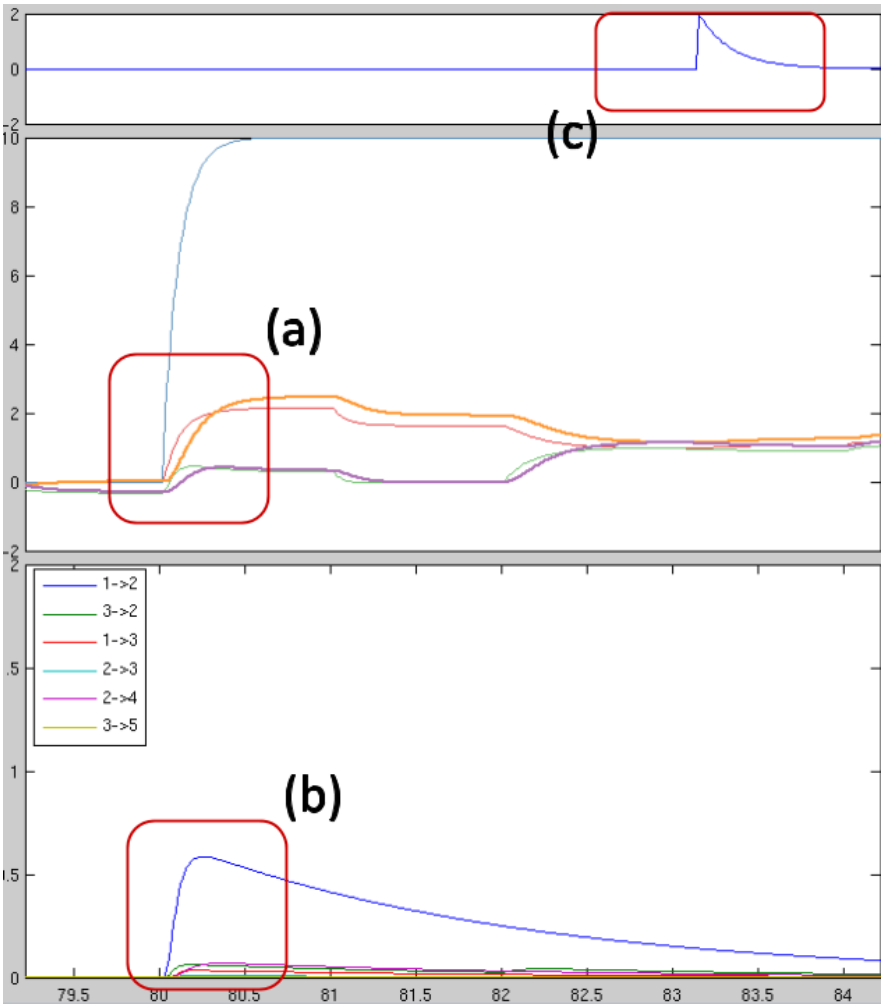


Figure 4.6: An instance of delayed reward. Node 1 leads to increased activation y in both of its connected nodes, 2 & 3. However, input noise leads to y_2, shown in red, being higher than y_3, shown in green (a). This leads to a higher tagging variable y_{12} (b) and eventually to delayed reward (c). The combination of tagging and reward will lead to an increase in the tagged connection's synaptic strength w_{12}.

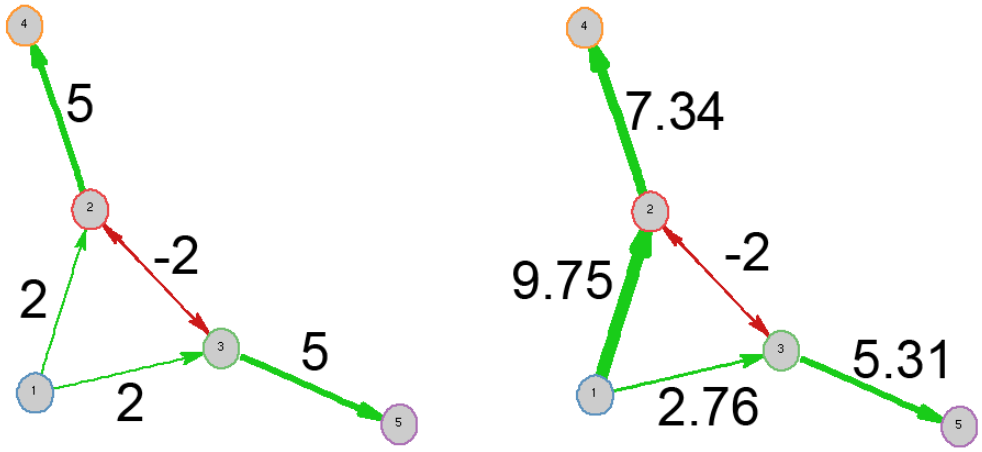


Figure 4.7: The weights of the "brain in a vat" network before (a) and after (b) 8 minutes of learning; initially, the weights of the connective pathway leading to "R" behaviour ($N1 \rightarrow N2 \rightarrow N4$) are the same as the weights of the pathway leading to "L" ($N1 \rightarrow N3 \rightarrow N5$). After learning, the weights of the "R" pathway are strengthened.

Fig. 4.8 shows the results of 10 minutes of learning. As we can see in Fig. 4.8a, the connections leading to "R" behaviour ($N1 \rightarrow N2 \rightarrow N4$) are gradually strengthened, with w_{12} reaching w_{max} around $t = 8m$. The weights of the network before and after learning are shown in Fig. 4.7.

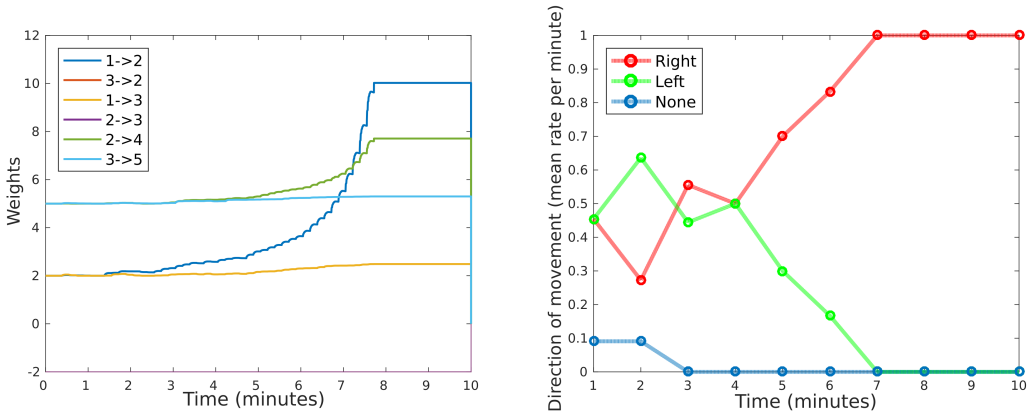
As a consequence of the weight changes, the rate of behaviour "R", initially equal to the "L" rate, increases (Fig. 4.8b); eventually the system reaches a 100% chance of responding with "R" behaviour when stimulated.

4.2 Isolated pair learning experiments

In section 4.1.3, we showed that our "tag & reward" algorithm successfully learns a stimulus-response association when the stimulus and reward are administered artificially. With the parameters that we chose, it took the system around 8 minutes to reach a success rate of 100%, but this might be considerably different in the less controlled case of agent to agent interaction. Learning success rate is a really important parameter in our system, especially in a population chain where one generation's learners become the next generation's teachers.

Our goal then for the rest of this chapter is twofold:

A. We want to show that our learning algorithm works in a joint action, task-oriented



(a) (b)

Figure 4.8: The results of 10 minutes of learning with the "brain in a vat" test system. The behaviour associated with node 4 ("R") is rewarded, so the pathway $N1 \rightarrow N2 \rightarrow N4$ is strengthened (a). After 8 minutes of learning, $w_{12} = w_{max}$ and the learning process stops. While initially the rates of both behaviours are similar, the probability of the system responding with the rewarded behaviour "R" increases over time, eventually reaching 1 (b). Rates are calculated over $1m$.

scenario;

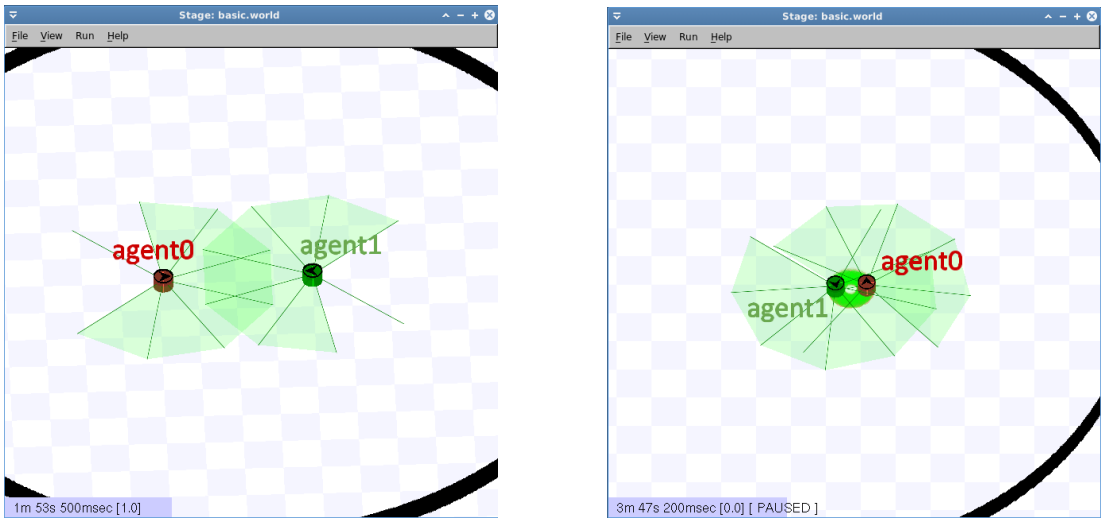
B. We also want to investigate how learning time interacts with final task success rates, the obvious hypothesis being that shorter learning times lead to lower success rates.

In the next sections we will describe an experimental setup that will allow us to investigate points A and B above.

4.2.1 Setup

The isolated pair learning experiments involve two agents interacting. One of the agents is a *learner*, a simulated e-puck agent with the same simple neural network as our "brain in a vat" experiment, shown in Fig. 3.4. The network is using the learning algorithm we detailed in section 4.1. The other agent is an *expert*, a simulated e-puck agent whose behaviour is preset: it wanders around until its proximity sensors detect something in range, then starts turning to the "Right" or "Left" (depending on whether it is an *R-expert* or an *L-expert*).

The two agents are initially placed in their starting positions, facing each other but out of proximity sensor range (Fig. 4.9a); they are then left to interact for a specific



(a) (b)

Figure 4.9: (a): Starting positions for agent0 (expert) and agent1 (learner) in the *Stage* simulator. The agents are facing each other but are not in sensory range. (b): The two agents successfully "dancing"; this is detected and rewarded.

length of time t_{learn}. The agents are automatically reset to their starting position every t_{reset}; each such reset usually leads to a *learning episode*, unless the agents' random walk does not bring them within sensory range of each other.

If the agents successfully dance (Fig. 4.9b; Fig. 4.10a, b), they are rewarded with $DA_r = 1$; if they crash into each other (Fig. 4.10c, d), they are punished with $DA_p = -0.5$. (There is also a small chance that the agents meet, fail to dance but do not crash either; this is usually due to a null response from either of the agents; an example is *agent1* in Fig. 4.10e.)

All response behaviour of the learning agent is logged; after each learning episode (so after t_{reset} seconds) we calculate a rolling mean rate of the last 10 responses for each possible response type ("Right", "Left", "None"). The mean rate of the response type that corresponds to the expert agent ("Right" for *R-expert*, "Left" for *L-expert*) is the *success rate* of the learning agent; the mean success rate in the last two minutes of the agents' interaction is the *final success rate*.

4.2.2 Results: successful learning example

Fig. 4.11 shows an example of successful learning in a pair of agents (*L-expert* and learner). Learning time t_{learn} is set to 15 minutes and t_{reset} is set to 15 seconds. The

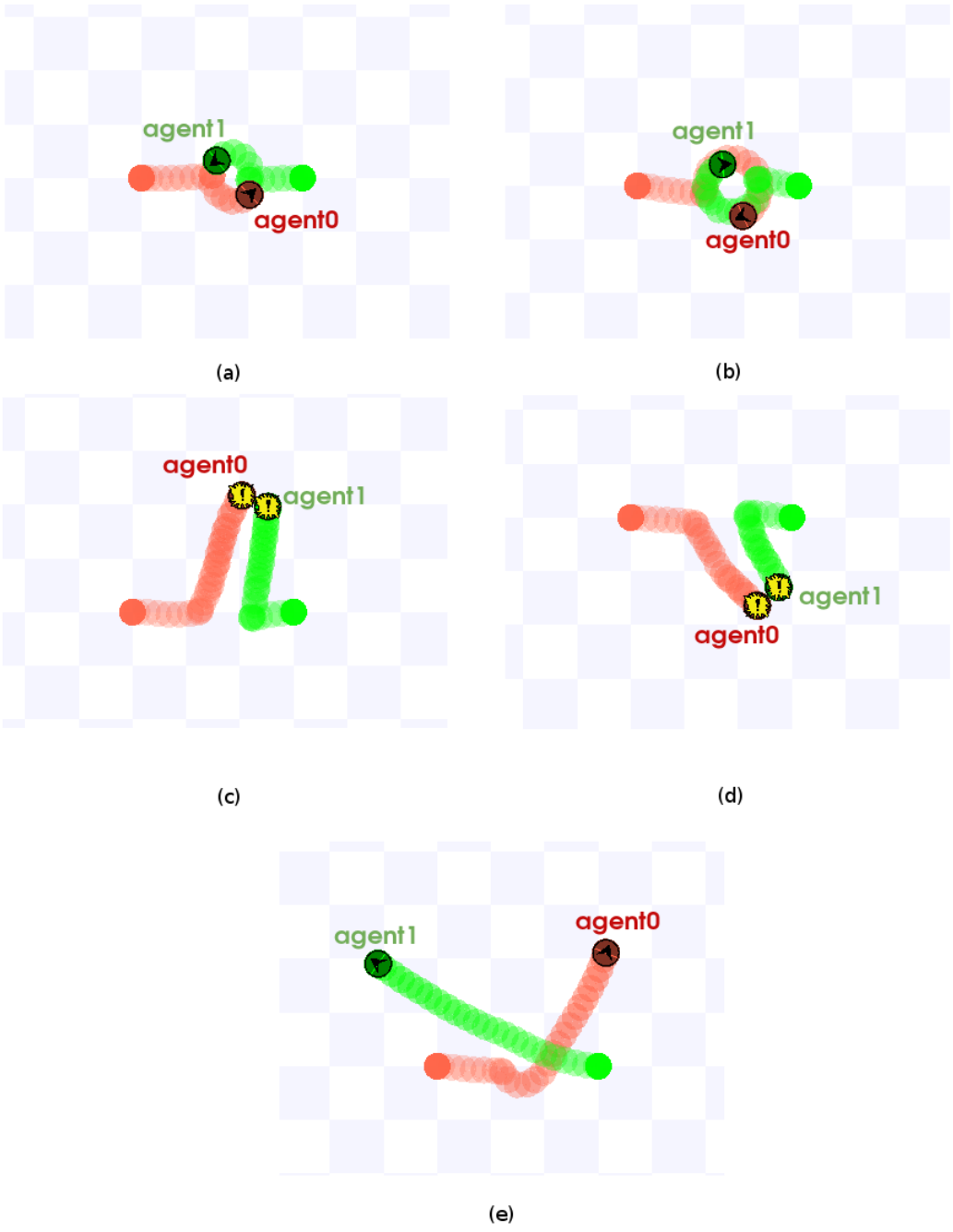


Figure 4.10: Some examples of possible different interactions between two e-puck agents (a teacher, *agent0* and a learner, *agent1*) trying to complete the L-R dancing task. The agents successfully dance if they turn in the same direction ("left-left", a; "right-right", b) and crash if they turn in opposite directions ("right-left", c; "left-right", d). Null responses occasionally lead to the agents missing each other (e; *agent1* produces a null response).

initial direction of the expert agent is reinforced, gradually increasing the weights of the relevant connections (w_{13}, w_{35}) more than the rest in the learning agent's nervous system (Fig. 4.11c). This leads the agent to respond with the rewarded behaviour with a higher rate (Fig. 4.11a, 4.11b). After 6 minutes of learning, the learning agent responds to the expert's proximity only with "Left", so the agents always successfully dance. After 11 minutes, $w_{13} = w_{max}$ and the weights stop increasing any further.

4.2.3 Results: learning time and success rate

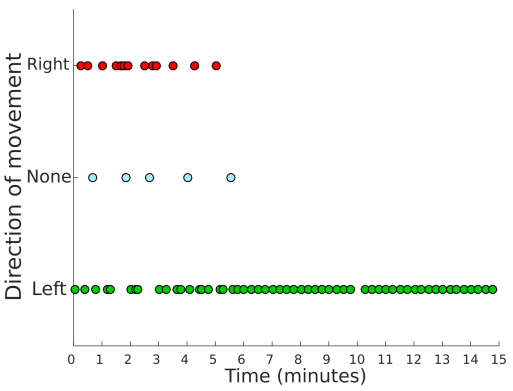
In order to investigate how learning time influences final task success rate, we ran 50 simulations for a number of different learning times from $1m$ to $20m$. We balanced the behaviour rewarded by using an *R-expert* for half of the simulations and an *L-expert* for the other half. In all cases, reset times were set to $t_{\text{reset}} = 15s$, reward values were $DA_r = 1$ and punishment values $DA_p = -0.5$. The distributions of the final success rates of all runs with $t_{\text{learn}} = 5m$, $7m$, $10m$ and $15m$ can be seen in Fig. 4.12.

As expected, longer learning times lead to higher success rates. For a learning time of $5m$ the average success rate is 65%, standard deviation $\sigma = 0.17$ (Fig. 4.12a) while for a learning time of $12m$ almost all agents successfully learn the task (average success rate is 99%, standard deviation $\sigma = 0.05$, Fig. 4.12d). For comparison, the baseline success rate of a non-learning agent is slightly lower than 50%, accounting for the possibility of "no-action" responses to stimulation.[3]

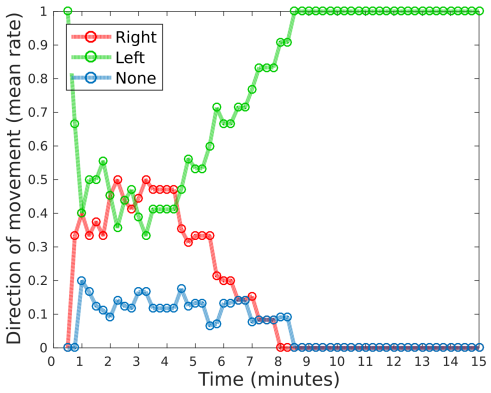
A better overview of this trend can be seen in Fig. 4.13. Fig 4.13a is a plot of the mean final success rate against the learning time t_{learn}; the relation appears to be linear before plateauing at success rates very close to 100% for learning times longer than $12m$. The error bars in Fig. 4.13a indicate the standard deviation.

Fig. 4.13b shows the probability density functions of the final success rates for all learning times tested. Once more, we can see that for shorter learning times the success rates are distributed around values close to 50% with a high spread; long learning times lead to higher mean success rates, with the spread being very low for $t_{\text{learn}} > 12m$.

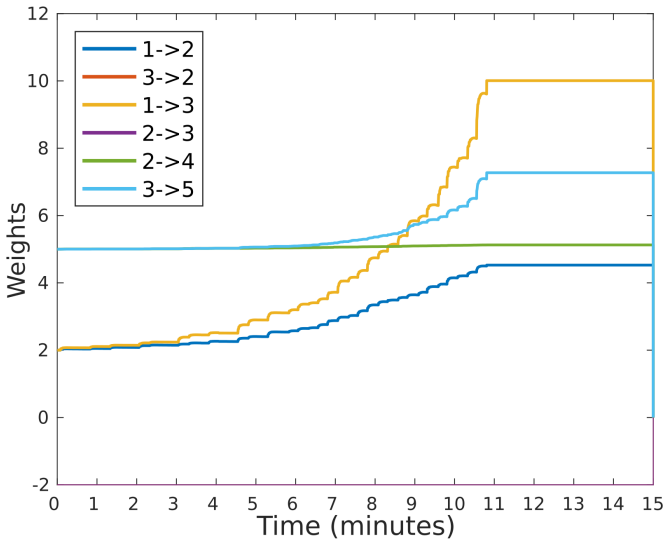
[3] A no-learning test run with artificial stimulation for 20 minutes gave a rate of 2% no-action responses.



(a)

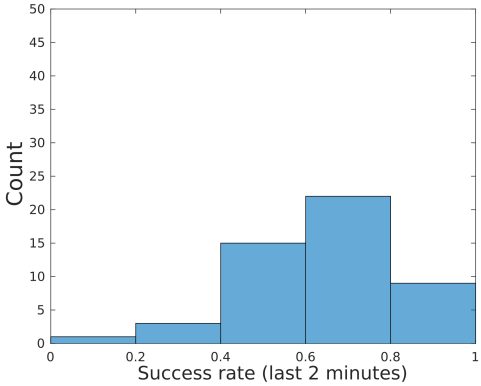


(b)

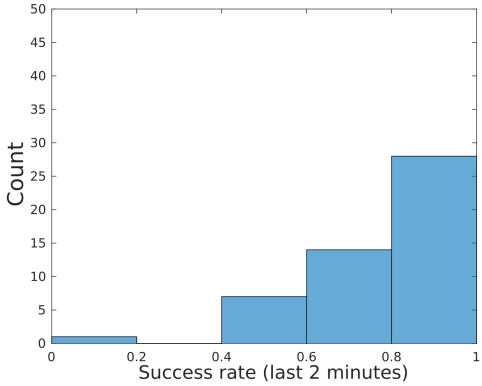


(c)

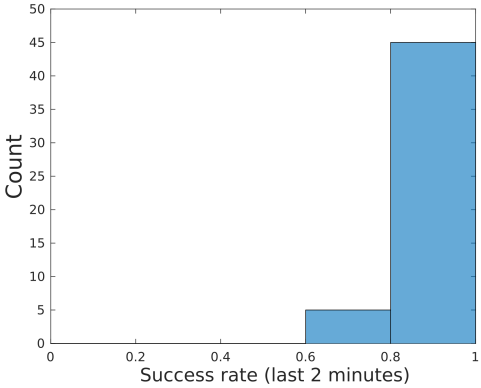
Figure 4.11: An example of a single agent learning by interacting with an *L-expert* agent for $t_{\text{learn}} = 15m$. Fig. (a) and (b) show a scatter plot and a rate chart of the learning agent's behaviour over time: initially the system's response includes all behaviours. As "Left" gets rewarded, the weights of the respective neural pathway w_{13}, w_{35} increase (c) and the system's behaviour gradually shifts to exclusively "Left".



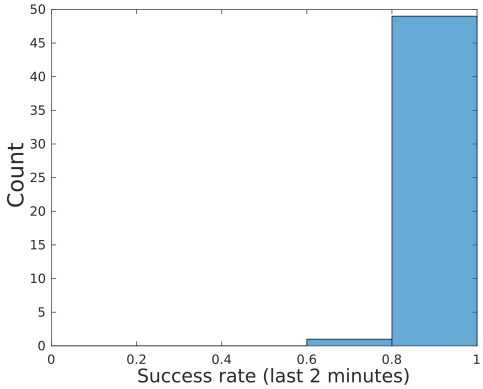
(a) Learning time $t_{\text{learn}} = 5m$. Mean success rate of 65%, standard deviation $\sigma = 0.17$.



(b) Learning time $t_{\text{learn}} = 7m$. Mean success rate of 78%, standard deviation $\sigma = 0.18$.

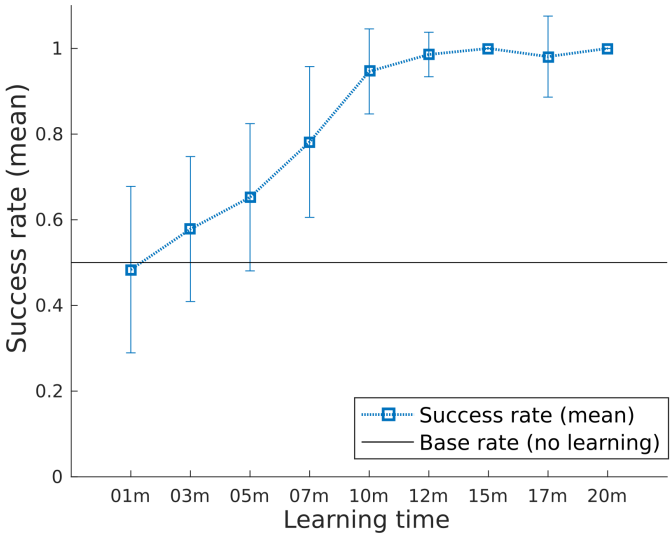


(c) Learning time $t_{\text{learn}} = 10m$. Mean success rate of 95%, standard deviation $\sigma = 0.10$.

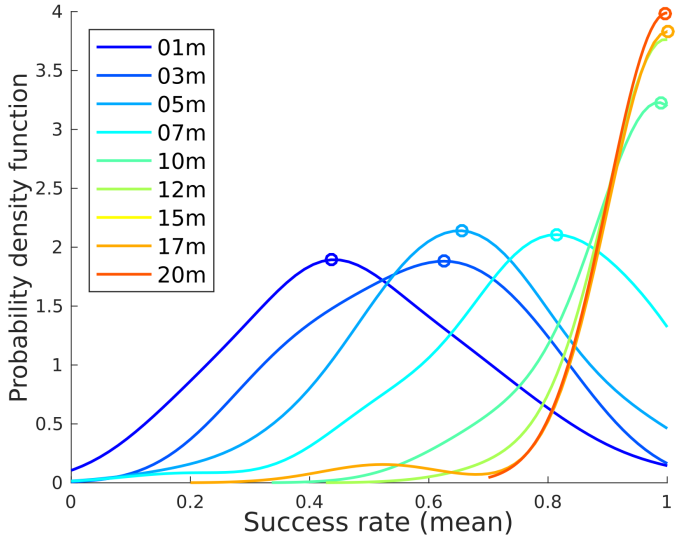


(d) Learning time $t_{\text{learn}} = 12m$. Mean success rate of 99%, standard deviation $\sigma = 0.05$.

Figure 4.12: Distributions of the final success rates for 50 learning simulations (25 using an *R-expert*, 25 using an *L-expert*) for various values of t_{learn}. As expected, longer learning times lead to higher success rates. In all experiments $t_{\text{reset}} = 15s$, $DA_r = 1$ and $DA_p = -0.5$. The baseline success rate of a non-learning agent would be slightly lower than 50%.



(a) A plot of the mean final success rate against the learning time t_{learn}. The relation appears to be linear before plateauing at success rates very close to 100% for learning times longer than $12m$. The error bars indicate the standard deviation.



(b) Probability density functions of the final success rates for all learning times tested. For shorter learning times, the success rates are distributed around values close to 50% with a high spread. Long learning times lead to higher mean success rates, with the spread being very low for $t_{\text{learn}} > 12m$.

Figure 4.13: Effect of learning time t_{learn} on final task success rates. Again, the data shown is the result of 50 learning simulations (25 using an *R-expert*, 25 using an *L-expert*) with $t_{\text{reset}} = 15s$, $DA_r = 1$ and $DA_p = -0.5$.

4.3 Discussion

In the first part of this chapter we detailed a "tag & reward" learning algorithm for CTRNNs, based on the STDP distal reward algorithm proposed by Izhikevich (2007). In the second part, we established that using this algorithm, learning works in isolated pairs of expert and learner agents; and that the success rate of the learning agents in the L/R dancing task depends on how long they interact with the expert agent for.

The implementation we discussed in this chapter has numerous shortcomings; we use a very simple neural network (Fig. 4.5); we make a number of assumptions that are not biologically plausible but that were needed to get a working system (detailed in section 4.1); we use a number of pre-wired behaviours (Right/Left dancing, wandering, crash avoidance). None of these, however, are fundamental aspects of our system, but rather "shortcuts" due to time constraints. Ways around them and future work that would make for a more principled approach will be discussed in Chapter 7.

One aspect of the learning algorithm that we have not touched on so far is its general applicability as a learning mechanism for continuous real-time recurrent neural networks. As we mentioned in Section 3.4, the reason we applied the "tag & reward" learning approach to CTRNNs was that we wanted to avoid the complexity of spiking networks, as used by Izhikevich (2007); but at the same time, implement ontogenetic learning, as opposed to the phylogenetic process of evolutionary computation methods that CTRNNs are usually trained with.

Since we did not test our learning algorithm in a more general setting, its applicability to recurrent neural networks of random topology and more general tasks remains to be seen. In retrospect, however, the loss of spike timing information in CTRNNs means that in order to implement the causal co-activation detection needed for Izhikevich's algorithm, we have to either use a measure that is potentially too simple to scale up to more complex use cases; or use measures that are computationally complex, negating the advantage over using actual spiking networks (see Section 4.1.1, "Causal activation detection").

That said, our goal is not to put forward a cutting-edge model of biologically plausible learning but to examine cultural transmission in the context of a non representational, autopoietic view of cognition. We would argue that, in view of this goal, the results we presented in this chapter are exactly what was needed as a first step: a minimal but self-contained, working system of an isolated episode of transmission of joint action behaviour; that uses biologically plausible mechanisms; and that makes as

few assumptions as possible that are clearly stated and not essential for the learning process.

Furthermore, despite the minimal nature of the system, it exhibits an interesting aspect: learning is not binary. Agents can learn the R/L dancing task in various measures of success, depending on both chance, as the system is stochastic, and, as we saw earlier in this chapter, on how long they interact with each other. This aspect, amplified through a chain of repeated transmission, can potentially lead to complex system behaviour.

In the next chapter, we will take another step towards our goal by building on this chapter's experiments and combining isolated learning episodes in transmission chains using an Iterated Learning framework.

Chapter 5
Transmission chains

In Chapter 4, we established that simulated e-puck *learner* agents can successfully learn the L/R dancing task from an *expert* agent after interacting for a sufficient amount of time. Our goals in this chapter are the following:

A. Show that it is possible to link isolated pair learning instances in cultural transmission chains;

B. Examine how interaction time affects the stability of these transmission chains. Since, as we saw, lower learning times led to worse task success rates, we expect this factor to be a major influence in the stability of the chains.

We will start by detailing the setup of an experiment that will allow us to initialise and investigate transmission chains. After discussing some characteristic examples of stable and unstable chains, we will determine how learning time interacts with chain stability and examine what exactly happens when chains break down. We will close the chapter by looking at cultural inheritance dynamics in transmission chains.

5.1 Setup

A transmission chain starts with the same setup we described in section 4.2.1: two simulated e-puck agents, an *expert* A_0 (a preset agent that always turns either to the left or to the right on proximity detection) and a *learner* A_1 (an agent that uses the "tag & reward" algorithm to learn the correct behaviour). Again, the two agents are placed opposite each other, outside proximity range, and are left to interact for t_{learn}. Every $t_{\text{reset}} = 15$ seconds, both agents are reset to their starting positions (Fig. 4.9a). Dancing is rewarded with $DA_r = 1$ while crashing is punished with $DA_p = -0.5$. The neural network used for the learning agents is the same as the one used for the experiments

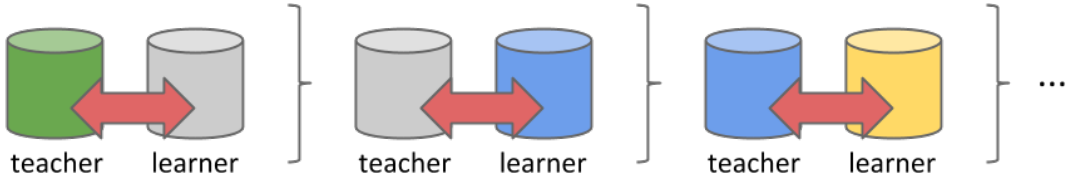


Figure 5.1: A joint action transmission chain. A pair of teacher-learner agents interact, with the teacher removed in each successive generation and the previous learner interacting as a teacher with a new learner. This process is repeated for N generations.

in Chapter 4, seen in Fig. 4.5. Initial weights are $w_{12} = w_{13} = 2$, $w_{23} = w_{32} = -2$ and $w_{24} = w_{35} = 5$.

At the end of the interaction time t_{learn}, the expert agent A_0 is removed and a new learner agent A_2 is introduced; the previous learner agent A_1 now assumes the role of the teacher and the agents are left to interact for a further t_{learn}. This process of gradually replacing agents while alternating learning and teaching roles for each agent is repeated for N generations (Fig. 5.1). The lifetime of each agent is thus $2 \cdot t_{\text{learn}}$, except for the initial expert agent A_0 and the last agent A_{N-1} who lack the learning and teaching phases respectively, so their lifetimes are t_{learn}. An algorithmic flowchart of the process is shown in Fig. 5.2.

Note that nothing internal to the agent changes when they assume the role of a teacher rather than a learner. In a way, there is "lifelong learning" as even when an agent is teaching, the learning algorithm is still functioning and the agent's synaptic weights can still change via reward or punishment. The only difference between the learning and teaching phases of an agent is how long the agent has been active for and whether their partner is "older" or "younger" than themselves.

This is a departure from the usual Iterated Learning models. In the simulation experiments (Section 2.2.1.1), the training and teaching phases are completely separate; once an agent becomes a teacher, the learning process stops. In the case of human experiments (Section 2.2.1.2), there is no contact between "teachers" and "learners". The subjects only operate as teachers implicitly, when the language they produce is used by the experimenters to train the next generation of participants.

However, we feel that this decision is in line with the "simplest approach" methodology we have adopted; switching between completely different teaching and learning behaviours requires further assumptions than letting the agents to continue learning even as teachers, so such a design decision would have to be justified. Furthermore,

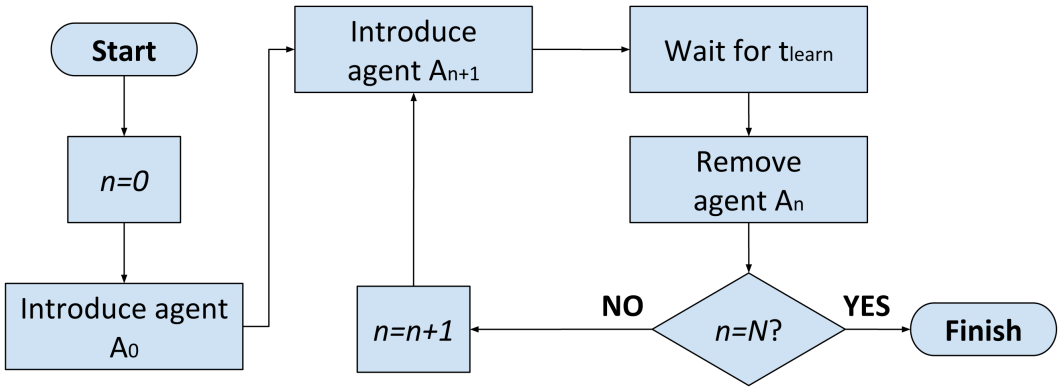


Figure 5.2: A flowchart of the transmission chain process. Pairs of agents interacting alternate between learning and teaching roles for N generations; in each generation, the older agent is replaced by a new learning agent.

as we will see later on in this chapter, "lifelong learning" allows agents to exhibit complex and interesting behaviour that would be otherwise suppressed if the agents stopped learning when they switched to their teaching role.

If all instances in this process of transmission are successful, by the end of the chain of agents the behaviour of the initial expert agent ("Right" or "Left" for *R-experts* and *L-experts* accordingly) should be preserved. As opposed to the isolated pair experiments, however, this time the teachers are not expert, pre-set agents that always react with a consistent response; instead, they are learners from the previous generation (with an exception, of course, for the initial agent A_0 starting the chain). In the next section we will see a number of transmission chain examples: some stable, some breaking down in different ways.

5.2 Examples of transmission chains

Fig. 5.3 shows five examples of different transmission chains resulting from the experimental setup we described. In all five cases the chains were run for $N = 10$ generations; the initial agent in cases (a), (d) and (e) was an *R-expert* and in cases (b) and (c) an *L-expert*. For each agent, we recorded the mean response rate for each possible behaviour ("Right", "Left", "None") over the last 2 minutes of its *learning* period (in other words, over the last two minutes of t_{learn}, the agent's first half of its lifetime; the other half being its *teaching* role).

(a) The most straightforward case is a working example of a stable chain for $t_{\text{learn}} = 20m$ (Fig. 5.3a). With such a long learning time, each agent's neural network is very likely to reach $w = w_{max}$ for the connective pathway that leads to the correct response. This means that all agents' behaviour is very robust, leading to a stable chain. While not very interesting, this example establishes that transmission chains are indeed possible in our system.

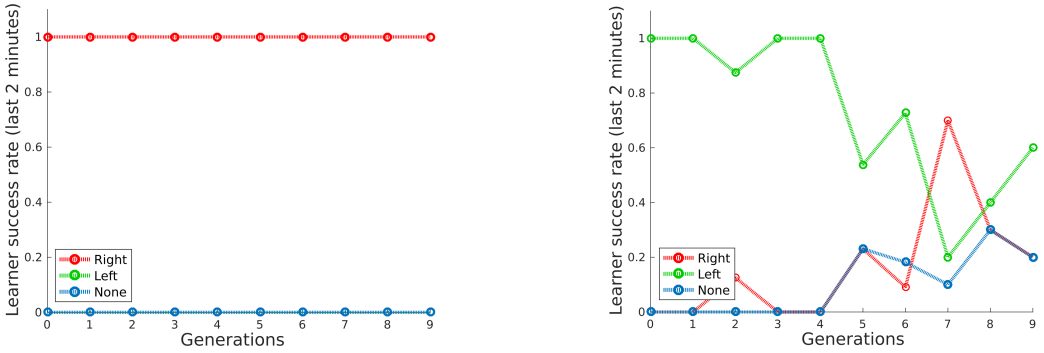
(b) The chain shown in Fig. 5.3b ($t_{\text{learn}} = 12m$) breaks down after 5 generations. For the rest of the chain after this breakdown, neither of the agents in each pair is proficient at the task; the agents, then, do not have anyone to learn from and revert to chance-level responses for generations 6 to 9. While the task success rate falls below 100% on generation 2, it does not drop further to chance-level rates; instead, it increases back to 100%, recovering the chain before its eventual collapse.

(c) Fig. 5.3c ($t_{\text{learn}} = 10m$) is another example of a successful chain recovery. While agents A_3 and A_4 drop to very low task success rates, agent A_5 reverses this drop and the chain remains stable until the end of the experiment. This is an example of the learners' ability to *regularise* and perform better than their teachers.

(d) On the other hand, Fig. 5.3d ($t_{\text{learn}} = 12m$) is an example of a stable chain that sustains a success rate lower than 100%. Learners in this chain do not regularise (increasing the success rate back to 100%) but instead seem to be *probability matching* by adopting similar success rates as their teachers. We will return to discuss regularising and probability matching behaviour later in this chapter.[1]

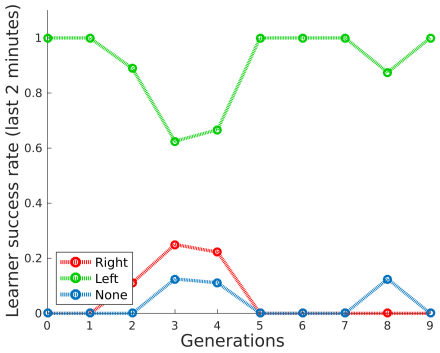
(e) Finally, Fig. 5.3e ($t_{\text{learn}} = 10m$) showcases an interesting case of *behaviour switching*: the transmission chain breaks down as soon as generation 5, but straight after the breakdown, both agents (A_5 and A_6) happen to turn in the same direction, opposite to the one the chain was initialised with. This leads to "Left" behaviour being reinforced instead of "Right", even though there was never any *L-expert* in the chain. This new "Left" chain persists for 3 generations, after which point it also breaks down. We will examine behaviour emergence in detail in Chapter 6.

[1] We are borrowing this terminology from Hudson Kam and Newport (2005).

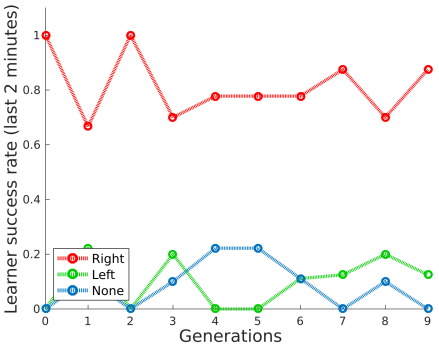


(a)

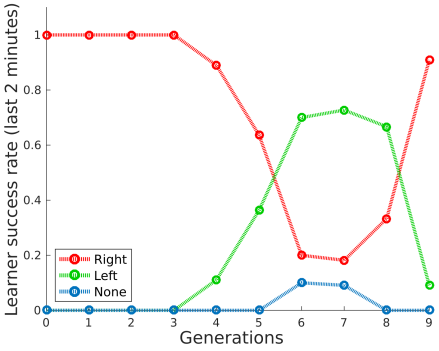
(b)



(c)



(d)



(e)

Figure 5.3: Five different transmission chain behaviours ($N = 10$ generations). *R-expert* initial agent in cases (a), (d) and (e) and *L-expert* in (b) and (c). (a): Stable chain with consistent 100% success rate; $t_{\text{learn}} = 20m$. (b): Chain breakdown after 5 generations; $t_{\text{learn}} = 12m$. (c): Task success rate drops after 3 generations but the drop is reversed and a breakdown prevented; $t_{\text{learn}} = 10m$. (d): Stable chain with $< 100\%$ success rate; $t_{\text{learn}} = 12m$. (e): Chain breakdown after 4 gen. with behaviour swap; $t_{\text{learn}} = 10m$.

5.3 Learning time and chain stability

As we saw in section 4.2.3, shorter learning times in isolated expert/learner pairs of agents lead to lower task success rates. It makes sense, then, that shorter chain interaction times will increase the odds of a chain breaking down in any given generation. In order to investigate this effect, we ran 10 transmission chain experiments, half of which were initialised with an *R-expert* (the other half being initialised with an *L-expert*) for a number of t_{learn} values ranging from 1 to 20 minutes. After each generation, we kept track of which of the chains had broken down (which we defined as having task success rates lower than 60%). The results are plotted in Fig. 5.4.

We can see a clear trend in the data: for any given generation, lower learning times lead to higher chain failure rates.[2] Almost all of the chains with $t_{\text{learn}} = 1m$ fail instantly (from generation 1), while chains with $t_{\text{learn}} = 3m$ or $5m$ mostly fail by generation 3. As the learning times increase, the fail rate curves grow less and less steep, becoming completely flat for $15m$ and $20m$ learning times (as none of those chains fail before the end of the experiments).

A better view of this trend is given by the half-life plot (Fig. 5.5). The *half-life* score associated with chains of a given learning time value t_{learn} is defined as the generation by which half of those chains have broken down. (The 12, 15 and 20 minutes t_{learn} chains do not have half-life scores, as more than half of the chains were still stable by the end of the experiments.) Again, Fig. 5.5 shows that longer learning times lead to more stable chains on average, an expected result that validates the hypothesis in the beginning of this chapter.

A more surprising result is the failure rate of $12m$ learning time chains: by generation 9, 4 out of the 10 chains have broken down. In the isolated pair learning experiments (Chapter 4), agents learning for $t_{\text{learn}} = 12m$ had a 99% success rate (Fig. 4.12d), the same as agents learning for $20m$; when linking the isolated experiments into chains, however, the shorter learning time seems to be making a difference. To investigate this effect we need to take a closer look into the agent interaction that leads to a chain breaking down.

[2]Note that for very short learning times of 1, 3 and 5 minutes the failure rate is calculated over the last 1 minute of agents' learning period, while for longer chains the failure rate is calculated over the last 2 minutes.

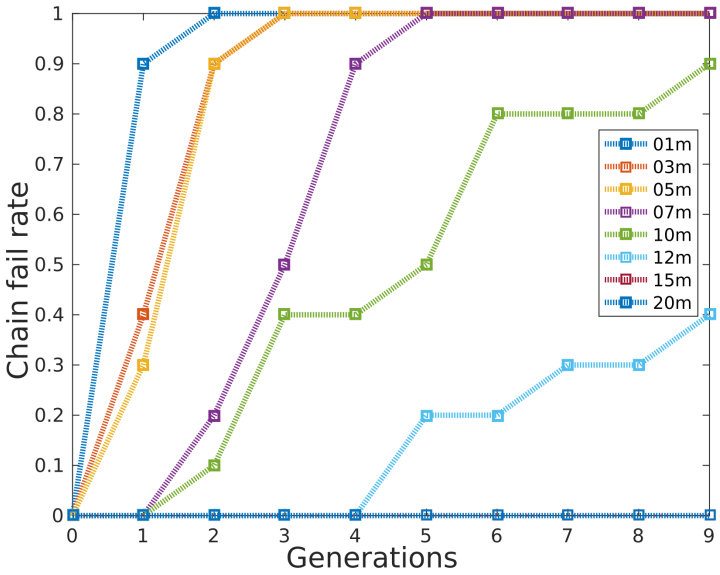


Figure 5.4: Chain fail rate by generation (lower is better). "Failed" chains are chains whose agent success rates have fallen below 60%. All of the chains with $t_{\text{learn}} = 1m$ fail from the 1st generation; none of the chains with $t_{\text{learn}} = 15m$ and $20m$ have failed by the end of the experiment (10th generation). Between these two extreme values, there is a clear trend of longer learning times leading to more stable chains.

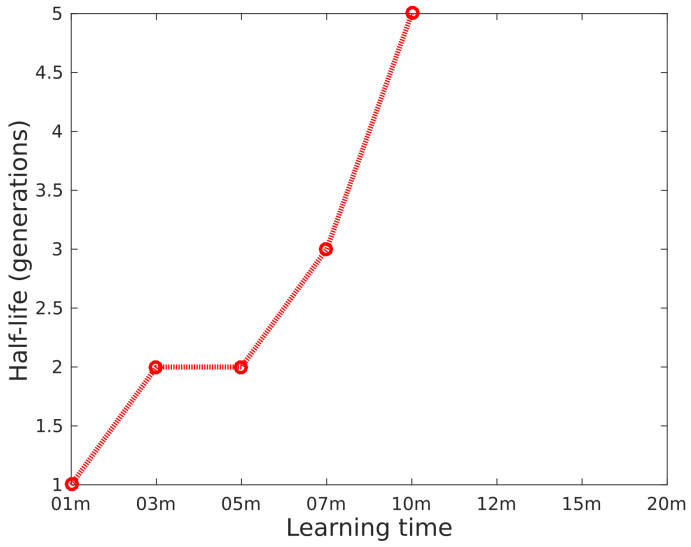


Figure 5.5: Half-life graph for chains of various learning times t_{learn}. The half-life score is the generation by which at least half of the chains with t_{learn} have broken down. 12, 15 and 20 minutes t_{learn} chains do not have half-life scores as more than half of the chains were still stable by the end of the experiments. Again, there is a clear trend of longer learning times leading to more stable chains.

5.4 Chain breakdowns

Fig. 5.6a is a plot of the success rates in one of the 12 minute learning time chains that break down. Generations 1 to 4 seem to be stable with very high success rates; the success rate of agent A_5, however, suddenly drops to 54%. In order to understand the cause of this drop, we need to look at a measure internal to the agents: the w_L/w_R ratio. This ratio, calculated after each agent's learning time, is a measure of the strength of the connective pathway leading to "Left" behaviour ($1 \rightarrow 3 \rightarrow 5$) compared to the strength of the pathway leading to "Right" behaviour ($1 \rightarrow 2 \rightarrow 4$). A high ratio leads to very stable "Left" responses from the agent, while a ratio close to 1 leads to an equal number of "Right" and "Left" responses.

The weight ratio in the above chain (Fig. 5.6b) starts high but falls abruptly in generation 2, and even further in generation 4. This is not reflected in the agents' task success rates: any ratio above a certain threshold leads to consistent behaviour. It does, however, affect their teaching ability and especially their resilience to "uncooperative" learners. (We place "uncooperative", here, in brackets as there is nothing inherent in the learning agent that makes them less cooperative in their learning role; the only stochastic element in the system is the noise, both internal to the neural network and external as part of the proximity sensors, and it is only "bad luck" that makes it so that a learning agent repeatedly produces the wrong response. In that way, a better description would be "unlucky" instead of "uncooperative" agents.)

Fig. 5.7 shows an overlapping timeline of the interaction between agents A_2 to A_6 in the same transmission chain. The weights of the agents change as they are rewarded or punished, both during their learning and teaching phases. (The weights shown are not individual weights, but rather combined weights for each connective pathway; w_R for the pathway leading to "Right" responses and w_L for the one leading to "Left" responses.) The switch between phases is indicated by vertical bars, while the coloured dots are points in time at which an agent's response does not lead to successful dancing; since the initial direction of this specific chain is "left", these incorrect responses are "right" (red dots) and "none" (blue dots).

The first two agents shown, A_2 and A_3, are successful as both learners and teachers. At the end of their learning period (minutes 12 & 24 respectively) they have high task success rates and their w_L/w_R weight ratios, while not very high, are above 1.5. During their teaching periods, both agents always respond correctly ("left") and by the end of their lifetime the weights of the connections leading to this response reach w_{max}.

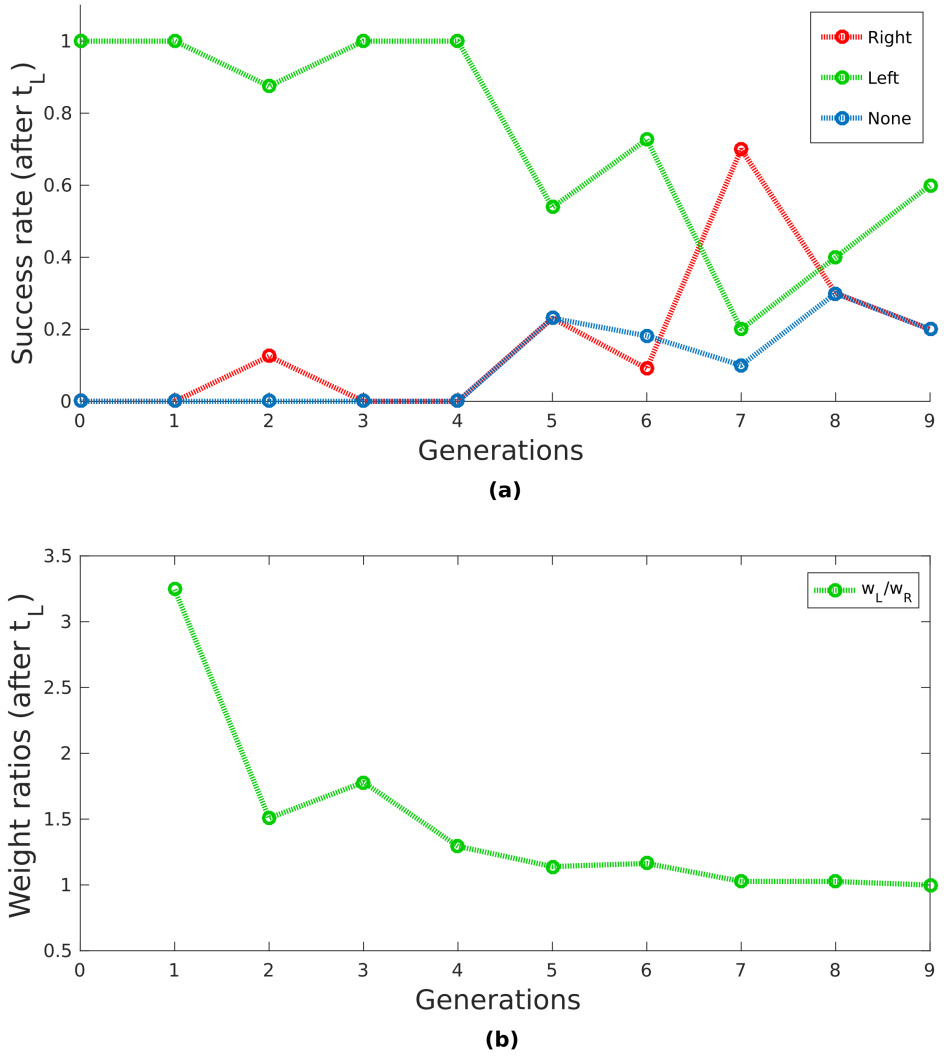


Figure 5.6: A $t_{\text{learn}} = 12m$ chain breaking down. (a): Task success rates (in this specific chain, "Left" response rates) are high for the first 4 generations then suddenly fall below 60%. (b): The weight ratio w_L/w_R is a measure of the strength of the connective pathway leading to "Left" behaviour compared to the strength of the pathway leading to "Right" behaviour. The weight ratio starts high, but falls abruptly in generation 2 and even further in generation 4. This is not reflected in the agents' task success rates but it affects their teaching ability. Both success rate and weight ratios correspond to the learning role of each agent.

This changes, however, with agent A_4: while at the end of the learning period A_4's task success rate is 100%, its w_L/w_R weight ratio is very low ($= 1.29$). This is reflected in A_4's performance as a teacher: paired with the uncooperative agent A_5, agent A_4 produces the wrong response on a number of occasions (shown as red and blue points in Fig. 5.7). This can have the effect of further decreasing the teacher's own weight ratio if the learning agent happens to produce the same response that leads to successful dancing in the "wrong" direction. Regardless of whether this happens or not, however, the inconsistency in the teacher's behaviour leads to low task success rates at the end of agent A_5's learning period. This further leads to both learning and teaching deficiencies for the next agents A_6 and A_7, the weight ratios of which get even closer to 1. The agent's responses become increasingly random and the chain eventually breaks down.

It seems, then, that there is a trait inherited from generation to generation of agents that is not reflected in the task success rates: an agent can be a good "learner" but a bad "teacher".[3] This answers the question we posed at the end of section 5.3 and explains how a learning time of $t_{\text{learn}} = 12m$ leads to perfect task success rates in isolated learning experiments (Fig. 4.13a) but is not enough for long-term stable chains (Fig. 5.4). In the next section we will take a closer look at the inheritance of both task success rates and weight ratios in transmission chains.

5.5 Cultural inheritance dynamics

Single chain plots, like the ones shown in Fig. 5.3, cannot be used to identify any trends in the transmission from generation to generation. In order to do that, we will look at cultural *inheritance dynamics* plots, state plots of a certain measure (in our case, success rate or weight ratio) inherited from one generation of agents to the next in transmission chains. Each point in the plot represents a transmission episode, the x-coordinate being the measure of the teacher (generation N) and the y-coordinate the equivalent measure of the learner (generation N+1). Fig. 5.8 shows a number of success rate inheritance plots for chains of various learning times. Each plot includes data from 10 separate chain experiments of the same learning time; each of those chain experiments use 9 learning agents, which means that there are 8 episodes of

[3] We used the weight ratio as a measure of the teaching ability of an agent, but a number of other measures could be used instead. One obvious example would be the count of "wrong" responses an agent produces during their teaching period; this has the advantage of being a behavioural measure, but it cannot be estimated before an agent actually goes through that teaching period.

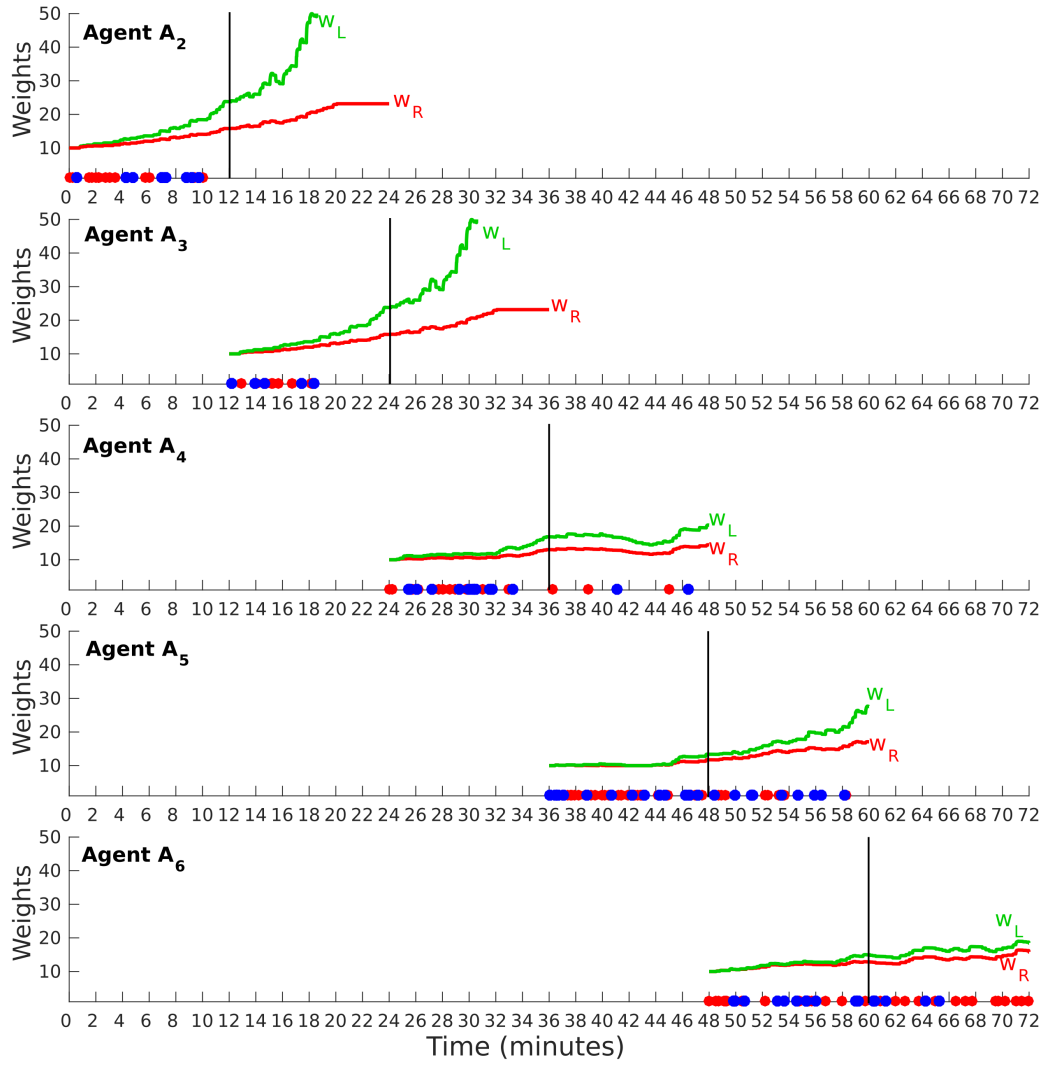


Figure 5.7: A series of graphs showing the interaction between agents A_2 to A_6 in a transmission chain. The weights of each agent, as they change while learning, are shown in the plotted lines; w_R is the combined weight of the connective pathway that leads to "Right" responses and w_L is the equivalent combined weight for "Left" responses. Vertical bars indicate the switch between learning and teaching roles. The coloured dots are points in time at which an agent's response does not lead to successful dancing ("right" responses in red and "none" responses in blue). The decreasing w_L/w_R weight ratio (A_4) leads to teachers that are less robust to uncooperative learners (A_5); eventually, the weight ratio becomes too low and the chain breaks down after agent A_6.

transmission, for a total of 80 data points per plot. Some observations:

1. The diagonal line drawn on each plot is the identity line ($x = y$) and it represents a perfect transmission of task success rates from teacher to learner. Unless the chain is perfectly stable, as in the case of the $t_{\text{learn}} = 20m$ chains where all agents have success rates of 1, most transmission points deviate from this line. Points *below* the line ($x > y$) represent the chain's success rate decreasing, since the success rate of generation N is higher than that of the next generation. Points *above* the line ($x < y$) represent the success rate increasing.

2. Stable chains are represented by points in the upper right corner of each plot. In perfect chains ($t_{\text{learn}} = 20m$) all points converge to $(x, y) = (1, 1)$; stable but not perfect chains ($t_{\text{learn}} = 15m$) include points with $x = 1$, $y < 1$ (drops in success rate) but also symmetric points with $x < 1$, $y = 1$ which indicate recoveries (Fig. 5.8a). Chains that have broken down are represented by points in the area around $(x, y) = (0.5, 0.5)$. The spread of these points however is high, since as we saw in section 4.2.3 (Fig. 4.12a) the responses of agents that operate at chance levels have high variance.

3. Vertical groups of points with $x = 1$ (Fig. 5.8b) represent success rates dropping from 100% to various lower values; horizontal groups of points with $y = 1$ (Fig. 5.8c) represent success rates increasing to 100% from various lower values. These groups are only present in chains with learning times high enough to have non-expert agents with 100% success rates but low enough to reliably drop from these high rates ($t_{\text{learn}} = 10$ to $12m$).

Fig. 5.9 shows a number of inheritance plots, this time of weight ratios instead of success rates. Again, each plot includes data from 10 separate chain experiments of the same learning time. The weight ratio we used for these plots was, as before, the w_L/w_R ratio. This time, however, since half of the chains of each learning time started with an *R-expert*, successful chains do not only correspond to high weight ratio values. Instead, successful "Right" agents have weight ratios that are much lower than 1, while successful "Left" agents have weight ratios that are much higher than 1. In order to accurately depict the difference between ratios, the plots are in logarithmic scale. Once again, some observations:

1. Low learning times ($t_{\text{learn}} = 1$, 3 and 5 minutes) lead to agents with small synaptic weight differences and $w_L/w_R \approx 1$. Inheritance data points are gathered

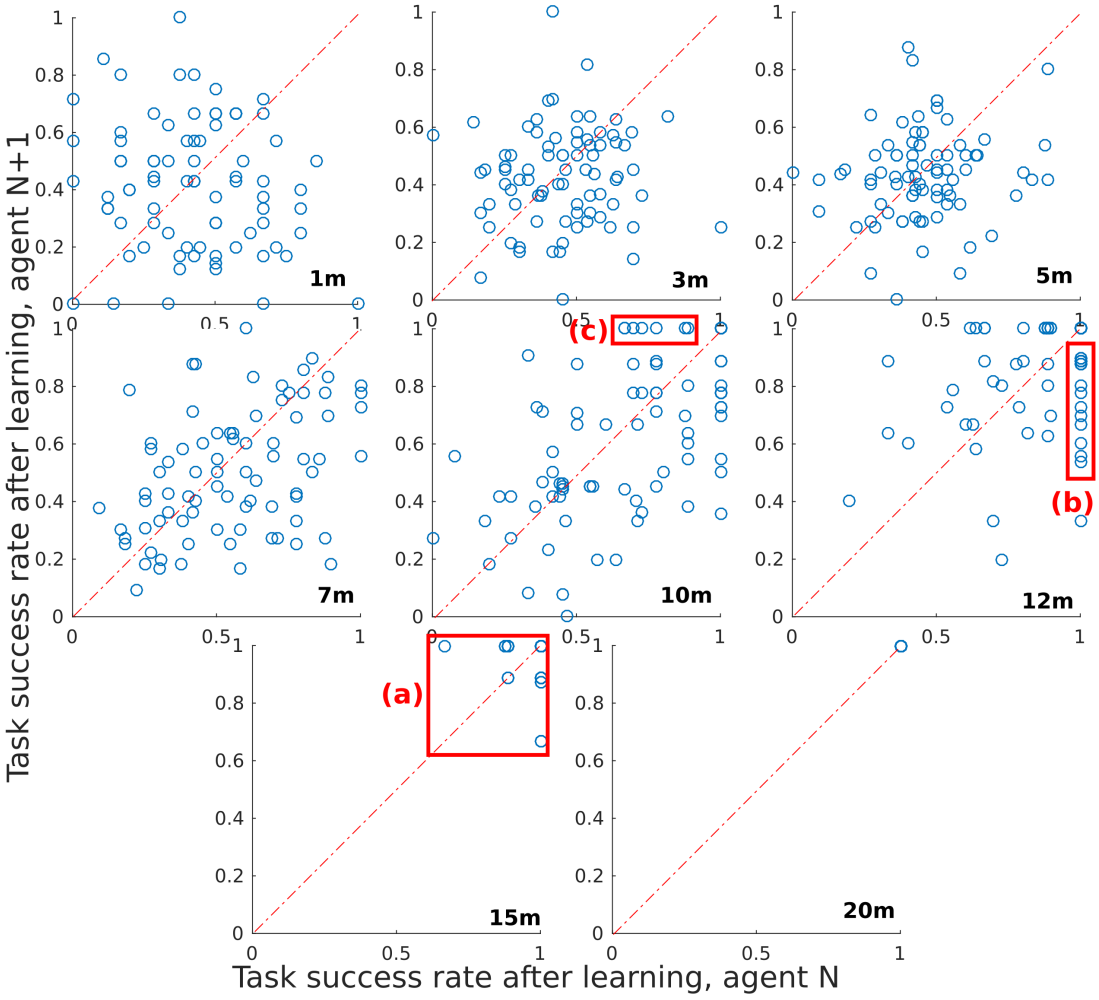


Figure 5.8: Success rate inheritance dynamics plots for chains of various learning times (state plots of the success rates inherited from one generation of agents to the next in transmission chains). Each point represents a transmission episode, the x-coordinate being the success rate of the teacher (generation N) and the y-coordinate the success rate of the learner (generation N+1). The diagonal identity line in each plot represents a perfect transmission of task success rates. Points below the identity line are drops in success rate, while points above the line are recoveries. (a): Points in the upper right corner indicate stable chains; each "drop" point is balanced out by a symmetric "recovery" point. (b,c): Vertical groups of points represent moments at which success rates drop from 100% to various lower values; horizontal groups represent recoveries to 100%.

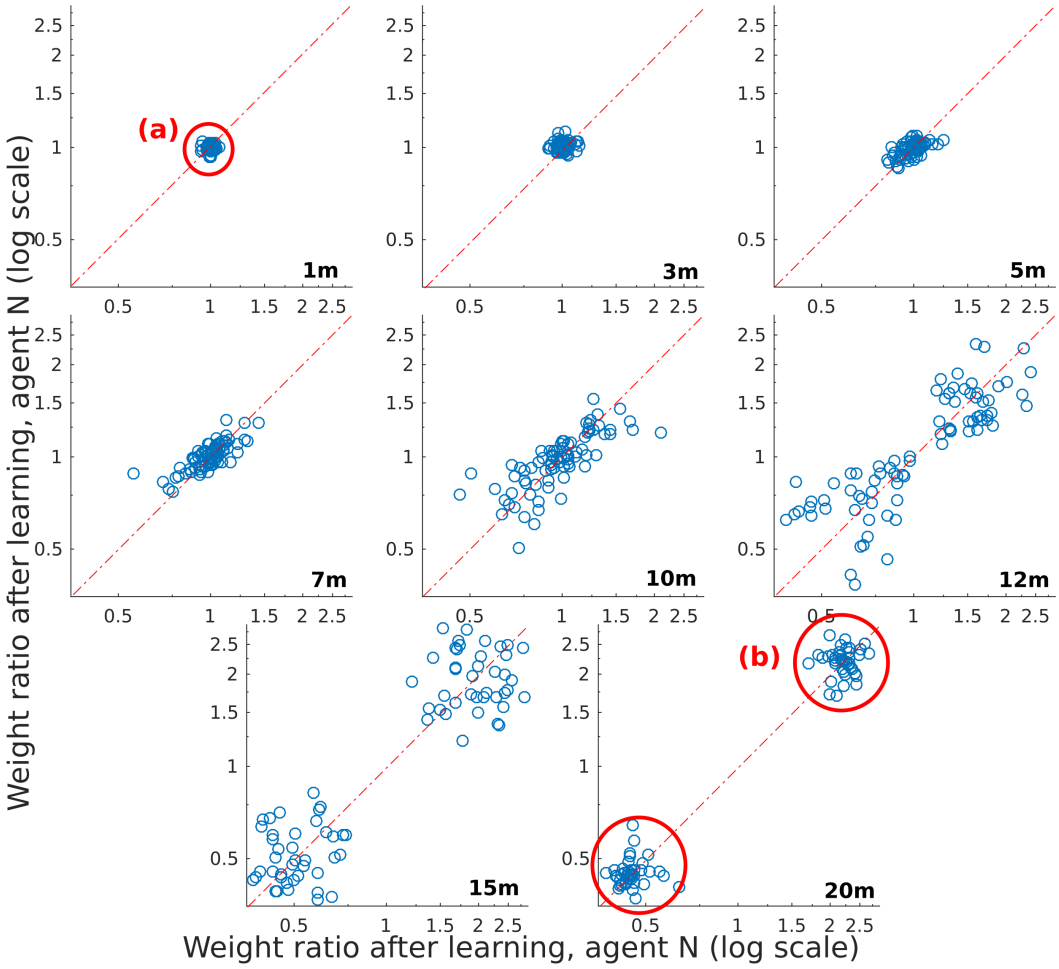


Figure 5.9: Weight ratio inheritance dynamics plots for chains of various learning times. Each point represents a transmission episode, the x-coordinate being the w_L/w_R weight ratio of the teacher (generation N) and the y-coordinate the w_L/w_R weight ratio of the learner (generation N+1). The plots are in logarithmic scale. (a): Chains with very low learning times break down instantly and operate at chance levels, with $w_L \approx w_R$ for all generations; all inheritance data points are gathered around the $(x, y) = (1, 1)$ attractor. (b): As learning time increases, two additional attractors appear in the upper right ("Left" behaviour) and bottom left ("Right" behaviour) corners.

around the $(x,y) = (1,1)$ attractor, with all chains failing quickly and agents reverting to chance response levels (Fig. 5.9a).

2. As learning time increases, the inheritance data points get pulled apart into the corners of the plot, indicating two new "successful chains" attractor points: "Right" in the lower left corner and "Left" in the upper right corner. Most $t_{\text{learn}} = 10m$ chains fail by generation 9 (Fig. 5.4) and all agents of those chains end up in the (1,1) attractor. For $t_{\text{learn}} = 12m$, however, the inheritance data points start getting drawn into the two new "Right" and "Left" attractors. Finally, for longer learning times ($t_{\text{learn}} = 15$ and 20 minutes), only the two stable "Right" and "Left" attractors remain; the weight difference is significant enough that none of the chains break down (Fig. 5.9b).

3. In comparison to the task success rate inheritance plots (Fig. 5.8), the inheritance data points in the weight ratio plots appear closer to the "perfect transmission" identity line. We can confirm that this is the case by comparing the R^2 score of the "fit" of the $x = y$ line for the points of each of the two conditions.[4] The distance for each of the learning times is given in Table 5.1 below; "SR" stands for success rate while "WR" is the weight ratio.

	1m	3m	5m	7m	10m	12m	15m	20m
SR	-1.2269	-0.9449	-0.8894	-0.3604	0.1025	-0.2467	-0.7068	1.0000
WR	-1.2068	-1.6615	-0.5618	0.2502	0.2221	0.6204	0.6995	0.9459

Table 5.1: R^2 scores for success rate (SR) and weight ratio (WR) inheritance data points. These scores show how good the fit of the $x = y$ line is to the data points; a good fit (with an R^2 score closer to 1) indicates a tendency of learners to more closely match their teachers.

For very short learning times (1, 3 and 5 minutes), the negative numbers indicate that the $x = y$ identity line is a very bad fit for the inheritance data points: high variance leads to very low inheritance from one generation to the next for both success rates and weight ratios. This, in fact, continues to be the case across

[4]The R^2 score is given by calculating the mean squared vertical distance of all points from the $x = y$ line; dividing it by the mean squared vertical distance from a horizontal line going through the mean y value of all points; and subtracting this quotient from 1. Note that despite the "R^2" is not actually squared; it can be a negative number if the fit is bad enough.

all learning times for the success rates, indicating that learning agents do not tend to match the success rates of their teachers. (As an exception, all agents for $t_{\text{learn}} = 20m$ have success rates equal to 100%, which leads to all the inheritance points being on the $x = y$ line and a perfect R^2 score of 1.) The comparatively higher R^2 scores for weight ratios for learning times of 7, 12 and 15 minutes, however, show that agents tend to match the weight ratios of their teachers more than their success rate.

5.6 Discussion

In this chapter, we documented an experiment that put together isolated pair learning episodes (as described in Chapter 4) in cultural transmission chains. After demonstrating a number of different examples of such chains (some stable, some breaking down) we examined the connection of agent learning time to chain stability; detailed the interactions between agents that lead to chains breaking down; and looked at the cultural inheritance dynamics of the transmission chains. Before moving on to the next chapter, let us make some brief comments related to the experiments we have described so far.

Stable chains are possible: The stable chain examples for learning times of $t_{\text{learn}} =$ 15 and 20 minutes are cultural transmission chains that show similarities to the chimpanzee transmission chains (Horner et al., 2006) we mentioned in Chapter 2: an agent trained in accomplishing a task using one of two possible behaviours ("Right", "Left") transmits this initial behaviour to posterior generations using a process of iterated learning through joint action. This effectively demonstrates the claim we put forward in section 2.3: it is possible to build a system that is consistent with an autopoietic approach to cognition and able to exhibit a chain of cultural transmission of behaviour. In Chapter 7, we will return to this point and examine it in relation to the autopoietic design points that we discussed in Section 2.1.2.2.

Teaching vs. learning: An interesting point that emerged through the chain experiments is that an agent's performance as a learner is not necessarily indicative of their performance as a teacher. For a given learning time, the learner success rates can be perfect (as we saw in section 4.2.3, Fig. 4.13a for learning times of 12 minutes); using these learners as teachers for the next generation, however,

leads to transmission chains that often break down (Fig. 5.4). We can identify two reasons that may cause this difference between an agent's learning and teaching abilities.

The first reason is the small sample size we use for our calculation of an agent's post-learning task success rate. We are calculating this success rate over the last 2 minutes of learning time; while learning, agents are reset to their initial positions after 15 seconds, which means that those last 2 minutes correspond to 8 interaction attempts, a sample size that might not be enough to identify smaller changes in task success rates. On the other hand, the results shown in Fig. 4.13a are calculated over 50 different experiments, effectively increasing the sample size to a point where even a small difference in response percentage would be visible.

A more plausible explanation is that a *thresholding effect* is applied to the agents' success rates. All weight ratios[5] above a certain threshold lead to ceiling success rates for a learner; further weight ratio increases are not visible when using success rate as a measure. During an agent's teaching role, however, uncooperative students often produce responses that lead to crashes. Since the teacher agent never stops learning, these crashes decrease the weights of the "correct" neural pathway, bringing in turn the teacher's weight ratio closer to 1. If the ratio was already close enough to 1, this further change might be enough to influence the teacher's correct response rates. At that point, as we saw in Fig. 5.7, the lower weight ratios eventually lead to the chain breaking down.

Probability matching vs. regularisation: A further question arising from the transmission chain results in this chapter is whether learning agents tend to inherit task success rates similar to their teachers, making them *probability matchers*, or tend to drive success rates towards binary values of success (100%) or failure (50%), making them *regularisers* (Hudson Kam and Newport, 2005).

This is an interesting question, as it has some parallels with language learning; when presented with variable forms in a language where their use is probabilistic rather than predictable, *adult* learners tend to produce those forms with the same probability they appear in the learned language ("probability matching", while

[5] We have been using the weight ratio w_L/w_R as a "biological", internal measure of how consistent an agent is in producing the right response to a stimulus; any other equivalent internal measure would be equally usable instead.

child learners tend to "regularise", producing only the form that appears with the highest probability (Smith and Wonnacott, 2010). (Once more, we will revisit this discussion in section 7.6.)

In any case, the weight inheritance plots seem to suggest that agents tend to regularise for success rates: vertical groups with $x = 1$ and horizontal groups with $y = 1$ (as shown in Fig. 5.8b and c respectively) represent drastic decreases and increases of success rates from and to 100%.

On the other hand, there are also examples of success rate matching (such as the chain shown in Fig. 5.3d). In order to get a clearer picture, we need to look at how stable longer probability matching chains are; we also need to look at what happens over time to chains that have broken down. In order to do that, 10-generation chains are not enough; we will return to this topic after looking at longer chains in the following chapter.

Learning, maintenance, construction: In a study of how learning biases determine the type of communication systems afforded by an iterated learning chain in a population of agents, Smith (2002) classifies the agents in three categories. *Learners* are agents that successfully learn a communication system; *maintainers* are agents that successfully maintain a communication system in an iterated learning chain; and *constructors* are agents that successfully evolve a communication system starting from random behaviour. These categories are not mutually exclusive, but form hierarchical sets: all constructors are maintainers and all maintainers are learners (but not the other way around).

Borrowing Smith's terminology, how can we classify our agents?[6] In Chapter 4, we demonstrated that they can be *learners*, as they can successfully learn to produce the correct response to a certain stimulus. In this chapter, we demonstrated that they can be *maintainers*, as they can form stable transmission chains that propagate a certain type of dancing behaviour from generation to generation, given long enough learning time.

However, "behaviour swapping" examples (as shown in Fig. 5.3e) indicate that the agents could also be *constructors*, able to "discover" dancing behaviour even without an expert agent to teach them. In the following chapter, we will set up

[6]Note that the specifics of Smith's experiments are very different from ours; we are only borrowing the agent classification terminology.

a series of experiments in order to investigate whether dancing behaviour can emerge (and be transmitted) in the absence of an initial expert agent.

Effect of weight saturation: One aspect of the learning algorithm that potentially plays an important role in the system's behaviour is the maximum weight cap, w_{max}, which, if reached by any of the synaptic connections in the neural network, stops all other connections from further increasing (Section 4.1.1, "Weight changes"). This does not directly influence the agents' learning success rates per se, as evidenced in Fig. 5.7 in agents A_2 & A_3, the learning success rates tend to reach 100% before any of the weights reach w_{max} and are stopped from further increasing (this happens after around 20 minutes for both agents).

What it can potentially influence, however, is how protected (or *resilient*) teachers are against "uncooperative" learners; a higher maximum weight ratio provides a larger "buffer" for weight ratio decreases before they actually influence an agent's behaviour. In this way, a higher value of w_{max} can lead to slower fail rates for transmission chains; the qualitative results we have discussed in this chapter, however, should remain the same. We will return to the discussion of the weight update mechanism and teacher resilience in Chapter 7 (Section 7.5).

Chapter 6
Behaviour emergence

In Chapters 4 and 5 we saw that simulated e-puck agents using the "tag & reward" learning algorithm can successfully *learn* the joint action L/R dancing task from an expert teacher; and that they can successfully *maintain* the expert's initial behaviour by passing it on from generation to generation, forming a transmission chain. Of the many open questions remaining from the experiments described in those chapters, we drew attention to two:

1. Can agents "discover" left or right dancing behaviour in the absence of an initial expert teacher?

2. Do agents tend to *regularise*, driving behaviour response rates to binary values of "success" (100% rates) or "failure" (50% rates)? Or do they tend to match their teachers' response rates?

In an attempt to give answers to these questions, in this chapter we will describe two experiments: one with the aim of finding out if behaviour emergence is possible (and if so, under what circumstances); the other with the aim of identifying trends in longer transmission chains.

6.1 Behaviour emergence

6.1.1 Setup

Any spontaneous "discovery" of one of the two possible dancing behaviours will happen in the context of an isolated pair of agents interacting; for this first behaviour emergence experiment, then, we will use two agents, both of them "naive" (not proficient at the dancing task), that interact for an extended period of time. The setup of the experiment is mostly identical to the isolated pair experiments we described in Chapter 4: two simulated e-puck agents (A_0 and A_1) are placed facing each other (but not in sensory range of each other) and left to interact for a period of time t_{learn}. Every $t_{\text{reset}} = 15s$ they are returned to those original positions (Fig. 4.9a); as before, successful dancing is rewarded with $DA_r = 1$ and crashing is punished with $DA_p = -0.5$.

The main difference in this experiment compared to the previous isolated pair experiments is that both agents are *learners*; there is no expert agent teaching a specific behaviour. Each learner is controlled by the simple neural network shown in Fig. 4.5; the weights of the network (initially set to $w_{12} = w_{13} = 2$, $w_{24} = w_{35} = 5$ and $w_{23} = w_{32} = -2$) are updated using the "tag & reward" learning algorithm detailed in section 4.1.

Another difference is the extended interaction time; we selected an initial value of $t_{\text{learn}} = 1$ hour to see how many (if any) pairs of agents would discover dancing in that time frame. We repeated the 1-hour interaction experiment for 60 isolated pairs of agents; for each of the agents, we recorded the rate of all three possible responses ("Right", "Left", "None") as well as the weight of each connection across the interaction time.

(Note that in this chapter, since there is no "correct" initial behaviour, we switch our terminology from "agent success rates" to "agent response rates".)

6.1.2 Examples of behaviour emergence

All of the 60 agent pairs ended up discovering dancing behaviour in the 1 hour of interaction; the time this took each pair ranged from quick (10 minutes or less) to slow (more than 40 minutes). The type of behaviour discovered was balanced: 27 (45%) of the pairs discovered "Right" dancing and 33 (55%) discovered "Left" (the two-tail P value of this outcome, given an equal probability for each direction, is 0.5190).

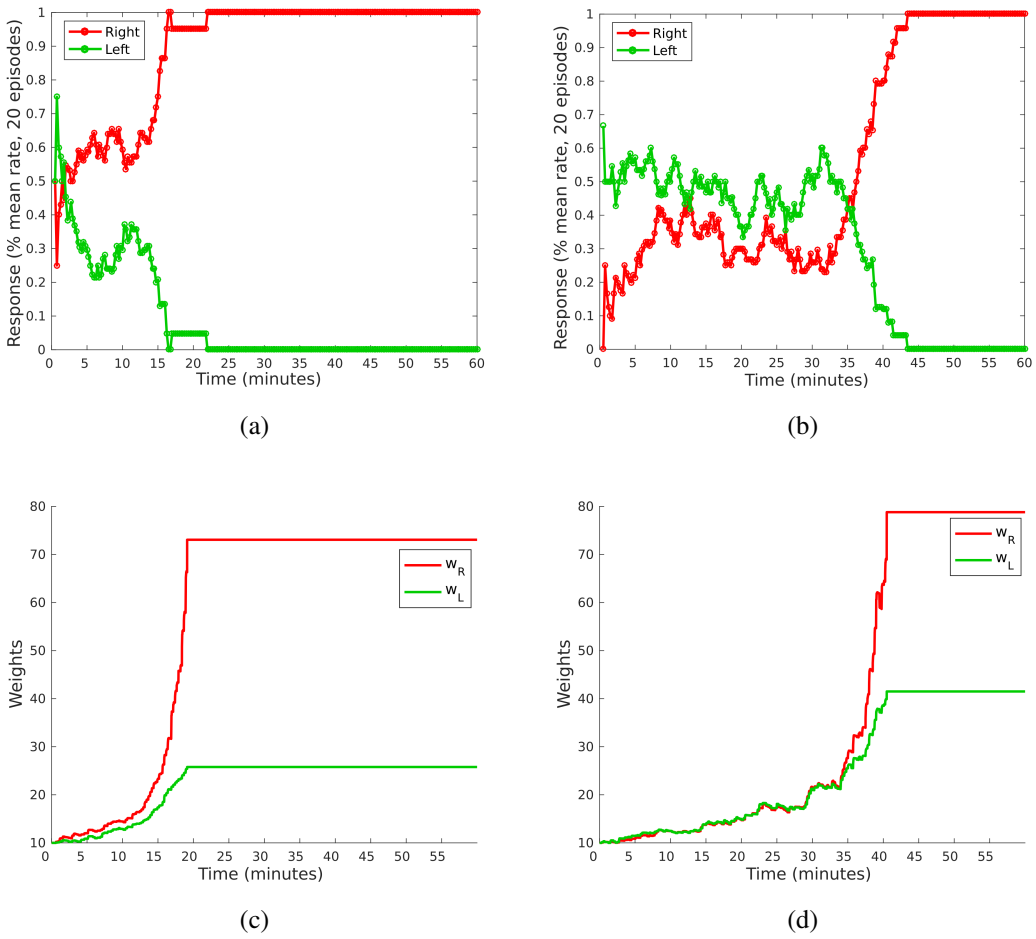


Figure 6.1: Two examples of non-expert agent pairs discovering dancing behaviour. The upper charts show the mean response rates of the first agent of each pair, calculated over 20 episodes of interaction: one pair discovers dancing around 15 minutes (a), the other pair around 40 minutes (b). The lower charts (c,d) show the cumulative weights of the two different connective pathways leading to "Right" (w_R) and "Left" (w_L) response. In both agent pairs, a initial period of "random walk" for all weights eventually leads to a chance event strengthening one of the weights more and to a learning feedback loop.

Fig. 6.1a and 6.1b show charts of the mean response rates of the first agent out of two different pairs that discover dancing (the response rates for the other agent of each pair follow identical patterns). The mean response rates are calculated over a rolling window of 20 episodes of interaction between the agents; only the "Right" and "Left" responses are shown, with "None" being omitted for clarity (null responses did not have any impact on the results). In the first pair of agents, behaviour emerges quite quickly (after around 15 minutes of interaction, Fig. 6.1a); the second pair of agents is an example of slower emergence (after around 40 minutes of interaction, Fig. 6.1b).

Fig. 6.1c and 6.1d show the changes in the cumulative weights of the two connective pathways leading to "Right" (w_R) and "Left" (w_L) responses. (Once more, the weights shown correspond to the first agent of each pair; the second agent's weights change in a similar way.) Both cases follow the same pattern; since agent responses are initially random, the weights of the pathways leading to either *R* or *L* response change in a "random walk" while keeping a weight ratio close to 1. When noise makes both agents turn in the same direction, all connections are strengthened; when they turn in opposite directions and crash, they are weakened. The weight increases are slightly more significant for the relevant connections that lead to the agents dancing, but without a steady response in either one or the other direction that would be provided by an expert agent, the total weights stay similar across all connections.

At some point, however, one of the weights happens to increase significantly more than its competitor. This could be due to repeated chance episodes of both agents turning in the same direction and dancing, or to an uncommonly quick dancing interaction between the two agents that leads to a quicker reward, changing the relevant weights more significantly. Regardless of the cause, this chance event creates a learning feedback loop that causes one behaviour to become more probable for both agents, both of which end up learning that behaviour even in the absence of a proper teacher.

6.1.3 Interaction time effect on emergence speed

Since the kind of behaviour emergence we described is caused by a chance event, it can happen at any point in time. We already saw two examples in which behaviour emerged at 15 and 40 minutes and mentioned that all of the 60 pairs in our experiment were dancing by the end of 1 hour of interaction. How long, however, can we usually expect a pair of agents to interact for before dancing behaviour emerges?

In Fig. 6.2, we provide an answer to this question by graphing the percentage of

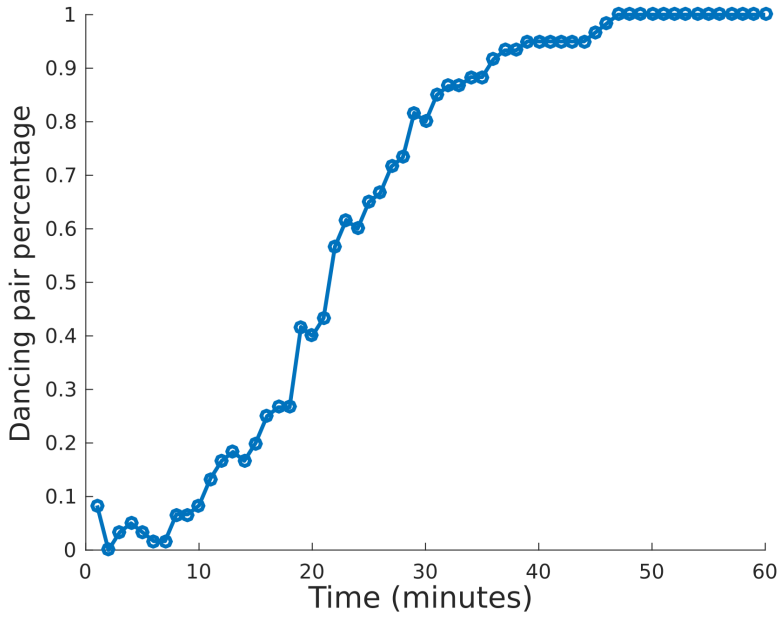


Figure 6.2: A graph of the percentage of 60 agent pairs that have discovered dancing at any given time point. (We count a pair as "dancing" if any response rate is higher than 90%). Both agents are non-expert. The first dancing behaviour emergence happens at $t = 8m$; after $t = 22m$ dancing has emerged in more than half of the pairs; after $t = 47m$ in all of them.

pairs that have discovered dancing at a given time point in the interaction. (We identify a pair as "dancing" when either the "Right" or "Left" mean response rates of the paired agents in the last 2 minutes are higher than 90%.) The first pair to discover dancing does so at $t = 8m$, while all pairs are dancing after $t = 47m$; between these two times, there seems to be a linear increase in the dancing behaviour emergence. After $t = 22m$ more than half of the 60 pairs have discovered dancing.

These results seem to suggest that, given enough attempts, dancing behaviour will eventually emerge even in agent pairs that interact for shorter times. Long transmission chains, like the ones we will examine in our next experiment, provide plenty of such attempts (albeit in a serial, not parallel way): even without a starting expert agent, then, long chains could lead to the emergence and maintenance of dancing behaviour. We will return to discuss this at the end of this chapter.

Before we go on to the next experiment, one last comment to make would be that since behaviour emergence depends on the weight difference of the two connective pathways w_R and w_L reaching a critical threshold, changing the values of dopamine

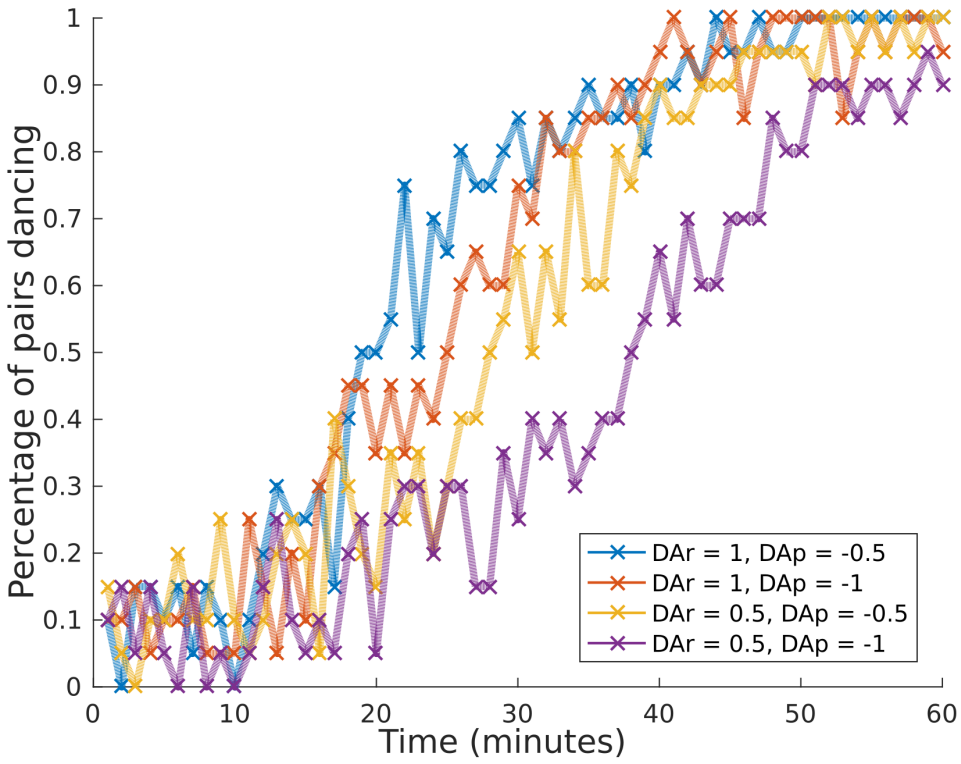


Figure 6.3: Plot of the percentage of agent pairs that have discovered dancing (identified by a maximum response rate of at least 80%) as a function of interaction time. Stronger punishment values lead to slower emergence of dancing.

rewards and punishments (DA_r, DA_p) allows us to manipulate how quickly a given pair of agents tends to discover dancing behaviour. Indeed, a test using 20 pairs of agents in each of four different reward conditions (DA_r, DA_p pairs of $1, -0.5$; $1, -1$; $0.5, -0.5$; $0.5, -1$) showed a trend of slower emergence for lower reward and higher punishment values. In all four conditions, however, all pairs of agents discovered dancing behaviour by the end of a 1-hour interaction.

Fig. 6.3 shows a plot of the percentage of agent pairs that have discovered dancing (identified by a maximum response rate of at least 80%) as a function of interaction time. In the condition where punishment is stronger than reward ($DA_R = 0.5$, $DA_P = -1$), dancing behaviour takes longer to emerge; still, even in that condition, almost all of the pairs have discovered dancing after 1 hour of interaction.

6.2 Long chains without experts

For the second part of this chapter, we will combine emergent behaviour, learning and transmission chains (or "construction", "learning" and "maintenance" in the terminology used by Smith, 2002) in a single experiment, in order to find out if stable *emergent behaviour chains* are possible in a transmission chain without experts. As we mentioned at the end of Chapter 5, we will also use longer chains in order to investigate potential probability matching or regularising biases that the agents might have.

6.2.1 Setup

The setup of this experiment is very similar to the chain experiments that we described in section 5.1: two agents A_0 and A_1 (both simulated e-pucks) are initially placed opposite each other but outside proximity sensor range and left to interact for t_{learn}. Every $t_{\text{reset}} = 15s$ the agents are reset to their starting position. After the interaction time t_{learn} has passed, agent A_0 is replaced with a new agent A_2, and agents A_1 and A_2 are left to interact for a further t_{learn}. This procedure is repeated for a number of generations N (Fig. 5.2).

The difference with the previous chain experiment is that agent A_0 is not an "expert", preset agent; all agents are learners with neural networks whose weights are updated using the "tag & reward" learning algorithm. The dopamine values used for reward and punishment are, once more, $DA_r = 1$ and $DA_p = -0.5$; the neural network used is the one shown in Fig. 4.5 with initial weights of $w_{12} = w_{13} = 2$, $w_{24} = w_{35} = 5$ and $w_{23} = w_{32} = -2$. Finally, a further difference is that for this experiment, the transmission process is repeated for a significantly higher number of generations ($N = 50$ instead of 10).

We ran 8 such transmission chain experiments, 4 using $t_{\text{learn}} = 12m$ and 4 using $t_{\text{learn}} = 15m$. The reason we picked these learning time values is that lower values (1 to 10 minutes) lead to "chaotic" chains that both break down very often (Fig. 5.4) and have a lower chance for each agent pair to discover dancing (Fig. 6.2). On the other hand, higher values (16 to 20 minutes) lead to very stable chains that quickly discover dancing and maintain success rates of 100% continuously; neither of those cases is helpful in trying to understand our agents' behaviour.

6.2.2 Examples of emergent chains

All 8 experiments we ran resulted in emergent chains very early (the quickest starting in generation 2, the slowest in generation 5). Fig. 6.4 shows two examples of emergent chains. In both cases, the top chart shows the mean response rates in the last 2 minutes of each agent's learning role; the bottom chart shows the w_L/w_R weight ratio (again, at the end of each agent's learning role).

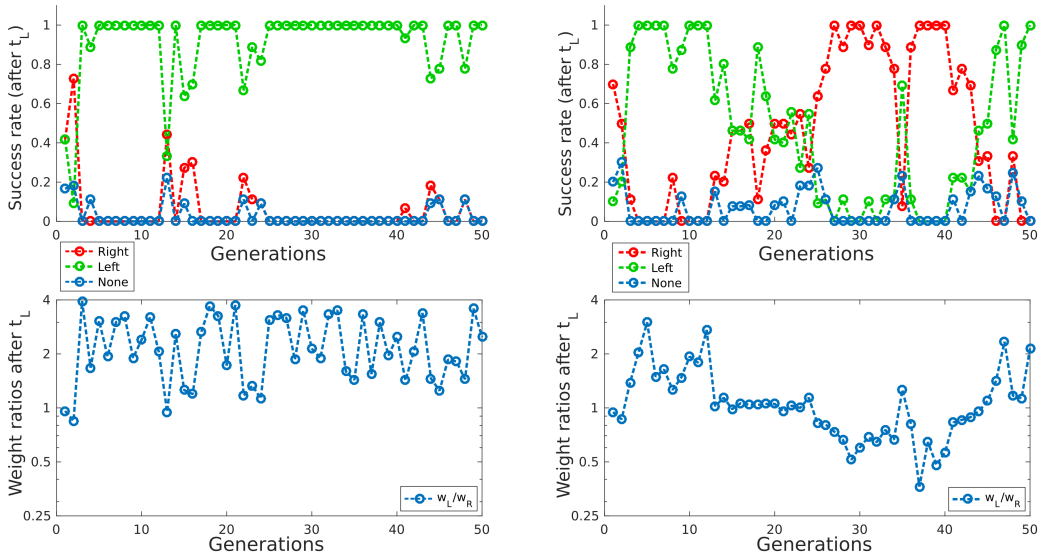
(a) Fig. 6.4a ($t_{\text{learn}} = 15m$) is an example of agents that discover "left" dancing very quickly (from the 3rd generation) and maintain a stable chain until the end of the experiment (50th generation). The success rates ("left" response rates) are 100% across almost all generations; occasional drops from this value are always followed by recoveries back to 100% rates. There is a lot more variation in the weight ratios, which seem to "jump around" (although their values consistently remain above 1, with weight ratios lower than 1.5 coinciding with the success rate drops).

(b) In the chain shown in Fig. 6.4b ($t_{\text{learn}} = 12m$), the agents again very quickly (generation 3) discover "left" dancing. The chain this time is a lot less stable; while success rates are very high (100%, except for a short drop) the chain breaks down in generation 12. This is followed by a period of non-learning agents with random responses and weight ratios very close to 1. Around generation 25, the agents "rediscover" dancing behaviour, this time "right" (in an example of behaviour swapping). This is maintained for around 20 generations and breaks down once more after generation 45. Interestingly, agents seem to match their teachers' weight ratios more closely than in the previous $t_{\text{learn}} = 15m$ example.

6.2.3 Trends in long chains

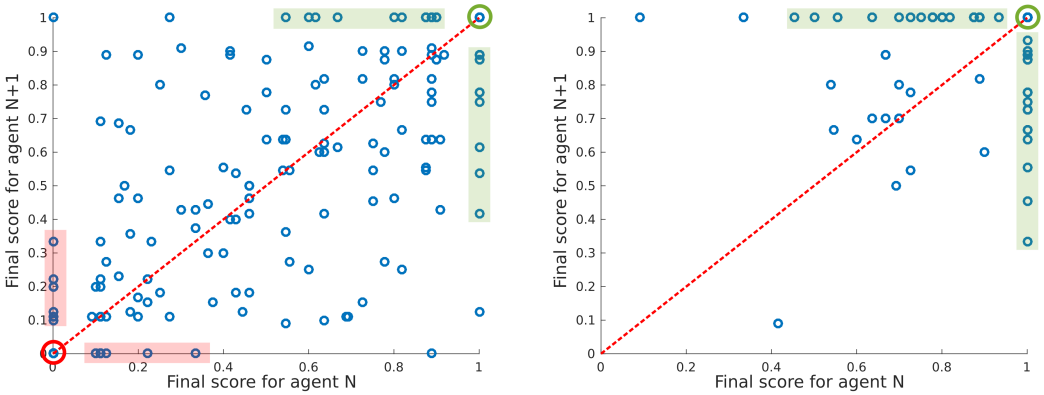
So far, we covered the first question we posed in the beginning of the chapter by showing that agents can indeed discover (and maintain) dancing behaviour, even in the absence of an expert agent initialising the transmission chain. Here we will attempt to answer the second question: do agents tend to be *regularisers* or to *probability match* their teachers?

The previous section's examples (Fig. 6.4) show a tendency of regularisation in long chains: the agents maintain the transmission chains with response rates that are mostly 100%; any drops in response rate are either reversed in recoveries, or lead to



(a) (b)

Figure 6.4: Two examples of emergent chains. In both cases, the initial agents for the chains are both learners; the pairs of learners quickly (gen. 3) discover dancing behaviour ("left" in both examples). (a): For $t_{\text{learn}} = 15m$, the emergent behaviour is maintained in a stable chain until the end of the experiment. Occasional drops from 100% response rates are always followed by recoveries. (b): For $t_{\text{learn}} = 12m$, the chain is less stable, breaking down in gen. 12. After some generations of non-learning agents, "right" behaviour emerges, persisting for 20 generations then breaking down.



(a) $t_{\text{learn}} = 12m$ (b) $t_{\text{learn}} = 15m$

Figure 6.5: Response rate inheritance dynamics plots for 50 generation long chains with no expert agents and learning times of 12m (a) and 15m (b). In both cases there are more points in the top right (green circle, "left" stable chains) and bottom left (red circle, "right" stable chains) compared to "probability matching" points on the $x = y$ line. Most drops in response rate (rectangles below the $x = y$ line) are followed by recoveries (rectangles above the line). In the 15m graph regularising behaviour is more pronounced; there is no "right" attractor, as "left" behaviour happened to emerge in all of the 4 chains we ran.

the chain breaking down. This trend is made clearer by looking at the response rates inheritance dynamics plots (Fig. 6.5) that we introduced in Chapter 5. Each point in the plot represents an episode of transmission; the x-coordinates are the response rates for the "teacher" generation N, while the y-coordinates are the response rates for the "learner" generation $N+1$. Fig. 6.5a includes the data from the chains with $t_{\text{learn}} = 12m$ and Fig. 6.5b the equivalent data for $t_{\text{learn}} = 15m$.

In both cases, the attractor points in the top right and bottom left corners, indicated by a green and red circle respectively, represent 100% response rate, stable "left" and "right" chains. The red line plotted through the points is the $x = y$ *identity* line, indicating *probability matching* agents. The red and green rectangles below the identity line highlight drops of response rates from 100% in stable chains; the rectangles *above* the identity line are equivalent recoveries to 100% response rates.

For learning times of 15 minutes (Fig. 6.5b), the chains are relatively stable. Most of the transmission episodes are between 100% response rate agents. All drops in response rate are followed by recoveries; other inheritance patterns are very infrequent. This is confirmed by looking at the number of inheritance points in each category (Fig.

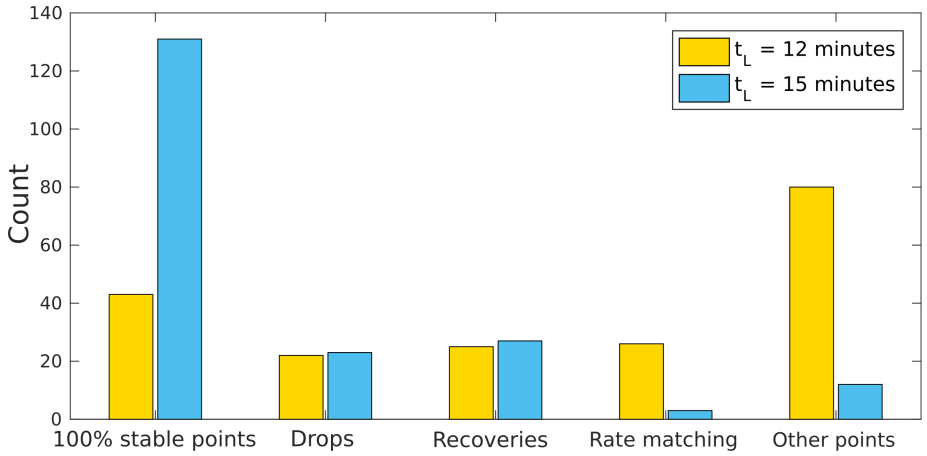


Figure 6.6: Frequency comparison for different possible inheritance behaviours (as identified in Fig. 6.5). *100% stable points* are points in the two extreme corners of the inheritance dynamics plot; *drops* from 100% are points shown in rectangles below the $x = y$ line; *recoveries* to 100% are points shown in rectangles above the $x = y$ line; *probability matching* points are points on the $x = y$ line except for (0,0) and (1,1). In both the 12m and 15m cases, the stable points are more frequent than rate matching and recoveries are more frequent than drops. (Recoveries include behaviour emergence.)

6.6, 15m): most of the points are stable and there are as many recoveries as drops. (In fact, the recovery number is slightly higher as it includes the initial discovery of dancing behaviour.) Note that there is no "right" attractor in Fig. 6.5b; this is because only "left" behaviour happened to emerge in the 4 chains we ran for $t_{\text{learn}} = 15m$.

The equivalent response rate inheritance data for the 12 minute chains (Fig. 6.5a) is harder to interpret visually. By comparing the frequency of inheritance points however (Fig. 6.6, 12m), we can see that while the "stable chain" points are definitely fewer compared to the 15m chains, they are still more numerous than the "probability matching" points that fall close to the $x = y$ identity line. Once more, the number of drops in response rate is matched by the recoveries. (Again, the number of recoveries is slightly higher, accounting for the behaviour emergence.)

6.3 Discussion

In this chapter we demonstrated through two experiments that dancing behaviour can emerge from non-expert agents interacting for long enough; and that this emerging behaviour can be maintained in transmission chains. In the next, final chapter we will take a step back and draw some conclusions from all the experiments we detailed in this thesis; before that, however, let us make a few short comments based on the findings of this chapter.

Emergent chains: With the parameters that we have been using, emergent behaviour chains are inevitable. Even when treating both agents in each pair in a transmission chain as new learners (so in effect, treating each generation as independent from the history of the chain, with a chance to discover dancing given by Fig. 6.2), an interaction time of 5 minutes is enough for an 80% probability of dancing emerging by generation 30. An interaction of 12 or 15 minutes, as in the chains we described in this chapter, almost always leads to the discovery of dancing by generation 20 (Fig. 6.7).

Of course, the generations are not independent as the "teacher" agent in each pair is influenced by its interaction as a "learner" in the previous generation. Any such influence that does not lead to response rates higher than 90% is not visible in the data shown in Fig. 6.2, and will lead to higher probabilities of emergent chains which in turn translates to quicker behaviour emergence; Fig. 6.7 only gives a lower bound.

Emergence speed: Even accounting for the "hidden" $< 90\%$ response rates does not seem to be enough to explain the speed at which behaviour actually emerges in the chains generated by our experiments: in most of the chains, emergence chains appear at (or before) generation 3. As we pointed out in Chapter 6, however, there is one more "hidden" transmissible factor at play: weight ratios. In the same way that weight ratios, transmitted from generation to generation as a hidden trait, can lead to chains suddenly breaking down, they can also lead to quicker emergence of behaviour.

Probability matching vs. regularisation: All of the longer chains produced by the experiment we described in section 6.2 tended to either have high response rates or break down. Any drops in response rate (especially in the case of the $t_{\text{learn}} = 15m$ chains) were followed by recoveries to high rates, not subsequent matching of the lower rates. This, along with the inheritance dynamics data shown in Fig. 6.5 and 6.6 seems to once more indicate that agents in the chain experiments we presented are regularising behaviour, not probability matching their teachers. We will come back to this discussion in the next chapter.

Behaviour swapping: In Figure 6.4b we can see an example of *behaviour swapping*; the initial transmission chain of "left" turning agents eventually breaks down and, in its place, a chain of "right" turning agents emerges instead. This behaviour is a combination of two effects we previously touched upon: the fact that, for learning times of 12 minutes and less, most transmission chains break down (see Section 5.4 for a description of the mechanism leading to these breakdowns); and the inevitable emergence of dancing behaviour between learning, non-expert agents (see Section 6.1.2).

"Behaviour switching" appears when these two behaviours are combined and it so happens that the new behaviour that emerges is different from the initial one that broke down; there is nothing systematic leading the agents to switch their behaviour from "left" to "right" (or vice versa).

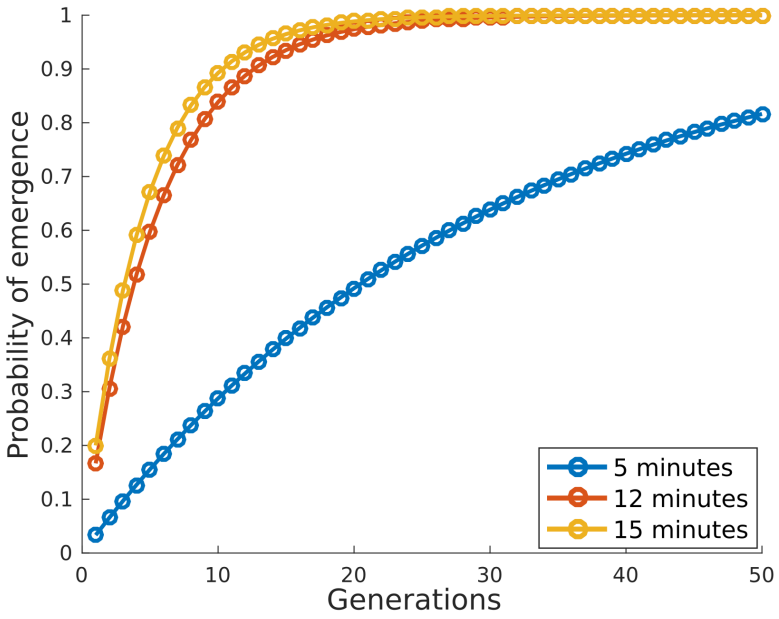


Figure 6.7: Probability of the emergence of dancing across generations for 3 different interaction times: 5, 12 and 15 minutes. The emergence rates for each interaction times are calculated from Fig. 6.2. Each pair of agents is treated as independent; any transmission of response rates lower than 90% is not visible in this data, so the probabilities shown in this graph are a lower bound of the actual probabilities of emergence.

Chapter 7
Discussion and future work

In the previous three chapters we detailed a series of experiments in which we showed that a population of simulated e-puck agents can *learn* the joint action "left/right dancing" task (Chapter 4); *maintain* an initial ("left" or "right") behaviour in a transmission chain (Chapter 5); and finally *construct* "left" or "right" dancing behaviour even in the absence of an initial expert (Chapter 6).

In this final chapter we will take a step back and discuss the results of all previous experiments as a whole, relating them to our initial motivation and goals. Through this discussion we will attempt to clearly delineate the contributions of the work described in this thesis both for autopoiesis as a theory of cognition and for the Iterated Learning research project. Finally, we will close with an analysis of the limitations of our work and proposals for future work that could be based on it, to either address those limitations or expand its scope.

7.1 Autopoiesis and Iterated Learning

Chapter 2, in addition to providing an introduction to autopoiesis as a non representational approach to cognition and to the Iterated Learning Model of language evolution ("ILM"), also contains what is essentially a *theoretical* contribution: the proposal of the iterated learning model as a *practical approach* to an autopoietic explanation of important structural features of language. In order to combine autopoiesis with iterated learning, our proposed system removes some components that most iterated learning experiments have in common.

The Iterated Learning components we remove are the following:

– Object concepts; these usually correspond to "meanings" in the ILM (Kirby et al., 2008).

– Lexicon, as a collection of meaning-symbol associations (Smith et al., 2003, p. 375).

– "Function independence" clause (Brighton and Kirby, 2005, p. 13).

Instead, we make the following additions:

+ Context, in the form of a specific joint-action task;

+ A biologically plausible (or at least, biologically informed) form of learning.

Some comments on this approach to Iterated Learning:

Novel contribution: Language (and the act of "languaging"; Maturana, 1978) is a concept of central importance in autopoiesis. To the best of our knowledge, however, all autopoietically inspired studies of language focus on theoretical explorations. These include philosophical extensions (in the "enactive" tradition) of Maturana's idea of languaging (Cuffari et al., 2015); a re-framing of concepts from linguistics (Bottineau, 2008); and a critique of "mainstream linguistics" (Kravchenko, 2011). (For a more comprehensive list, see Cuffari et al., 2015, pp. 1090-1091.)

Our approach is novel in that it aims for a bottom-up, practical investigation in the "understanding by building" tradition (Pfeifer et al., 2005); in addition, we use a biologically plausible model of learning, adding a layer of ecological validity to this investigation. Di Paolo's simulation studies of communication (Di Paolo, 1997) and social coordination (Di Paolo, 2000) are also examples of practical approaches, but their focus is not on a linguistic system.

Representations: Although our approach seems to firmly place itself on the non-representational side of the "representation debate" we mentioned in Chapter 1, it actually provides a constructive viewpoint as it does not discount the use of representations in human cognition. The only strong claim that we are making is that representations are grounded in language instead of the other way around.

Moreover, in the combination of autopoiesis and the ILM, we are providing a potential pathway towards language, and in consequence also towards a grounded

theory of representations. Of course, the question of grounding object concepts and representations in language is an entirely different research topic that we have not touched at all in our work (for an example, see Steels, 2012).

Dependence on autopoiesis: While the work in this thesis is motivated by the theory of autopoiesis, the fact remains that there is nothing inherently "autopoietic" about the system we described, apart from its design being informed by principles stemming from our understanding of autopoiesis. The theory of autopoiesis, for the purposes of this thesis, could be replaced with a different non-representational approach to cognition without substracting from the value of the results that we have described.

In fact, even if someone completely rejects the arguments of non-representational theories of cognition in general, an autopoietic approach can be of *methodological* use to the Iterated Learning model in two (related) ways. First, adopting an autopoietic view forces us to be more rigorous: none of the assumptions that we make (such as, for example, any assumptions about the *function* of language) exist in a vacuum: they influence the rest of the system and must thus be made explicit. At the same time, by getting rid of as many assumptions as possible and seeing if the conclusions of Iterated Learning models still hold, we can examine the boundary conditions of any results and make the Iterated Learning methodology more robust.

That said, we still believe that the autopoietic approach is an exciting research direction for cognitive science; and since our results stem directly from the adoption of an "autopoietic" perspective, they also provide some support to the methodological value of autopoiesis as a theory of cognition.

First step: There is a long way to go from the model we propose in this thesis to an Iterated Learning model of language, as the cultural transmission chains we describe only transmit *behaviour*, not an evolvable communication system. We will come back to this at the end of this chapter (section 7.7.1).

7.2 Experimental framework

The experimental framework we detailed in Chapter 3 is a *methodological* contribution: the design and implementation of a system that allows us to run joint action transmission chain experiments using simulated robots. The following components are novel contributions:

1. The description of a simulated e-puck agent and a basic environment (Fig. 3.3) in the form of a custom world file for the *Stage* simulator (Vaughan, 2008).

2. The implementation (as described in section 3.4.2) of a neural network controller and a biologically plausible form of reinforcement learning (for learning agents), as well as a rules-based controller (for expert agents).

3. The description of a joint action task (more specifically, the "left/right dancing task") in the form of a *reward server*, an external "observer" that checks the state of the simulation and rewards (or punishes) the agents taking part in the task.

4. Finally, a *script controller* that can remove or add agents to the simulation as well as move them around in space, thus controlling both learning episodes and transmission chain experiments.

One advantage of this system is the modularity it provides. Since, as we saw in Chapter 3, all the system's components communicate through a common interface (ROS, Quigley et al., 2009), any of the components can be changed as long as the new component communicates using the same interface. In this way, the simulated world or agents can be easily changed (possibly for a more detailed simulator); the task can be swapped for a completely different one; or new types of experiments can be set up; all without having to change the rest of the system's components.

An example of the system's modularity comes from a contribution we did not mention so far, as we ended up not using it for any of the experiments: a custom ROS controller for physical e-puck robots, completely interchangeable with the simulated e-pucks in *Stage*. This allows us to substitute the simulated e-pucks for physical robots without the need of any further adjustments. We will return to this topic in section 7.7.4, discussing ideas for the use of physical e-puck robots in transmission chain experiments.

Of course, the most substantive contributions in this thesis are the experiments and results described in Chapters 4, 5 and 6. We will discuss these in the next sections.

7.3 Cultural transmission chains "in the wild"

At the end of Chapter 2 (section 2.3), we made the following claim:

> It is possible to build a system that is both,
>
> 1. Consistent with the non-representational, autopoietic design principles that we outlined in section 2.1.2.2 and,
> 2. Able to exhibit a chain of cultural transmission of behaviour.

The results we presented in Chapter 5 confirm this initial claim, as they establish that a system of agents that keeps clear of representations or object concepts as foundational "building blocks" can indeed generate cultural transmission chains.

Comparing our results to the chimpanzee transmission chain experiment that inspired our claim in the first place (Horner et al., 2006), we find a number of similarities:

Task: Both the "chimpanzee chains" and our "robot chains" involve a certain task (food retrieval and dancing, respectively) that can be solved in two different ways (*sliding* or *lifting* a door for the chimpanzees; dancing *right* or *left* for the robots).

Experts: In both experiments, a transmission chain starts with an "expert" agent, taught by the experimenters to successfully complete the task consistently in one of the two possible ways.

Learning & chains: After the initial agent is trained, a second agent learns how to perform the task from this expert; a third agent learns from the second one, a fourth from the third, *et cetera*. This process, repeated for a number of generations, creates a transmission chain.

We must note, however, that the types of learning are different: the chimpanzees learn by watching the previous agent in the chain successfully solve the food retrieval task. This is a form of *observational* learning: there is no direct interaction between the agents. In contrast, the robots learn by performing the dancing task with the previous agent in the chain and getting reinforcement in the form of dopamine "injections" (a form of *operant conditioning*).

Both types of learning are found in nature (Hoppitt and Laland, 2008); we are, however, more interested in joint action as, contrary to observational learning,

it provides a "platform" on which to build a communication system (see section 7.7.1).

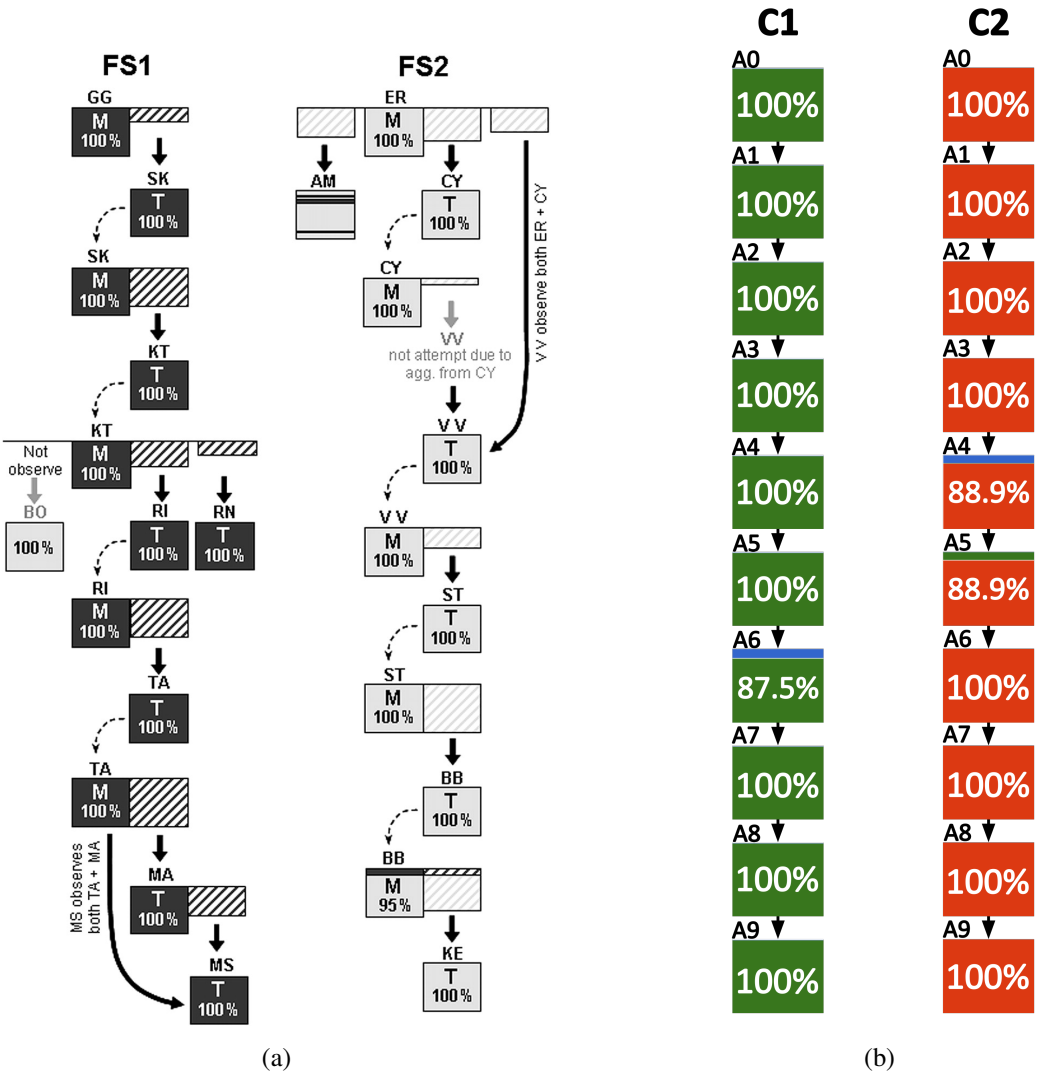
Distinct role periods: Both the robot agents and the chimpanzees go through at least two distinct phases: a phase where they learn a certain behaviour from the previous agent in the chain, and a phase where they "perform" for the next agent in the chain to learn ("model" phase in Horner et al., 2006). This happens either directly (in the case of the robots interacting) or indirectly (in the case of chimpanzees observing).

A difference here is that the chimpanzees go through an intermediate phase as well ("test" phase), in which they are tested for consistent behaviour: only observers who managed to successfully complete the task 10 times are allowed to become "teachers" for the next agent (Horner et al., 2006, p. 13879). In the case of our robots, there is no testing phase: after a set time t_{learn} each learner adopts a teaching role, regardless of its success rate at the dancing task.

Behaviour persistence: Finally, the results of both experiments are very similar in that the original behaviour taught to the expert agent persists in a cultural transmission chain. To illustrate this, we re-formatted the success rates of one of the stable $t_{\text{learn}} = 15m$ chains that resulted from the experiment we described in section 5.1 (Fig. 7.1). Of course, depending on the learning time values chosen, some of the robot chains fail; and even in stable chains, some of the agents have success rates significantly lower than 100%.

We have established that cultural transmission chains, an element crucial to all Iterated Learning approaches, can be built without any assumptions that would be incompatible with autopoietic theory. This is a significant result, as it means that we have accomplished the primary goal of this thesis: a first step towards an autopoietic account of language evolution through Iterated Learning.

Furthermore, despite the simplicity of all of the components that we made use of (agent bodies with one type of sensor and actuator; very basic neural networks; minimal, non-communicative left/right joint action task), our system exhibits complex behaviour that has interesting parallels with other Iterated Learning research; we will discuss some of these parallels in the next sections.



(a) (b)

Figure 7.1: Behaviour persistence in chimpanzee and robot transmission chains.
(a): Two chimpanzee transmission chains, starting with two different initial behaviours (Horner et al., 2006). Dark shaded areas represent “lifting” behaviour; light shaded areas are “sliding” behaviour. The initial behaviour in each chain persists for 6 generations. Each of the chimpanzees go through an observation phase (bold arrows), a testing phase (“T” boxes) and a model phase (“M” boxes).
(b): Two robot transmission chains, again starting with two different behaviours. Red areas represent “right” dancing behaviour, green areas represent “left” and blue areas show null responses. Again, the initial behaviour in each chain persists for 10 generations.

7.4 Learning, maintenance, construction

In Chapters 4 and 5, we saw that our e-puck robot agents can be classified as learners (since they are able to learn the correct dancing behaviour using the "tag & reward" learning algorithm) and maintainers (since they are able to maintain consistent behaviour through generational transmission).[1] These two agent attributes, *learner* and *maintainer*, are sufficient to support cultural transmission chains. As we saw in Chapter 6, however, they can also be classified as *constructors*, since they can discover consistent dancing behaviour from initially random responses.[2] The fact that our agents are constructors is significant for a number of reasons; we will comment on three of them that are especially relevant.

Emergent social behaviour: In section 3.1.2, we discussed an experiment by Quinn (2001) in which a genetic algorithm transforms a formerly non-communicative behaviour into a communicative signal, evolving a communication channel in a system that initially had none. Quinn (2001, pp. 358-359) stresses the importance of starting from a non-communicative system if we want to provide a convincing explanation of how communication *evolved* in the first place.

There is a certain similarity between Quinn's experiment and the emergence of dancing in our system. What initially starts as a behaviour that has some non-social function ("turning around" an object) is transformed into something different ("dancing") when placed in a social context. There is no need to explicitly include social behaviour in our system; we only need to include a "basic" behaviour, more easily explainable in evolutionary terms, and "dancing" as a social behaviour emerges naturally from that.

Fewer assumptions: The fact that two learning agents can "discover" dancing by interacting allows us to discard one of the original assumptions: that somehow, there is an initial expert dancing agent initialising each transmission chain. By removing this assumption we are giving a more systemic account: a non-social behaviour ("turning around"), placed in a social context, leads to the emergence of a social behaviour ("dancing"). This social behaviour is then maintained in a

[1] Once more, we are borrowing this terminology —and only the terminology— from Smith (2002).

[2] We should note here that this is one point where our definition of "constructors" differs from Smith's: in his experiment, construction happens through the Iterated Learning chain process; in our experiments in Chapter 6, the emergence of dancing behaviour happens between two isolated agents, without the need of a transmission chain.

	Stable chains		No stable chains
	Emergence	**No emergence**	
Breakdowns	Type A *(10m, 12m)*	Type C *(?)*	Type E *(1m, 3m, 5m)*
No breakdowns	Type B *(15m, 20m)*	Type D *(?)*	

Table 7.1: Possible types of societies depending on the presence of emergence, stable chains and chain breakdowns. Values in brackets are learning times that lead to societies of a respective type. Some examples are given in Fig. 7.2.

cultural transmission chain from generation to generation of learners and teachers.

Possible "society types": There are three different possible scenarios in the cultural transmission chains we saw in our experiments: behaviour can emerge from previously random-response agents; it can be maintained in stable chains; finally, these stable chains can break back down into random-response agents. If we treat each of these as a possible *aspect* of a society ("emergence", "breakdowns" and "stable chains") we can distinguish $2^3 = 8$ different social "profiles" (Table 7.1).

- 4 out of these 8 society types have no stable chains (for example, populations with t_{learn} equal to 1, 3 or 5 minutes). These are degenerate societies full of random-response agents; there are no learners, maintainers or constructors. We will group all of them together and classify them as *Type E*. An example of a Type E society, made up of random agents without any stable chains can be seen in Fig. 7.2c.

- *Type C* and *Type D* societies have stable chains but no emergent behaviour; agents belonging to these societies are learners and maintainers but not constructors. When initialised with a certain behaviour, these society types either maintain the behaviour indefinitely with no breakdowns (Type C) or break down at some point and degenerate into Type E societies. If there is no initial expert, since there is no emergence they are again indistinguishable from Type E societies.

  We have no examples of such societies in our experiments, as any interaction time that is long enough for learning (which is a prerequisite for stable chains) is also long enough for behaviour to eventually emerge. Significantly altering

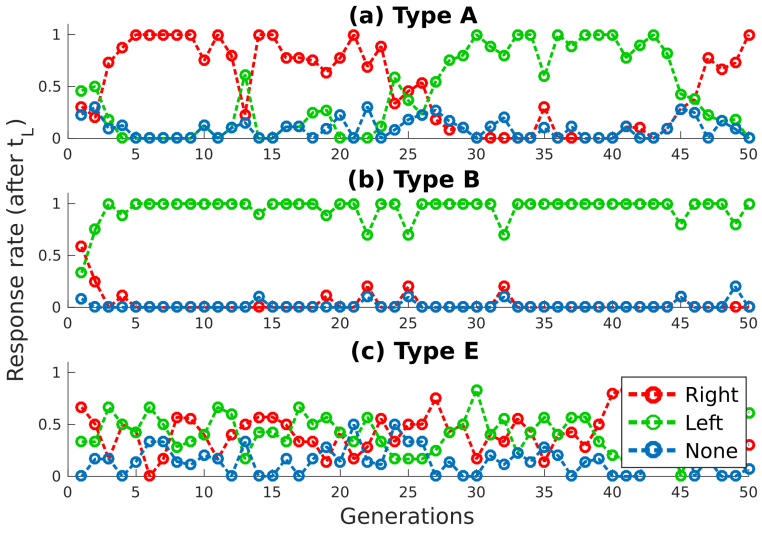


Figure 7.2: Three examples of different society types. (a): Type A societies include behaviour emergence, stable chains and breakdowns; different types of behaviour emerge, are maintained in chains and break down. (b): In Type B societies, behaviour (in this case, "left" dancing) emerges and is maintained indefinitely in stable chains. (c): Type E societies consist of random-response agents; there are no successful chains.

some aspects of our learning algorithm (for example, dopamine DA_r and DA_p values; see section 6.1.3) could lead to such chains; however, without further work we cannot say whether Type C or Type D societies are eventually possible.

- *Type B* societies are societies in which one type of behaviour eventually emerges and, since there are no breakdowns, is maintained indefinitely. Agents are learners, (perfect) maintainers and constructors. Longer learning times (15 and 20 minutes) lead to such societies; an example for $t_{\text{learn}} = 15m$ is seen in Fig. 7.2b.

- Finally, the most interesting societies are *Type A* societies, in which either "right" or "left" behaviour emerges and is maintained for some time in a transmission chain. Agents in Type A societies are learners, constructors and *imperfect* maintainers. At some point the chain breaks into random-response agents, but eventually behaviour (not necessarily of the same type as before) emerges again and the cycle restarts. Type A chains societies appear for learning times between 10 and 14 minutes; an example for $t_{\text{learn}} = 12m$ is given in Fig. 7.2a.

7.5 Teacher resilience

An interesting point that we raised in Chapter 5 was the presence of a "hidden" trait of agents that is not always visible in their response rates as learners but potentially influences their "teaching" ability. We used the weight ratio w_L/w_R to measure this trait (w_L being the combined weights of the connective pathway leading to "left" responses and w_R the equivalent measure for "right" responses). The weight ratio directly influences response rates, but after these reach ceiling values (100% right or left responses) any further changes of the ratio in the same direction have no further effect during the agent's *learning* phase.

During its *teaching* phase, however, the agent is paired with a possibly "uncooperative" learner, leading to crashes, punishment and lowering of weights for both learner and teacher. If the weight ratio value is close enough to the threshold, any further weight decreases can lead to visible changes in the teacher's response rates. The weight ratio, then, is a measure of how easily influenced an agent is by an uncooperative dancing partner. We can call the equivalent trait an agent's *resilience*.

Breakdown prediction: Low or decreasing teacher resilience can be a predictor of a chain that is about to fail. Looking at success rates only, a chain can appear perfectly stable; if teacher resilience is low, however, this stability is fragile and can be easily influenced by an "uncooperative" learner. (We gave a detailed account of a case where this happens in section 5.4.)

Determinants for resilience: There are two main factors that determine an agent's resilience; the first being the *time spent dancing* while the second being the *maximum value* that an agent's weight ratio is allowed to reach.

1. The more time a learning agent spends dancing with their teacher, the higher the weights of the "correct" neural pathway; this leads to a high weight ratio and high resilience. Longer t_{learn} times translate to more time spent dancing; t_{reset}, however, also has an effect: resetting the agents' positions too often leads to more time spent trying to find each other and less time spent dancing.
2. After any of the weights reach the w_{max} value, all weights stop increasing (see section 4.1.1). This maximum weight limit also sets a limit for an

agent's weight ratio[3]; higher values of w_{max}, then, would also lead to higher agent resilience.

Theoretically, changing the value of w_{max} would alter the stability of transmission chains for given learning time ranges. Removing w_{max} altogether could lead to a very low range of learning times for which the system exhibits interesting behaviour such as the "Type A" societies we described in the previous section.

In practice, it remains to be seen how much of an effect it would have; this could be a potential direction for future work, although in section 7.7 we will make a case against the relevancy of parameter exploration for our thesis.

Behavioural measure of resilience: We have been using the weight ratio w_L/w_R as an internal, "biological" measure of agent resilience. This, as well as the use of any equivalent internal measures, is only possible because of the simplicity of the agent's neural network (which conveniently has very distinct "left response" and "right response" pathways). More complex neural networks (which are needed for any widening of this work's scope, as we discuss in section 7.7.3) will make it difficult (but not impossible; see Beer, 2000), to cleanly associate network measures with behavioural traits.

A behavioural, rather than internal, measure that could be used to determine an agent's resilience comes from reversing the context the agent is in by pairing it with an expert agent dancing in the opposite direction. In this scenario, the non-expert agent's behaviour will eventually swap from "left" to "right" or vice versa; how long this takes is a direct measure of the agent's resilience.

Masked traits: In a discussion about biological and linguistic co-evolution, Deacon (2003) introduces the notion of evolutionary *masking* of genetic change. A masked genetic factor (this can be a gene, an allele, or even a trait) is hidden, or "shielded", from natural selection. Some obvious ways that this can happen

[3]This limit, however, is not straightforward to calculate as it also depends on the rate at which the weights are increasing. As an example, if w_L is increasing at a high rate, the value of w_R when w_L reaches w_{max} will be low and the maximum weight ratio w_L/w_R will be high; a lower weight increase rate would lead to a lower maximum weight ratio. In turn, the rate of weight increase depends on how quickly the agents start dancing after they make a "decision" about which way to turn; this is stochastic, as it can change depending on the orientation of the agents when they happen to wander in sensory range of each other, which is random.

include unexpressed genetic change and recessive alleles whose effect is suppressed by a dominant allele.

However, even if a genetic change is actually expressed (for example, as the degradation of one of the organism's former functions), it can be hidden from natural selection in a low-competition environment or niche where this loss of function does not lead to vastly different fitness (Deacon, 2003, p. 10). In the same way, environmental changes can also lead to the *unmasking* of genetic changes that were neutral before, but now somehow interact with the agent's fitness (Deacon, 2003, p. 12).

Of course, there can be no direct equivalent to this process in our system, as there is no selective pressure on any of our agents' traits. Nevertheless, there are definite similarities between traits masked from natural selection and teacher resilience as a "hidden" trait. As long as resilience (measured as the w_L/w_R weight ratio) is above a certain threshold, it is a *masked* trait: any changes or degradation in resilience do not affect the agent's behaviour (response rate) and are thus not "expressed".

Significant enough degradation, however, combined with environmental factors (an "uncooperative" dancing partner) lead to resilience being *unmasked* as a trait: it now affects behaviour, causing the agent's response rates to drop; since this propagates to the next agents, it potentially leads to a chain breakdown.

7.6 Regularisation

One of the questions we tried to answer in both Chapter 5 and Chapter 6 was whether the agents in a transmission chain tend to *match* the response rates of their teachers or *regularise* to high rates of either "right" or "left" behaviour. In order to explain why this is relevant, we need to look at an important property of language: *predictability*.

Most natural languages are regular; linguistic forms are of course variable, but they are predictable in their variability (Smith and Wonnacott, 2010). As can be seen during the formation of novel languages, even initially unpredictable languages go through a process of *regularisation* (Hudson Kam and Newport, 2005). In the same way, even when a language is transmitted through non-native speakers who introduce unpredictable variation ("probabilistic grammatical tendencies", see Hudson Kam and Newport, 2009), this variation disappears through the process of learning and the learned language is regularised.

On the other hand, experiments show that humans match the variability of languages they learn as adults (Hudson Kam and Newport, 2005, p. 153):[4] if, for example, two interchangeable linguistic forms A and B are encountered in a non-predictable way with probabilities equal to p_A and p_B respectively in a language, adult learners of that language tend to produce the forms with those same probabilities (Smith and Wonnacott, 2010, p.444). If that is the case, how is it that languages become regular?

One answer comes from the experiments of Hudson Kam and Newport (2005); these show that while adults tend to match the variation of languages they are taught, this is not the case for children, who typically regularise. Languages, then, become regular when "passing through" young learners. Another, complementary answer comes from the iterated learning experiments of Smith and Wonnacott (2010), which show that an initially *unpredictable* semi-artificial language, passed from generation to generation of learners in a transmission chain, keeps its variability but becomes fully *predictable*. The process of iterated learning, then, can amplify any regularising tendencies speakers might have and lead to predictable languages.

Returning to our system, are our robot agents probability matching (similar to adult language learners) or do they regularise (similar to children learners)? As we discussed in Chapters 5 and 6, there are several signs that they are biased towards *regularisation*:

[4]There is evidence that adults also tend to regularise when the stimuli they are exposed to are too complex to learn (Hudson Kam and Newport, 2005, p. 157). This points to adults having (weak) biases towards regularisation.

(i) Looking at chain examples for learning times of 12, 15 and 20 minutes, most stable chains seem to have 100% "right" or "left" response rates (Fig. 5.3a, c; Fig. 6.4a, b). Any drops in response rates do not persist; chains tend to either break down or go back up to 100% rates. This is more pronounced for 15 minute learning times, while 20 minute learning times never fall below 100% rates. Note, however, that there *are* some examples of chains that seem to be probability matching, as they maintain relatively stable sub-100% response rates over several generations (Fig. 5.3d).

(ii) Response rate inheritance dynamics plots for experiments with regular stable chains ($t_{\text{learn}} = 12m$ and $15m$) tend not to have many points proximal to the $x = y$ identity line that represents probability matching (Fig. 5.8, 6.5). This is confirmed by the relatively low R^2 values, indicating that the $x = y$ line fits poorly to the inheritance data (Table 5.1).

(iii) Finally, another indication that most agents regularise is the frequency of the different types of inheritance dynamics shown in Fig. 6.6: for both 12 minute and 15 minute learning times (although especially so for the latter), the number of agents matching their teacher's rate is relatively low.

However, there are also a lot of unanswered questions. These results, especially for 12 minute learning time chains, are not conclusive; the inheritance dynamics plots are not easy to interpret, and as we mentioned there are examples of agents that seem to be probability matching. More importantly, we do not know if agents learning in an isolated pair context would probability match or regularise; further testing is needed to determine that, as maybe the transmission of behaviour through chains is the cause of any regularisation effect. Finally, it is also plausible that "lifelong" learning also plays an important role, as it means that teacher response rates can *also* change in response to the learner.

7.7 Future work

The work we described in this thesis is only a first step towards the more ambitious goal of a full autopoietic account of language evolution; it also has numerous constraints and potentially eliminable assumptions. The upside to this is that it provides a plethora of open questions to be answered by future research. Broadly speaking, we could divide these open questions into two groups. The first group would be *parameter variation*: determining how various parameters of the system — weight freezing and w_{max}, learning times, dopamine reward values, "lifelong" learning and many others — influence the system behaviour. The second group would be *exploratory expansion*, trying to build on the system we have detailed instead of trying to better understand it.

The goal of this thesis, however, is not to fully explain or understand how the mechanism of transmission chains of joint behaviour works; rather, the goal is to provide a "proof of concept" that there *do exist* parameters that enable such transmission chains to appear in the system. Therefore, instead of focusing future work on an exploration of how varying system parameters would influence aspects of the system, we would suggest that it makes more sense, after having established that transmission chains are possible, to next try and expand the system's behaviour.

In this spirit, in the rest of this chapter we will detail four possible "exploratory" research directions: *communication games*, *populations of agents*, *complex networks* and *physical robots*.

7.7.1 Communication games

There are two components that are common in most of the Iterated Learning experiments: a cultural transmission chain and a communication system (usually, but not always, a collection of meaning to signal associations). This communication system is initially stochastic, which makes it difficult to learn; as it goes through the cultural transmission chain, a pressure for learnability leads the communication system to evolve and become more structured in some way. This increase in structure also makes the learners progressively better at whatever task they are using the language for. (Of course, as we saw in section 2.2.1.3, this is a simplistic view of iterated learning research; more recent research explores more factors that just this pressure for learnability.

So far, we have established that an autopoietic "version" of the cultural transmis-

sion chain is possible. The next step in an autopoietic account of Iterated Learning must then be the transmission of an evolvable communication system. This means that we need to make a transition from a simple joint action task ("L/R dancing") to a *communication game*. That said, adapting our task to one that makes use of signalling is not straightforward, as we need a signalling system that supports at least two signals if we want the system to be evolvable. A signalling system that is only comprised of one signal is binary in its success; it can either fail or fully succeed, and there is no potential for a middle ground and thus no potential for a gradual evolution.

We will give here an example of a minimal communication game that still supports an evolvable communication system: the *L/R/* dancing game*. The game is based on the L/R dancing task, with two additions: a new agent role ("follower", as opposed to the left or right "leader" agents); and a signalling system made of two signals, S_A and S_B.

L/R/* dancing game description

Leader role: The role of leaders is similar to the L/R dancing task; an ideal leader still always responds with "right" or "left" behaviour. In addition to this, however, they also produce a respective signal — for example, S_A when turning right and S_B when turning left.

Follower role: The role of followers is to appropriately react to the leader's signal; for example, by turning right for the S_A signal and left for S_B. In other words, followers are flexible and adapt to the leader's turning direction.

Assigning roles: As agents need to learn both leader and follower roles from their teachers, the roles need to be randomly assigned for each interaction attempt. This random assignment could also depend on an environmental factor, giving some degree of ecological plausibility.

Initial and ideal communication systems: An ideal communication system would be the one we described above: a one-to-one correspondence between the signal space (S_A, S_B) and the behavioural space ("turn right", "turn left"). The initial communication system, however, should be ambiguous or stochastic. (An example of ambiguous behaviour would be the production of both S_A and S_B by a leader turning "right"; a stochastic behaviour could involve the follower turning "right" 80% of the time and "left" 20% of the time when receiving an S_A signal.)

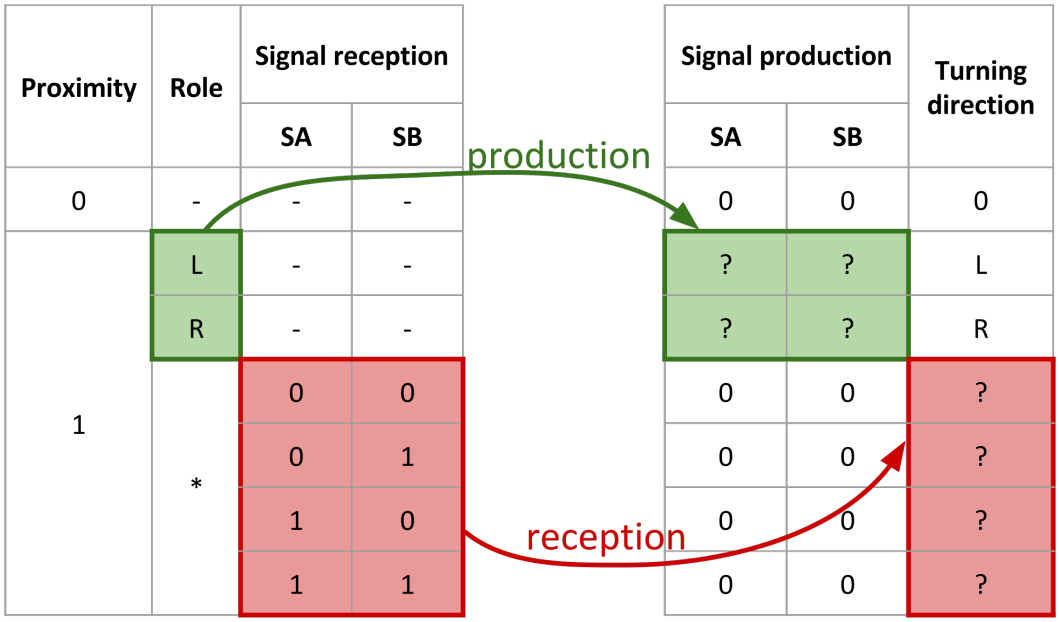


Figure 7.3: A truth table describing the L/R/* dancing game. Some values (indicated with question marks) are variable, as they depend on the production rules for leaders (shown in green) and reception rules for followers (shown in red).

This initial imperfection is required if we want the communication system to be potentially evolvable.

Neural network inputs and outputs: A neural network capable of learning all three possible roles ("R-leader", "L-leader", "*-follower") needs four input nodes (two nodes for S_A and S_B signal reception; one node for proximity detection; and one node that corresponds to the agent's current role) and four output nodes (two nodes for turning "right" and "left", two nodes for S_A and S_B signal production).[5] A truth table describing the L/R/* dancing game is shown in Fig. 7.3. (Note that the production and reception rules are encoded in the neural network.)

We should note at this point that, despite the fact that a communication system would seem like a straightforward addition to our system that would bring it closer to the Iterated Learning model and to the study of less controversially "linguistic" phenomena, such an addition is more challenging than it may appear. Despite the simplicity of the task that we used for the experiments in this thesis, we had to make a

[5] Of course, the network will need to be more complex than the minimal one we have been using so far; see section 7.7.3 for a relevant discussion.

number of concessions in the biological plausibility of our learning algorithm in order for the task to be learnable by out agents.

The leap from this task that could be learned by agents with a 1-input, 1-output network to one that involves signalling and requires a 4-input, 4-output task is very steep. In essence, what we are encountering here is a manifestation of the "scaling up" issue of non-representational cognitive science that we described in section 1.4: while explaining simple tasks in non-representational terms is often possible, the real challenge comes in trying to scale the task complexity.

7.7.2 Populations of agents

The transmission chains we described in our experiments are supposed to be "miniature societies" of agents, mimicking the transmission of behaviour from generation to generation in cycles of learning and teaching. However, in real societies transmission does not only happen *vertically*, from generation to generation, but also *horizontally*, within the members of a generation itself.[6] Equivalently, each generation has a *population* of agents (instead of just one agent in our case).

It would be intriguing, then, to see what kind of population effects would emerge by adding a horizontal transmission aspect to our system. The basis for this addition is actually already present:

Locomotion: Agents can already move around in physical space; the default agent behaviour is random wandering.

Dancing initiation: Agents already spontaneously initiate dancing; there is no need for a mechanism that randomly chooses agents and forces them to interact.

Periodic agent creation and removal: Each agent already has a predetermined lifetime ($2 \cdot t_{\text{learn}}$), after which they are removed from the simulation and replaced with a new learner.

However, just adding more agents to each generation would not work; the following aspects need to be modified first:

Spawning point: All new agents that enter the simulation are now placed in the same position; this would have to change. New agents could be placed in random positions, or in the former positions of the agents they are replacing.

[6] We mentioned some horizontal (Steels, 2015) and integrative (Theisen-White et al., 2011) approaches to experimental semiotics in section 2.2.

Decoupling from dancing: In the existing system, agents are forced to decouple from dancing after a time of t_{reset} and reset to their initial positions. This only works since just two agents are interacting at any time; for a population of agents, we would need to have a way for agents to "naturally" decouple from dancing.

The choice of spawning point for new agents is especially important: many interesting population effects, such as subcultures of "right" or "left" agents in different parts of the world or border effects where these subcultures meet, could depend on the initial positioning of agents in space.

7.7.3 More complex & spiking networks

One of the main shortcomings of our system is the simplicity of the neural network we used, which only has 5 nodes and is barely recurrent; while it is enough for the simple L/R dancing task, any more complicated task (such as the L/R/* game we described in section 7.7.1) will also need a more complex network. However, how would that network be designed? Randomly instantiating networks does not work, as most random networks have very unbalanced initial response rates and are thus not suitable for learning. There are at least two possible solutions to this issue:

Evolutionary search: One possible approach is to search the space of random networks for ones that are good candidates for learning using genetic algorithms. From an ecological perspective, it makes sense: organisms that can learn do not have "randomly instantiated" nervous systems. Rather, learning is a function that has evolved over long timescales and that requires specialised nervous system structures.

Of course, evolving learning as a function is an extremely ambitious goal and a research project in itself. What we are suggesting, though, is not to attempt to evolve learning, but rather to evolve networks by selecting them for their learning capacity, while still using the "tag & reward" learning mechanism we detailed in section 4.1. We have already tested this approach with mixed results; however, we used an earlier, flawed version of the learning algorithm. We did not continue pursuing it due to both time constraints and the fact that a simpler, hand-crafted network was sufficient for the minimal version of the L/R dancing task.

Spiking networks: A more principled and biologically plausible approach, however, would be to switch to large scale spiking networks instead of the recurrent neu-

ral network (RNN) implementation we are currently using. In general, while the RNN implementation of the distal reward algorithm (Izhikevich, 2007) works well for our basic network, it requires "stop gap" measures that nullify its advantage over spiking implementations (namely, computational complexity) and make its scaling ability uncertain.

As the working model of reinforcement learning by Izhikevich (2007) demonstrates, even randomly instantiated spiking networks can successfully learn. We postulate two reasons for this: firstly, the response decision (for example, for "right" and "left" behaviour) is not based on the activation of single neurons, but on the averaged activation of areas of neurons. Secondly, the scale of the network means that, despite random connections, all areas of the network have similar average activations.

7.7.4 Physical robots

Finally, as we mentioned in section 7.2, the modularity of our system allows us to easily replace the simulated e-pucks we have been using for all our experiments with physical robots. The use of physical robots can have certain advantages over simulated ones (see section 3.2.1 for a discussion), so this constitutes an interesting future possibility (and was also our original plan). There are two issues that need to be addressed, however:

Infrared sensors: E-puck robots use infrared sensors for proximity detection; however, a combination of the limited range of the sensors and the clear plastic build of the e-pucks (Fig. 3.2) made it so that they cannot reliably "see" each other. Possible ways around this are discussed in Longchamp et al. (2007).

Agent manipulation: In a simulator, it is easy to remove or add agents automatically; with physical robots, they have to be manipulated directly by the experimenter. This can be difficult, especially for longer chain experiments. A way around this would be to "virtually" replace agents by only overwriting their nervous system; this means, however, that since the robots cannot be manually moved around in space, dancing agents need to have a "decoupling" mechanism that stops them from dancing and moves them temporarily away from each other. Furthermore, it means that that new learner agents will be necessarily placed in the same space as the agents they are replacing.

7.8 A closing note

In the very beginning of this thesis, we stated our aspiration to do "philosophy of mind using a screwdriver" (Harvey, 2000). From the system design we described in Chapter 3, to the experiments of Chapters 4, 5 and 6 and finally the discussion in this chapter, it is apparent (we hope!) that we *have*, indeed, been using a screwdriver. Does this leave us any wiser, however, regarding the "philosophy of mind" part?

We would like to think that it does; by showing that transmission chains can be built from the bottom up, with as few operational assumptions as possible, we have placed the first screw in the construction of a bridge between pre-symbolic cognition and language. Admittedly, it is but a small screw, and the bridge will need many more to be stable. As we saw in this last chapter, however, the next step —an evolvable communication system— is definitely within reach, and it will place our experiments in proper "iterated learning" ground. There is a plan (and we have a screwdriver!).

Bibliography

Amodei, D., Anubhai, R., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Chen, J., Chrzanowski, M., Coates, A., Diamos, G., et al. (2015). Deep speech 2: End-to-end speech recognition in english and mandarin. *arXiv preprint arXiv:1512.02595*.

Becker, A. (1991). Language and languaging. *Language*, II(l):33–35.

Beer, R. (1995a). A dynamical systems perspective on agent-environment interaction. *Artificial Intelligence*, 72(1-2):173–215.

Beer, R. (1996). Toward the evolution of dynamical neural networks for minimally cognitive behavior. In Maes, P., editor, *From Animals to Animats 4: Proc. 4th Int. Conf. on Simulation of Adaptive Behavior*, pages 421–429. MIT Press.

Beer, R. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3):91–99.

Beer, R. D. (1995b). On the Dynamics of Small Continuous-Time Recurrent Neural Networks. *Adaptive Behavior*, 3(4):469–509.

Beer, R. D. and Gallagher, J. C. (1992). Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior*, 1(1):91.

Beer, R. R. D. (2003). The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11(4):209.

Belta, B. Y. C., Bicchi, A., Egerstedt, M., Frazzoli, E., Klavins, E., and Pappas, G. J. (2007). Symbolic planning and control of robot motion [grand challenges of robotics]. *IEEE Robotics & Automation Magazine*, 14(1):61–70.

Berwick, R. C., Pietroski, P., Yankama, B., and Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science*, 35(7):1207–1242.

Blynel, J. and Floreano, D. (2003). Exploring the t-maze: Evolving learning-like robot behaviors using ctrnns. In Cagnoni, S., editor, *Applications of Evolutionary Computing*, volume 2611, pages 593–604. Springer, Berlin.

Bottineau, D. (2008). Language and enaction. In *Enaction: Towards a New Paradigm for Cognitive Science*, pages 1–67. MIT Press.

Brighton, H. and Kirby, S. (2005). Cultural selection for learnability: three principles underlying the view that language adapts to be learnable. *Language origins: Perspectives*, pages 1–16.

Bullock, S. (2004). Making room for representation. *Adaptive Behavior*, 11(4):279–280.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. MIT Press.

Chomsky, N. (1992). On the nature, use and acquisition of language. In Putz, M., editor, *Thirty Years of Linguistic Evolution: Studies in Honour of René Dirven on the Occasion of His Sixtieth Birthday*, pages 3–29. John Benjamins Publishing.

Chomsky, N. (1995). *The Minimalist Program*. The MIT Press.

Chomsky, N. (2007). Of Minds and Language. *Biolinguistics*, 1(June 2006):9–27.

Chomsky, N. (2014). Minimal recursion: exploring the prospects. In *Recursion: Complexity in cognition*, pages 1–15. Springer.

Chow, S. J. (2013). What's the Problem with the Frame Problem? *Review of Philosophy and Psychology*, 4(2):309–331.

Claidière, N., Smith, K., Kirby, S., and Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1797).

Clark, A. (1997). The dynamical challenge. *Cognitive Science*, 21(4):461–481.

Clark, A. (1998). Magic words: how language augments human computation. In Carruthers, P. and Boucher, J., editors, *Language and thought: Interdisciplinary themes*, pages 162–183. Cambridge University Press.

Clark, A. and Toribio, J. (1994). Doing without representing? *Synthese*, 101(3):401–431.

Cliff, D., Husbands, P., and Harvey, I. (1993). Explorations in Evolutionary Robotics. *Adaptive Behavior*, 2(1):73–110.

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, 482(7383):85–88.

Cuffari, E. C., Di Paolo, E., and De Jaegher, H. (2015). From participatory sense-making to language: there and back again. *Phenomenology and the Cognitive Sciences*, 14(4):1089–1125.

Deacon, T. W. (2003). Multilevel selection in a complex adaptive system: The problem of language origins. [References]. *Evolution and Learning: The Baldwin Effect Reconsidered. Life and mind*, pages 81–106.

Di Paolo, E. (1997). An Investigation into the Evolution of Communication. *Adaptive Behavior*, 6(2):285–324.

Di Paolo, E. (2000). Behavioral Coordination, Structural Congruence and Entrainment in a Simulation of Acoustically Coupled Agents. *Adaptive Behavior*, 8(1):27–48.

Di Paolo, E. (2003). Evolving spike-timing-dependent plasticity for single-trial learning in robots. *Philosophical Transactions. Series A, Mathematical, physical, and engineering sciences*, 361(1811):2299–2319.

Di Paolo, E. (2004). Crawling out of the simulation: Evolving real robot morphologies using cheap, reusable modules. In *Artificial life IX: proceedings of the Ninth International Conference on the Simulation and Synthesis of Artificial Life*, volume 9, page 94. The MIT Press.

Di Paolo, E. (2005). Autopoiesis, adaptivity, teleology, agency. *Phenomenology and the Cognitive Sciences*, 4(4):429–452.

Di Paolo, E. (2009). Editorial: The social and enactive mind. *Phenomenology and the Cognitive Sciences*, 8(4):409–415.

Donald, M. (2000). The central role of culture in cognitive evolution: A reflection on the myth of the isolated mind.. In Nucci, L., Saxe, G., and Turiel, E., editors, *Culture, thought, and development*, pages 19–38. Lawrence Erlbaum.

Dreyfus, H. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical Psychology*, 20(2):247–268.

Dreyfus, H. and Dreyfus, S. (1988). Making a mind versus modeling the brain: Artificial intelligence back at a branchpoint. *Daedalus*, 117(1):15–43.

Edelman, S. (2003). But will it scale up? Not without representations. A commentary on The dynamics of active categorical perception in an evolved model agent by R. Beer. *Adaptive Behavior*, 11(4):273–275.

Elman, J. (1998). *Rethinking Innateness: Connectionist Perspective on Development*. MIT Press.

Fay, N., Garrod, S., Roberts, L., and Swoboda, N. (2010). The interactive evolution of human communication systems. *Cognitive Science*, 34(3):351–86.

Froese, T. (2011). From second-order cybernetics to enactive cognitive science: Varela's turn from epistemology to phenomenology. *Systems Research and Behavioral Science*, 28(6):631–645.

Froese, T. and Di Paolo, E. (2011). The enactive approach: Theoretical sketches from cell to society. *Pragmatics & Cognition*, 19(1):1–36.

Froese, T. and Di Paolo, E. a. (2009). Sociality and the lifemind continuity thesis. *Phenomenology and the Cognitive Sciences*, 8(4):439–463.

Froese, T. and Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, 173(3-4):466–500.

Funahashi, K. and Nakamura, Y. (1993). Approximation of dynamical systems by continuous time recurrent neural networks. *Neural networks*, 6(6):801–806.

Galantucci, B. (2009). Experimental Semiotics: A New Approach for Studying Communication as a Form of Joint Action. *Topics in Cognitive Science*, 1(2):393–410.

Galantucci, B. and Garrod, S. (2011). Experimental semiotics: a review. *Frontiers in human neuroscience*, 5(February):11.

Gallagher, S. (2008). Are minimal representations still representations? *International Journal of Philosophical Studies*, 16(3):351–369.

Gallagher, S. (2009). Two problems of intersubjectivity. *Journal of Consciousness Studies, 16*, (6):289–308.

Garrod, S., Fay, N., Lee, J., Oberlander, J., and MacLeod, T. (2007). Foundations of representation: Where might graphical symbol systems come from? *Cognitive Science*, 31(6):961–987.

Garzón, F. (2008). Towards a general theory of antirepresentationalism. *The British Journal for the Philosophy of Science*, 59(3):259.

Gold, E. (1967). Language identification in the limit. *Information and control*.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11):419–429.

Griffiths, T. and Kalish, M. (2005). A Bayesian view of language evolution by iterated learning. *Proceedings of the 27th annual conference of the cognitive science society*, pages 827–832.

Griffiths, T. L. and Kalish, M. L. (2007). Language evolution by iterated learning with bayesian agents. *Cognitive Science*, 31(3):441–80.

Harvey, I. (2000). Robotics: Philosophy of mind using a screwdriver. *Evolutionary robotics: From intelligent robots to artificial life*, 3:207–230.

Harvey, I., Husbands, P., Cliff, D., Thompson, A., and Jakobi, N. (1997). Evolutionary robotics: the Sussex approach. *Robotics and Autonomous System*, 20(2-4):205–224.

Hebb, D. O. (1949). *The Organization of Behaviour*.

Hoppitt, W. and Laland, K. (2008). Social processes influencing learning in animals: a review of the evidence. *Advances in the Study of Behavior*, 38.

Hoppitt, W. J. E., Brown, G. R., Kendal, R., Rendell, L., Thornton, A., Webster, M. M., and Laland, K. N. (2008). Lessons from animal teaching. *Trends in Ecology and Evolution*, 23(9):486–493.

Horner, V., Whiten, A., Flynn, E., and de Waal, F. B. M. (2006). Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the United States of America*, 103(37):13878–83.

Hudson Kam, C. and Newport, E. (2005). Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Language Learning and Development*, 1(2):151–195.

Hudson Kam, C. L. and Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1):30–66.

Ikegami, T. and Zlatev, J. (2007). From pre-representational cognition to language. In Ziemke, T., Zlatev, J., and Frank, R., editors, *Body, language and mind. Volume 1, Embodiment*. Mouton De Gruyter.

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070.

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17(10):2443–2452.

Jaeger, H. (2002). A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the "echo state network" approach. Technical report, Tech. Rep. GMD Report 159, 2002.

Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-organization of Brain and Behavior*. MIT Press.

Kirby, S. (2002a). Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Kirby, S. (2002b). The emergence of linguistic structure: An overview of the iterated learning model. *Simulating the evolution of language*.

Kirby, S., Cornish, H., and Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10681–6.

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141:87–102.

Kravchenko, A. (2004). Essential properties of language from the point of view of autopoiesis. *Signs*.

Kravchenko, A. (2011). How Humberto Maturana's biology of cognition can revive the language sciences. *Constructivist Foundations*, 6(3):352–362.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Lally, A. and Fodor, P. (2011). Natural language processing with prolog in the ibm watson system. *The Association for Logic Programming (ALP) Newsletter*.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Leydesdorff, L. (2000). Luhmann, Habermas and the theory of communication. *Systems Research and Behavioral Science*, 288(August 1998):273–288.

Loetzsch, M. and Spranger, M. (2010). Why robots? In *The Evolution of Language (EVOLANG 8)*, pages 222–229.

Longchamp, V., Roberts, J. F., and Bonani, M. (2007). Infrared relative positioning with the e-pucks. Technical Report June, EPFL.

Luhmann, N. (1986). The autopoiesis of social systems. *Sociocybernetic paradoxes*, 6(2).

Magnenat, S., Waibel, M., and Beyeler, A. (2011). Enki: The fast 2d robot simulator. *URL http://home.gna.org/enki*.

MATLAB (2014). Release 2014a.

Maturana, H. (1975). The organization of the living: a theory of the living organization. *International Journal of Man-Machine Studies*, (June 1974):149–168.

Maturana, H. (2002). Autopoiesis, structural coupling and cognition: a history of these and other notions in the biology of cognition. *Cybernetics & Human Knowing*, 3(4):5–34.

Maturana, H. R. (1978). Biology of language: The epistemology of reality. *Science*.

Maturana, H. R. and Varela, F. (1992). *Tree of Knowledge*. Shambhala.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Springer.

McMullin, B. (2004). Thirty years of computational autopoiesis: A review. *Artificial Life*, 10(3):277–295.

Michel, O. (2004). Webots TM : Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, 1(1):39–42.

Minsky, M. (1977). Frame-system theory. In Johnson-Laird, P. N. and Wason, P. C., editors, *Thinking*, pages 355–376. Cambridge: Cambridge University Press.

Mondada, F., Bonani, M., Raemy, X., Pugh, J., Cianci, C., Klaptocz, A., Zufferey, J.-c., Floreano, D., and Martinoli, A. (2006). The e-puck, a robot designed for education in engineering. *Robotics*, 1(1):59–65.

Morse, A. F. and Ziemke, T. (2007). Cognitive robotics, enactive perception, and learning in the real world. In *Proceedings of the Cognitive Science Society*, volume 29.

Newell, A. and Simon, H. a. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126.

Noë, A. (2004). *Action in Perception*. MIT Press.

Noe, A. (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Hill & Wang.

O'Regan, J. K. and Noë, A. (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese*, 129(1):79–103.

Oudeyer, P.-Y. (2005). How Phonological Structures Can Be Culturally Selected for Learnability. *Adaptive Behavior*, 13(4):269–280.

Oudeyer, P.-Y. (2013). Self-Organization: Complex Dynamical Systems in the Evolution of Speech. In Binder, P.-M. and Smith, K., editors, *The Language Phenomenon: Human Communication from Milliseconds to Millennia*, pages 191–216. Springer Berlin Heidelberg.

Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Duke University Press.

Pfeifer, R., Iida, F., and Bongard, J. (2005). New robotics: Design principles for intelligent systems. *Artificial Life*, 11(1-2):99–120.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Science*, 13(4):707–784.

Pullum, G. and Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review*, 19(i):9–50.

Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., Wheeler, R., and Ng, A. Y. (2009). Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, volume 3, page 5. Kobe.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J. and Sosík, P., editors, *Advances in Artificial Life: 6th European Conference, ECAL 2001 Prague, Czech Republic, September 10–14, 2001 Proceedings*. Springer Berlin Heidelberg.

Routledge, P., Wheeler, M., Routledge, P., and Wheeler, M. (2008). Minimal representing: A response to Gallagher. *International Journal of Philosophical Studies*, 16(3):371–376.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Pearson.

Schumacher, P. (2011). *The Autopoiesis of Architecture: A New Framework for Architecture*. Wiley.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65–84.

Smith, K., Kirby, S., and Brighton, H. (2003). Iterated learning: A framework for the emergence of language. *Artificial Life*, 9(4):371–86.

Smith, K. and Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3):444–449.

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems*, 16(5):16–22.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7):308–312.

Steels, L. (2012). Grounding language through evolutionary language games. In Steels, L. and Hild, M., editors, *Language Grounding in Robots*, pages 1–22. Springer.

Steels, L. (2015). *The Talking Heads experiment: Origins of words and meanings*, volume 1. Language Science Press.

Steels, L. and Kaplan, F. (1999). Bootstrapping grounded word semantics. In Briscoe, T., editor, *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*. Cambridge University Press.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112.

Suzuki, K. and Ikegami, T. (2009). Shapes and self-movement in protocell systems. *Artificial Life*, 15(1):59–70.

Tamariz, M. and Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, 8:37–43.

Theisen-White, C., Kirby, S., and Oberlander, J. (2011). Integrating the horizontal and vertical cultural transmission of novel communication systems. In *CogSci 2011 Proceedings: Cognitive Science Society Conference*, volume 1, pages 956–961.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21(5):615–628.

Van Gelder, T. and Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In Port, R. F., editor, *Mind as Motion: Explorations in the Dynamics of Cognition*, pages 1–43. Bradford Books.

Varela, F. G., Maturana, H. R., and Uribe, R. (1974). Autopoiesis: The organization of living systems, its characterization and a model. *BioSystems*, 5(4):187–196.

Varela, F. J., Rosch, E., and Thompson, E. (2000). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press.

Vaughan, R. (2008). Massively multi-robot simulation in stage. *Swarm Intelligence*, 2(2-4):189–208.

Verhoef, T., de Boer, B., and Kirby, S. (2012). Holistic or synthetic protolanguage: Evidence from iterated learning of whistled signals. In Scott-Philips, T. C., Tamariz, M., Cartmill, E. A., and Hurford, J. R., editors, *The Evolution of Language: Proceedings of the 9th International Conference (EVOLANG9)*, pages 368–375. World Scientific.

Verhoef, T., Kirby, S., and de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43(1):57–68.

Verhoef, T., Kirby, S., and Padden, C. (2011). Cultural emergence of combinatorial structure in an artificial whistled language. In Carlson, L., Holscher, C., and Shipley, T., editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 483–488. Cognitive Science Society.

Vervaeke, J., Lillicrap, T. P., and Richards, B. A. (2012). Relevance realization and the emerging framework in cognitive science. *Journal of Logic and Computation*, 22(1):79–99.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence*, 167(1):206–242.

Weber, A. (2002). Life after Kant: Natural purposes and the autopoietic foundations of biological individuality. *Phenomenology and the Cognitive Sciences*, 1(2):97–125.

Wellens, P., Loetzsch, M., and Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, 20(2-3):173–191.

Wheeler, M. (2005). *Reconstructing the Cognitive World: The Next Step*. MIT Press.

Yamauchi, B. M. and Beer, R. D. (1994). Sequential behavior and learning in evolved dynamical neural networks. *Adaptive Behavior*, 2(3):219–246.

Zeleny, M. and Pierre, N. A. (1976). Simulation of self-renewing systems. In Jantsch, E. and Waddington, C., editors, *Evolution and Consciousness*. Addison Wesley.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*, pages 51–58. MIT Press.