The Ecology of Open-Ended Skill Acquisition

Clément Moulin-Frier

HABILITATION

of the University of Bordeaux

Specialty: Computer Science

Defended by

Clément Moulin-Frier

The Ecology of Open-Ended Skill Acquisition

Computational framework and experiments on the interactions between environmental, adaptive, multi-agent and cultural dynamics

---

Prepared at Inria Bordeaux, Flowers Team

defended on December 7th, 2022 at **Inria Bordeaux Sud-Ouest** Center

---

Composition of the jury

Reviewers:	Dr. Jean-Baptiste Mouret	-	Inria, Nancy, France
	Prof. Yukie Nagai	-	University of Tokyo, Japan
	Prof. Dan Dediu	-	Univeristy of Barcelona, Spain
Examinators:	Prof. Nicolas Bredeche	-	Sorbonne University, France
	Dr. David Ha	-	Stability AI, Japan
	Dr. Emmanuel Dupoux	-	EHESS, Paris, France
Mentor:	Dr. Pierre-Yves Oudeyer	-	Inria, Bordeaux, France

Abstract

An intriguing feature of the human species is our ability to continuously invent new problems and to proactively acquiring new skills in order to solve them: what is called open-ended skill acquisition (OESA). Understanding the mechanisms underlying OESA is an important scientific challenge in both cognitive science (e.g. by studying infant cognitive development) and in artificial intelligence (aiming at computational architectures capable of open-ended learning). Both fields, however, mostly focus on cognitive and social mechanisms at the scale of an individual's life. It is rarely acknowledged that OESA, an ability that is fundamentally related to the characteristics of human intelligence, has been necessarily shaped by ecological, evolutionary and cultural mechanisms interacting at multiple spatiotemporal scales.

In this thesis, I present a research program aiming at understanding, modeling and simulating the dynamics of OESA in artificial systems, grounded in theories studying its eco-evolutionary bases in the human species. It relies on a conceptual framework expressing the complex interactions between environmental, adaptive, multi-agent and cultural dynamics. Three main research questions are developed and I present a selection of my contributions for each of them.

- What are the ecological conditions favoring the evolution of skill acquisition?
- How to bootstrap the formation of a cultural repertoire in populations of adaptive agents?
- What is the role of cultural evolution in the open-ended dynamics of human skill acquisition?

By developing these topics, we will reveal interesting relationships between theories in human evolution and recent approaches in artificial intelligence. This will lead to the proposition of a humanist perspective on AI: using it as a family of computational tools that can help us to explore and study the mechanisms driving open-ended skill acquisition in both artificial and biological systems, as a way to better understand the dynamics of our own species within its whole ecological context.

This document presents an overview of my scientific trajectory since the start of my PhD thesis in 2007, the detail of my current research program, a selection of my contributions as well as perspectives for future work.

Acknowledgments

To my wife, Jasmine Garside, the most fascinating human I have ever met.

To my families: the Moulin-Frier/Kryzak family and the Garside family. To the missed ones who left way too early: Hannah-Rose and Tomy.

To my friends, who are impossible to list here. They will anyway hardly reach these lines but if so, they will recognize them if they met me in the main cities where I have lived. In chronological order: Lans-en-Vercors, Grenoble, Los Angeles, Paris, Bordeaux and Barcelona.

To my scientific mentors: Pierre-Yves Oudeyer and Jean-Luc Schwartz. Many other people have had a strong positive influence on my scientific life (see below), but this thesis would have never come to existence without these two.

To my dear colleagues, who again are hard to enumerate but let's give it a quick try by quickly adapting my co-author list (sorry for those I missed): Julien Diard, Pierre Bessiere, Marti Sanchez-Fibla, Xerxes D. Arsiwalla, Jordi Ysard Puigbò, Giovanni Maffei, Diogo Pata, Klaudia Grechuta, Vasiliki Vouloutsi, Ismael T. Freire, Ivan Herreros, Paul Verschure, Tristan Karch, Mayalen Etcheverry, Cédric Colas (thanks for the Latex template!), Tobias Fischer, Pierre Rouanet, Fabien Benureau, Sao Mai Nguyen, Michael Arbib, Ricard Solé, Steve N'Guyen, Jacques Droulez, Louis-Jean Boë, Pascal Barla, Xavier Hinaut, Nathalie Robin.

To all the other important people in my life that I probably forgot in these acknowledgments written too rapidly.

To the complexity of nature which never stops to fascinate me and is driving most of my scientific interests. Let's cite Sean M. Carroll and his interesting philosophy of *poetic naturalism* (Carroll, 2016):

> *The universe doesn't care about us, but we care about the universe. That's what makes us special, not any immaterial souls or special purpose in the grand cosmic plan. Billions of years of evolution have created creatures capable of thinking about the world, forming a picture of it in our minds and holding it up to scrutiny. We are interested in the world, in its physical manifestations and in our fellow humans and other creatures. That caring, contained inside us, is the only source of "mattering" in any cosmic sense. Whenever we ask ourselves whether something matters, the answer has to be found in whether it matters to some person or persons. We take the world and attach value to it, an achievement of which we can be justly proud. (The Big Picture, by Sean M. Carroll (Carroll, 2016))*

Bordeaux, November 2, 2022

Contents

Acronyms

AI artificial intelligence

DNN deep neural network

DRL deep reinforcement learning

HBE human behavioral ecology

IM intrinsic motivation

IMGEP intrinsically motivated goal exploration process

MARL multi-agent reinforcement learning

META-RL meta reinforcement learning

ML machine learning

OESA open-ended skill acquisition

PCV Pulsed Climate Variability

RL reinforcement learning

Chapter 1

Introduction

1.1 Outline

This thesis is organized into two parts. In Part I I first detail my scientific trajectory in Chapter 2, from the start of my PhD thesis in 2007 to my recruitment as a permanent researcher in the Flowers group at Inria in 2019. Then, in Chapter 3, I present the research program I have initiated since my recruitment as a permanent researcher. This research program aims at understanding, modeling and simulating the dynamics of open-ended skill acquisition in artificial systems, grounded in theories studying its eco-evolutionary bases in the human species. It relies on a conceptual framework expressing the complex interactions between environmental, adaptive, multi-agent and cultural dynamics.

Then, in Part II, I present selected contributions from my entire career – attempting to reinterpret them in light of the proposed framework – as well as perspectives for future work.

1.2 How to read this thesis

I obviously recommend reading this thesis from the start to the end. However, I provide below recommendations for readers who have limited time to read it or are looking for specific information.

The main contribution of this thesis is developed in Chapter 3, which takes the form of a new, self-contained position paper where I explain in detail my current research program. I propose a conceptual framework for *grounding artificial intelligence in the origins of the human behavior*, proposing to understand, model and simulate the emergence of open-ended skill acquisition in humans and machines through an artificial ecology approach. This chapter details the motivation behind this framework and its structure, as well as its position with respect to the existing literature in biological and artificial open-ended skill acquisition. The core arguments of the framework can be found in Section 3.3. It is structured around three main research questions:

- What are the ecological conditions favoring the evolution of skill acquisition? (Section 3.4)

- How to bootstrap the formation of a cultural repertoire in populations of adaptive agents? (Section 3.5)
- What is the role of cultural evolution in the open-ended dynamics of human skill acquisition? (Section 3.6)

In Part II, I summarize selected contributions realized during my career. I take the occasion to reinterpret these past and present contributions in light of the framework proposed in Chapter 3, structuring them according to the three research questions above (in Chapters 5, 6 and 7, respectively). Therefore, these contributions were selected according to several criteria: their impact of course, but also their potential to illustrate how the multiple parts of the proposed framework interact together. I conclude each chapter of this part with concrete propositions for future work.

Finally, information on my curriculum is located in Chapter 2, which provides a summary of my scientific trajectory as well as my full Curriculum Vitae (including my publication list).

Part I

From my scientific trajectory to my current research program

Chapter 2

Scientific trajectory

Contents

2.1 Short curriculum

My research mostly focuses on understanding and modeling the origins of complex social behavior in humans and machines. The underlying theoretical question pertains to the origins of cooperation and communication and the role of self-organization mechanisms in the co-structuration of sensorimotor skills and social interactions. My approach aims to analyze and model those mechanisms in their evolutionary, developmental and socio-cultural contexts in order to extract strong computational principles. In turn, those principles are implemented and simulated to support, invalidate or reframe existing theoretical hypotheses from the literature. For this aim, I rely on theories and hypotheses from life science (language evolution, speech science, infant development and motor control, human evolution, behavioral ecology, evolutionary biology) and use methods from computer science as modeling tools (cognitive modeling, control theory, machine learning, multi-agent systems, complex systems, developmental robotics, evolutionary computation). As a result of this modeling and simulation effort, testable hypotheses on the origins of social behaviors, as well as real-world applications in artificial intelligence and robotics, can be proposed. My scientific trajectory evolved from a focus on the evolution and acquisition of speech and language, to a broader scope on the origins of open-ended skill acquisition in humans and machines. I detail below this scientific trajectory, step by step (I am co-author of all the references cited in this section).

My research career started in 2007 when I started my PhD at the Gipsa-Lab[1] in Grenoble, under the supervision of Jean-Luc Schwartz, Pierre Bessière and Julien Diard. At this point, I became interested in modeling and simulating the emergence of speech

[1] http://www.gipsa-lab.grenoble-inp.fr/en/home.php

and language in populations of sensorimotor agents (Moulin-Frier et al., 2008, 2011, 2015a; Schwartz et al., 2015, see Section 6.3 for my main contribution on this topic). I studied the theories of the origin of language as well as the neuroanatomical circuits of speech perception and production, notably during a 6-month visit at the University of Southern California in Los Angeles with Michael A. Arbib (Arbib & Moulin-Frier, 2010).

At the end of my thesis in September 2011, I realized a short post-doctoral project of 4 months at the Laboratoire de Physiologie de la Perception et de l'Action (LPPA) of the Collège de France[2] with Jacques Droulez, where I worked on probabilistic methods of optimal control (N'Guyen et al., 2013).

I was then recruited in January 2012 for a 20-month post-doctoral position in the Flowers[3] research team at Inria Bordeaux Sud-Ouest in France. My contract will be extended by 16 months (for a total of 36 months) as a research scientist. With Pierre-Yves Oudeyer, director of the team, I worked on the main topics forming the research program of developmental robotics in the context of his ERC Starting Grant EXPLORERS. Making the link with my thesis work, I applied in a novel way the algorithmic architectures for intrinsically motivated autonomous exploration (Moulin-Frier & Oudeyer, 2013a,b; Moulin-Frier et al., 2014b) to the modeling of early vocal development in humans and machines (Moulin-Frier & Oudeyer, 2012, 2013b; Moulin-Frier et al., 2014a, 2017a, see Section 6.2 for my main contribution on this topic).

As of January 2015, I was recruited at the Synthetic, Perceptive, Emotive and Cognitive Systems (SPECS)[4] laboratory of the Universitat Pompeu Fabra in Barcelona as a post-doctoral researcher coordinating the scientific and technological integration of the European project WYSIWYD[5], with Paul Verschure. This project involved six research labs on an ambitious project focusing on language acquisition in human-robot interaction (Fischer et al., 2018; Moulin-Frier et al., 2016, 2018; Puigbò et al., 2015c,b,a, 2016; Sánchez-Fibla et al., 2017). During this period, I also worked on emerging social conventions in multi-agent reinforcement learning (Freire et al., 2020; Moulin-Frier et al., 2015b) and made several theoretical contributions in artificial intelligence and cognitive science (Moulin-Frier et al., 2017b; Arsiwalla et al., 2018; Moulin-Frier & Verschure, 2016).

Then, in October 2017, I joined the US-based artificial intelligence research company Cogitai[6] (now Sony-AI America), founded by renowned scientists in reinforcement learning (Peter Stone, Satinder Bajeva, Mark Ring) and whose goal is to build the first operational architecture for continuous learning from experience. I was trained in the methods resulting from recent advances in artificial intelligence, in particular deep reinforcement learning and its variations, including those inspired by developmental robotics that I had studied during my projects at Inria-Flowers.

Finally, in October 2019 I was recruited as a permanent researcher in the Flowers[3] research group at Inria in France. I am co-supervising 3 PhD students and 1 postdoctoral researcher on topics ranging from multi-agent reinforcement learning (Moulin-Frier &

[2] https://www.college-de-france.fr/site/en-college/index.htm
[3] https://flowers.inria.fr/
[4] https://specs-lab.com/
[5] http://wysiwyd.specs-lab.com/
[6] https://cogitai.com/

Oudeyer, 2021; Nisioti et al., 2022, see Section 7.3 for a contribution on this topic), language-based reinforcement learning (Colas et al., 2020, 2022a; Karch et al., 2020, see Section 7.2 for a contribution on this topic), artificial life (Etcheverry et al., 2020; Hamon et al., 2022, see Section 5.4 for a contribution on this topic) and evolutionary computation (Nisioti & Moulin-Frier, 2022; Colas et al., 2021, see Section 5.2 for a contribution on this topic). More generally, I have initiated a long-term research program aiming at understanding and modeling how open-ended skill acquisition emerges in populations of artificial agents through the complex interaction between environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms operating at multiple spatiotemporal scales. For this aim, I have obtained funding from Inria (Exploratory Action ORIGINS and Cordi programs) as well as the French National Agency for Research (ANR, project ECOCURL). Chapter 3, which can be considered the main contribution of this thesis, is explaining in detail the motivation behind this research program.

2.2 An integrative effort

During this career, a recurrent pattern has emerged: I have a particular interest in integrating different scientific trends into unified conceptual or computational frameworks. Here are a few examples. The model I developed during my PhD thesis, originally for simulating the emergence of speech sound systems in agent populations, has actually contributed to the unification of major theories in speech production and perception processing in a unified Bayesian formulation (Moulin-Frier et al., 2012, 2010). During my post-doc position at Flowers, I have proposed a probabilistic unification of the major exploration strategies proposed in developmental robotics (Moulin-Frier & Oudeyer, 2013a). During my post-doc position at SPECS, I have proposed that major existing architectures in artificial intelligence (from top-down symbolic reasoning, to behavioral robotics, to modern deep reinforcement learning) can be expressed within the context of an integrated layered cognitive architecture (Moulin-Frier et al., 2017b).

The conceptual framework I will propose in Chapter 3 is the last and most developed instantiation of this integrative effort. It highlights strong relationships between recent contributions in artificial intelligence and hypotheses in human behavioral ecology, a research field studying how ecological constraints and opportunities could have shaped human behavior throughout its evolutionary history. This leads to the proposition of a conceptual framework emphasizing that open-ended skill acquisition is not only implemented as a cognitive mechanism, but is the product of complex interactions between environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms.

2.3 Curriculum Vitae (including publication list)

CLÉMENT MOULIN-FRIER, PHD

Address:	27 rue de Grassi 33000 Bordeaux, France	**Date of Birth:**	27th of May 1981 41 year old
Email:	clement.moulin-frier@inria.fr	**Nationality:**	French
Web:	http://clement-moulin-frier.github.io	**Phone:**	(+33) 6.62.56.42.89

Last update: October 25, 2022

CURRENT POSITION

Since October 2019

PERMANENT RESEARCHER (CRCN)

Inria, Bordeaux, FLOWERS research group, France

Research: *From ecological constraints to the emergence of language: Grounding Artificial Intelligence in the origins of human behavior*
Website: https://flowers.inria.fr
Funding:
- Permanent position at Inria funded by the French government
- Inria Exploratory Action ORIGINS (2020-2022)*: Grounding AI in the Origins of Human Behavior*
- French National Research Agency (ANR) (2021-2025): *Emergent Communication through Curiosity-driven Multi-Agent Reinforcement Learning*

PREVIOUS POSITIONS

October 2017 – July 2019

RESEARCH SCIENTIST

Cogitai, Inc. (Now Sony-AI America), Orange County, USA

Research: Cogitai, Inc. is dedicated to building artificial intelligences (AIs) that learn continually from interaction with the real world. Our goal is to build the brains, i.e., the continual-learning AI software, that will let everyday things that sense and act get smarter with experience. This experience will be shared across devices and domains to allow the rapid scaling-up of learning.
Executives: Peter Stone, Satinder Singh, Mark Ring
Website: https://www.cogitai.com

January 2015 – October 2017

POST-DOCTORAL RESEARCHER

SPECS research group, Universitat Pompeu Fabra, Barcelona, Spain

Research: *Adaptive cognitive architectures for robotics and the emergence of social behaviors*
Supervision: Paul Verschure
Funding:
- *What You Say Is What You Did,* WYSIWYD project (FP7 ICT 612139)
- *Socialising Sensori-Motor Contingencies,* socSMC project (641321-H2020FETPROACT-2014).
- *Role of Consciousness in Adaptive Behavior* ERC's CDAC project (ERC-2013ADG 341196)

January 2012 – November 2014

RESEARCHER

FLOWERS research group, Inria, Bordeaux, France.

Research: *Curiosity-driven learning applied to robotics and emergent communication*
Supervision: Pierre-Yves Oudeyer
Funding: ERC Starting Grant EXPLORERS 240 007, then Inria institute

September 2011 – December 2011

POST-DOCTORAL RESEARCHER

LPPA (Physiology of Perception and Action), Collège de France CNRS, Paris, France.

Research: *Bayesian models of decision making for bipedal walking control*
Supervision: Jacques Droulez
Funding: French government

January 2009 – July 2009

VISITING SCHOLAR

University of Southern California, Los Angeles, USA.

Research: *Recognizing speech in a novel accent: The motor theory of speech perception reframed*
Supervision: Michael A. Arbib
Funding: Explora-Doc French scholarship

ACADEMIC EDUCATION

September 2007 – June 2011

PHD STUDENT

Gipsa-Lab, Speech and Cognition department, Grenoble University

Research: *Emergence of communication systems in Bayesian vocal agent populations*
Supervision: Jean-Luc Schwartz, Pierre Bessière, and Julien Diard
Research stay: 6 months with M.A. Arbib at University of Southern California, Los Angeles, USA.
Funding: French ministry research scholarship

September 2006 – July 2007

MASTER DEGREE IN COGNITIVE SCIENCE

Grenoble Institute of Technology, France

With honors
Grenoble Institute of Technology

September 2005 – July 2006

MASTER DEGREE IN COMPUTER SCIENCE

With honors
University Joseph Fourier, Grenoble, France

Scientific Responsibilities

2020-2022

Co-organizer of the *SMILES* workshop

International Conference on Development and Learning, ICDL (virtual conference)
https://sites.google.com/view/smiles-workshop/

2019 – 2020

Co-edition of the Frontiers Research Topic *"Emergent Behavior in Animal-Inspired Robotics"*

https://www.frontiersin.org/research-topics/13627/emergent-behavior-in-animal-inspired-robotics

2015

Program chair

International Conference on Development and Learning, ICDL/Epirob, Providence, RI, USA

www.icdl-epirob.org

2015-present

Associate Editor

International Conference on Development and Learning, ICDL/Epirob

www.icdl-epirob.org

2015-present

Chair of the Language and Cognition Task Force

IEEE Technical Committee on Cognitive and Developmental Systems
https://sites.google.com/view/ieeetflanguageandcognition/home

2014 – 2015

Co-editor of the Special Issue "On the cognitive nature of speech sound systems"

Journal of Phonetics

First author of the target article, see *Publications*

2015

Member of the program committee

Workshop "Sensorimotor Contingencies for Robotics" at IROS 2015, Hamburg, Germany

http://www.iri.upc.edu/groups/perception/sensorimotorIROS15/

Awards, Honors, Grants and Competitions

2020 – 2022

Inria Exploratory Action

ORIGINS: *Grounding Artificial Intelligence in the Origins of Human Behavior*
Funding of a 2-year post-doc position

2021 – 2025

French National Research Agency (ANR)

Emergent Communication through Curiosity-driven Multi-Agent Reinforcement Learning

Funding of a 3-year PhD position and a 18-month Research Engineer position

2020 – 2022

Inria Cordi Grant

Emergent Communication through Curiosity-driven Multi-Agent Reinforcement Learning

Funding of a 3-year PhD position

2017 – 2020

Plan Nacional (Spanish research grant)

INSOCO project (DPI2016-80116-P)

Social interactions based on sensorimotor contingencies
In collaboration with Marti Sanchez-Fibla.
http://specs.upf.edu/projects/3159

November 2012

Best paper Award

International Conference on Development and Learning, ICDL/Epirob, San Diego, USA.

Category: Computational models of development
Paper: Curiosity-driven phonetic learning, see *Publications*

2012

Qualification as Associate Professor (Renewed in 2018)

French ministry of research.

Domain: Computer Science

2009

Best teaching project

Grenoble CIES (French center for university-level teaching)

2008

Explora-doc scholarship

French Rhone-Alpes region

Funding for a 6-months visit at the University of Southern California, Los Angeles, USA

2006

Master degree fellowship

French government

Invited Talks

October 2022

“Grounding Artificial Intelligence in the Origins of Human Behavior: The ORIGINS project”

Annual joint workshop Inria-DFKI at Inria Bordeaux Sud-Ouest, France

Invitation from Frederic Alexandre, Inria

December 2021

“OPEN-ENDED SKILL ACQUISITION IN HUMANS AND MACHINES: AN EVOLUTIONARY AND DEVELOPMENTAL PERSPECTIVE”

Brains@Bay meetup (organized by Numenta, USA)

Invitation from Numenta

September 2021

“THE ROLE OF SELF-ORGANIZATION MECHANISMS IN THE EMERGENCE OF BEHAVIORAL REGULARITY AND DIVERSITY”

Preprogrammed: Innateness in Neuroscience and AI (Symposium organized by the Nencki Institute, Poland)

Invitation from Mateusz Kostecki

April 2021

“BEYOND THE UTILITARIAN APPROACH TO EMERGENT COMMUNICATION: THE ROLE OF SELF-ORGANIZATION AND COMPOSITIONAL IMAGINATION IN LANGUAGE EVOLUTION AND CULTURAL INNOVATION”

Deepmind reading group

Invitation from Florian Strub and Julien Pérolat (Deepmind, Paris)

November 2019

“SELF-ORGANIZATION OF COMMUNICATION SYSTEMS IN LEARNING AGENTS”

Neurorobotics workshop @ ENSEIRB, Bordeaux, France

Invitation from Xavier Hinaut (Inria)

October 2019

“ACTIVE LEARNING AND CURIOSITY IN ROBOTICS”

French national days of robotics research (JNRR), Vittel, France

Invitation from David Filliat (ENSTA Paristech)

June 2017

“COGNITION, EMBODIMENT AND SELF-ORGANIZATION: AN INTEGRATED VIEW TO ARTIFICIAL INTELLIGENCE”

Machine Learning group at Universitat Pompeu Fabra

Invitation from Hector Geffner, head of the group

March 2017

“COGNITIVE ARCHITECTURES FOR SOCIAL ROBOTICS”

European Robotics Forum, Edinburgh, Scotland

Empathic Human-Robot Interaction Workshop
Invitation from Kerstin Dautenhahn, organizer of the workshop

August 2015

“EVOLUTION AND DEVELOPMENT OF VOCAL COMMUNICATION STRUCTURES”

Princeton University, Developmental Neuromechanics & Communication Lab, USA

Invitation from Asif Ghazanfar, head of the group

November 2014

“EXPLORATION STRATEGIES IN DEVELOPMENTAL ROBOTICS”

Humanoids conference, Madrid, Spain. Workshop

Workshop “Active Learning in Robotics: Exploration Strategies in Complex Environments”
Organisers: Johannes Kulick, Herke van Hoof, Marc Toussaint, and Jan Peters

October 2014

“POPPY: A ROBOTIC PLATFORM FOR CODERS, MAKERS, ARTISTS AND RESEARCHERS”

Pycon conference, Lyon, France

Invitation by Françoise Conil, co-organizer of the conference

August 2013

“EXPLORATION STRATEGIES IN DEVELOPMENTAL ROBOTICS”

Honda Research Institute, Tokyo, Japan

Invitation by Angelica Lim, visiting scholar and now researcher at Softbank Robotics

TEACHING ACTIVITIES

2015-2022

RESPONSIBLE PROFESSOR (77.5 HOURS)

Universitat Pompeu Fabra, Barcelona, Spain

Course “Real-time Interaction in Cognitive and Social Systems” (2015-2017) and “System Design, Integration and Control” (2017-2022)
Cognitive Systems and Interactive Media (CSIM) Master

2020

IA2 / ROBOTICS COURSE (12 HOURS)

Centre de Recherche Interdisciplinaire (CRI, Paris)

Master AIRE: Interdisciplinary Approaches in Research and Education

2010--2011

TEACHING ASSISTANT IN COMPUTER SCIENCE (92 HOURS)

UFR IMAG, University Joseph Fourier, Grenoble, France

Computer Science and Applied Mathematics
Bachelor and Master degrees

2007--2010

TEACHING ASSISTANT IN COMPUTER SCIENCE (192 HOURS)

Université Stendhal, Grenoble, France

3 years of teacher training

STUDENT SUPERVISION

2022 – 2025

PHD THESIS SUPERVISION

Flowers group, Inria, France

Student: Gautier Hamon
Grounding Artificial Intelligence in the Origins of Human Behavior

2020 – 2023

PHD THESIS SUPERVISION

Flowers group, Inria, France

Student: Julius Taylor
Emergent Communication through Curiosity-driven Multi-Agent Reinforcement Learning

2020 – 2023

PHD THESIS SUPERVISION

Flowers group, Inria, France

Student: Mayalen Etcheverry
Automated Discovery in Complex Systems

2019 – 2022

PHD THESIS CO-SUPERVISION

Flowers group, Inria, France

Student: Tristan Karch
Grounded language learning and curiosity-driven exploration with deep reinforcement learning

2021 – 2022

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Erwan Plantec, Université de Lorraine, Nancy, France
Flow Lenia: Mass conservation for the study of virtual creatures in continuous cellular automata

2021 – 2022

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Elías Masquil, MVA Master, Paris-Saclay, France
Intrinsically Motivated Goal-Conditioned Reinforcement Learning in Multi-Agent Environments

2021 – 2022

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Yoann Lemesle, Université Paris Dauphine- PSL
Self-organization of shared graphical languages in groups of agents using multimodal contrastive deep learning mechanisms

2020 – 2021

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Mateo Mahaut, Ecole Nationale Supérieure de Cognitique, Bordeaux, France
Towards an ecologically valid simulation environment for reinforcement learning

2020 – 2021

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Katia Jodogne del Litto, École Polytechnique, Paris, France
Role of the structure of social networks on collective innovation in multi-agent reinforcement learning

2019 – 2020

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Valentin Villecroze, École Polytechnique, Paris, France
Emergence of communication systems as a way to maintain cooperative networks in multi-agent simulations

2019 – 2020

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Younès Rabii, Ecole Normale Supérieure de Cognitique, Bordeaux, France
Simulation environment and ecologically valid interaction scenarios for multi-agent reinforcement learning

2015-2017

PHD THESIS CO-SUPERVISION

SPECS group, Universitat Pompeu Fabra, Spain

Student: Jordi-Ysard Puigbo
Value modulation in cortical visual processing and application to robotic control

2016 – 2017

MASTER PROJECT SUPERVISION

SPECS group, Universitat Pompeu Fabra, Spain

Student: Ismael Tito Freire González, CSIM Master, UPF, Spain
Modeling the formation of social conventions in agent populations

2015 – 2016

MASTER PROJECT SUPERVISION

SPECS group, Universitat Pompeu Fabra, Spain

Student: Yasin Can Akmehmet, CSIM Master, UPF, Spain
Autonomous development of turn-taking behaviors in robot populations

2014 – 2015

MASTER PROJECT, THEN PHD THESIS CO-SUPERVISION

Flowers group, Inria, France

Student: Sébastien Forestier, Ecole Normale Supérieur, Paris, France
Active learning strategies for the modelling of infant vocal development

2014 – 2013

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Marie-Morgane Paumard, Ecole Normale Supérieur de Cachan, France
Learning the manipulation of flexible tools in developmental robotics: a fishing robot

2013 – 2014

MASTER PROJECT SUPERVISION

Flowers group, Inria, France

Student: Jules Brochard, Ecole Normale Supérieur de Cachan, France
Emergent maturations in early vocal development. Journal article, see *Publications*

2010 – 2011

MASTER PROJECT SUPERVISION

GIPSA-Lab, Grenoble Institute of Technology, France

Student: Raphaël Laurent, Master MoSIG, ENSIMAG, Grenoble, France

A computational model to study quantitatively motor, sensory, and sensorimotor model responses in Speech Recognition. 3 co-authored publications, including a journal paper

PUBLICATIONS

All my publications are available open-access on my Google Scholar profile.

JOURNAL ARTICLES

Colas, C., Karch, T., Moulin-Frier, C., & Oudeyer, P. Y. (2022). Vygotskian Autotelic Artificial Intelligence: Language and Culture Internalization for Human-Like AI. *arXiv preprint arXiv:2206.01134*. To appear in *Nature Machine Intelligence*.

Colas, C., Hejblum, B., Rouillon, S., Thiébaut, R., Oudeyer, P. Y., Moulin-Frier, C., & Prague, M. (2021). Epidemioptim: A toolbox for the optimization of control policies in epidemiological models. *Journal of Artificial Intelligence Research*, *71*, 479-519.

Demirel, B., *Moulin-Frier, C.*, Arsiwalla, X. D., Verschure, P. F., & Sánchez-Fibla, M. (2021). Distinguishing Self, Other, and Autonomy From Visual Feedback: A Combined Correlation and Acceleration Transfer Analysis. *Frontiers in Human Neuroscience*, 443.

Kusters, R., Misevic, D., Berry, H., Cully, A., Le Cunff, Y., Dandoy, L., ..., Moulin-Frier, C., … & Wehbi, F. (2020). Interdisciplinary Research in Artificial Intelligence: Challenges and Opportunities. *Frontiers in Big Data*, 45.

Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D., & Verschure, P. F. M. J. (2020). Modeling the formation of social conventions from embodied real-time interactions. *PLOS ONE*, *15*(6), e0234434. https://doi.org/10.1371/journal.pone.0234434

Sanchez-Fibla, M., Forestier, S., Moulin-Frier, C., Puigbo, J.-Y. and Verschure, P. F. (2019) “From motor to visually guided bimanual affordance learning,” *Adaptive Behavior*. https://doi.org/10.1177/1059712319855836

Moulin-Frier, C., Fischer, T., Petit, M. Pointeau, G., Puigbo, J.-Y., Pattacini, U., Low, S.C., Camilleri, D., Nguyen, P. Hoffmann, M. Chang, H.J., Zambelli, M., Mealier, A.-L., Damianou, A., Metta, G. Prescott, T., Demiris, Y., Dominey, P.-F. and. Verschure, P. (2018). DAC-h3: A Proactive Robot Cognitive Architecture to Acquire and Express Knowledge About the World and the Self. *IEEE Transactions on Cognitive and Developmental Systems*. 10(4): 1005 – 1022.

Fischer, T., Puigbò, J.-Y., Camilleri, D., Nguyen, P. D. H., Moulin-Frier, C., Lallée, S., Metta, G., Prescott, T. J., Demiris, Y., & Verschure, P. F. M. J. (2018). ICub-HRI: A Software Framework for Complex Human–Robot Interaction Scenarios on the iCub Humanoid Robot. *Frontiers in Robotics and AI*, *5*, 22. https://doi.org/10.3389/frobt.2018.00022

Moulin-Frier, C., Brochard, J., Stulp, F., & Oudeyer, P.-Y. (2017). Emergent Jaw Predominance in Vocal Development through Stochastic Optimization. *IEEE Transactions on Cognitive and Developmental Systems*. Early access: https://ieeexplore.ieee.org/document/7955101

Acevedo Valle, J. M., Angulo, C., & Moulin-Frier, C. (2017). Autonomous Discovery of Motor Constraints in an Intrinsically-Motivated Vocal Learner. *IEEE Transactions on Cognitive and Developmental Systems*. 10(2): 314 – 325.

Moulin-Frier, C., Diard, J., Schwartz, J.-L., and Bessière, P. (2015). COSMO (“Communicating about Objects using Sensory-Motor Operations”): a Bayesian modeling framework for studying speech communication and the emergence of phonological systems. *Journal of Phonetics*. 53: 5–41 **Target paper of a special issue**.

Moulin-Frier, C., Nguyen, S. M., and Oudeyer, P.-Y. (2013). Self-organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Psychology (Cognitive Science)*, 4(1006).

Moulin-Frier, C. and Arbib, M. A. (2013). Recognizing speech in a novel accent: The motor theory of speech perception reframed. *Biological Cybernetics*, 107 (4):421–447.

N’Guyen, S., Moulin-Frier, C., and Droulez, J. (2013). Decision Making under Uncertainty: A Quasimetric Approach. *PLoS ONE*, 8(12).

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L., and Diard, J. (2012). Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: an exploratory Bayesian modeling study. *Language and Cognitive Processes*. 27(7-8): 1240–1263. Special Issue: Speech Recognition in Adverse Conditions.

INVITED COMMENTARIES IN INTERNATIONAL JOURNALS

Moulin-Frier, C., & Verschure, P. (2016). Two possible driving forces supporting the evolution of animal communication: Comment on “Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain” by Michael A. Arbib. *Physics Of Life Reviews*, 16, 88–90.

Schwartz, J.-L., Barnaud, M.-L., Bessière, P., Diard, J., & Moulin-Frier, C. (2016). Phonology in the mirror: Comment on “Towards a Computational Comparative Neuroprimatology: Framing the language-ready brain” by Michael A. Arbib . *Physics Of Life Reviews*, 16, 93–95.

Laurent, R., Moulin-Frier, C., Bessière, P., Schwartz, J.-L., & Diard, J. (2013). Integrate yes, but what and how? A computational approach of sensorimotor fusion in speech. Commentary In *Behavioral and Brain Sciences*, 36(4):36–37.

BOOK CHAPTERS

Ten, A., Oudeyer, P. Y., & Moulin-Frier, C. (2022). Curiosity-Driven Exploration: Diversity of mechanisms and functions. Chapter in *The Drive for Knowledge: The Science of Human Information Seeking*, 53.

Moulin-Frier, C., Schwartz, J., Diard, J., and Bessière, P. (2011b). Emergence of articulatory-acoustic systems from deictic interaction games in a "Vocalize to Localize" framework. Chapter in Primate communication and human language: Vocalisations, gestures, imitation and deixis in humans and non-humans. *Advances in Interaction Studies'* series by John Benjamins Pub. Co.

INTERNATIONAL CONFERENCES – FULL PAPERS

Nisioti, E., & Moulin-Frier, C. (2022). Plasticity and evolvability under environmental variability: the joint role of fitness-based selection and niche-limited competition. *Proceedings of the 2022 Genetic and Evolutionary Computation Conference (GECCO 2022).*

Taylor, J., Nisioti, E., & Moulin-Frier, C. (2022). Socially Supervised Representation Learning: the Role of Subjectivity in Learning Efficient Representations. *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2022)*

Karch, T., Teodorescu, L., Hofmann, K., Moulin-Frier, C., & Oudeyer, P. Y. (2021). Grounding Spatio-Temporal Language with Transformers. *Advances in Neural Information Processing Systems*, *34*.

Nisioti, E., Jodogne-del Litto, K., & Moulin-Frier, C. (2021). Grounding an Ecological Theory of Artificial Intelligence in Human Evolution. In *NeurIPS 2021-Conference on Neural Information Processing Systems / Workshop: Ecological Theory of Reinforcement Learning*.

Colas, C., Karch, T., Lair, N., Dussoux, J. M., Moulin-Frier, C., Dominey, P., & Oudeyer, P. Y. (2020). Language as a cognitive tool to imagine goals in curiosity driven exploration. *Advances in Neural Information Processing Systems*, *33*, 3761-3774.

Etcheverry, M., Moulin-Frier, C., & Oudeyer, P. Y. (2020). Hierarchically organized latent modules for exploratory search in morphogenetic systems. *Advances in Neural Information Processing Systems*, *33*, 4846-4859.

Karch, T., Colas, C., Teodorescu, L., Moulin-Frier, C., & Oudeyer, P.-Y. (2020). Deep Sets for Generalization in RL. *Beyond "Tabula Rasa" in Reinforcement Learning (BeTR-RL) Workshop*. International Conference on Learning Representations (ICLR 2020). http://arxiv.org/abs/2003.09443

Moulin-Frier, C., Puigbò, J. Y., Arsiwalla, X. D., Sanchez-Fibla, M., & Verschure, P. F. (2017). Embodied Artificial Intelligence through Distributed Adaptive Control: An Integrated Framework. *International Conference on Development and Learning, ICDL/Epirob, Lisbon, Portugal*.

Arsiwalla, X.D., Herreros, I., Moulin-Frier, C., Verschure, P.F.M.J (2017). Consciousness as an Evolutionary Game-Theoretic Strategy. In *Conference on Biomimetic and Biohybrid Systems*, 509-514

Sanchez-Fibla, M., Moulin-Frier, C., Arsiwalla, X. and Verschure, P. (2017) Social Sensorimotor Contingencies: Towards Theory of Mind in Synthetic Agents, in *Recent Advances in Artificial Intelligence Research and Development: Proceedings of the 20th International Conference of the Catalan Association for Artificial Intelligence*, vol. 300, p. 251.

Sanchez-Fibla, M., Moulin-Frier, C. and Verschure, P. (2017) A sensorimotor account of visual and tactile integration for depth perception: An iCub robot experiment. *International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 2017, pp. 86–91.

Moulin-Frier, C., Arsiwalla, X. D., Puigbò, J.-Y., Sánchez-Fibla, M., Duff, A., and Verschure, P. F. M. J. (2016). Top-Down and Bottom-Up Interactions between Low-Level Reactive Control and Symbolic Rule Learning in Embodied Agents. In *Proceedings of the Workshop on Cognitive Computation: Integrating neural and symbolic approaches. 30th Annual Conference on Neural Information Processing Systems (NIPS 2016)*.

Puigbò, J.-Y., Moulin-Frier, C., and Verschure, P. F. M. J. (2016). Towards Self-controlled Robots Through Distributed Adaptive Control. In *Conference on Biomimetic and Biohybrid Systems* (pp. 490–497). Springer.

Arsiwalla, X. D., Herreros-Alonso, I., Moulin-Frier, C., Sánchez-Fibla, M., and Verschure, P. F. M. J. (2016). Is Consciousness a Control Process? In *Proceedings of the 19th International Conference of the Catalan Association for Artificial Intelligence*.

Acevedo Valle, J. M., Angulo Bahón, C., Moulin-Frier, C., Trejo Ramírez, K. A. (2016). The role of somatosensory models in vocal autonomous exploration. In *Revista Internacional de Investigación e Innovación Tecnológica* 4 (23), 1-11

Moulin-Frier, C., Sanchez-Fibla, M., and Verschure, P. F.M.J (2015b). Autonomous development of turn-taking behaviors in agent populations: a computational study. In *International Conference on Development and Learning, ICDL/Epirob, Providence (RI), USA*.

Puigbò, J.-Y., Moulin-Frier, C., Vouloutsi, V., Sanchez-Fibla, M., Herreros, I., and Verschure, P. F. M. J. (2015). Skill refinement through cerebellar learning and human haptic feedback: an iCub learning to paint experiment. In *IEEE-RAS Conference on Humanoids Robots (Humanoids 2015)*, Seoul, Korea.

Puigbò, J.-Y., Herreros, I., Moulin-Frier, C., and Verschure, P. F. M. J. (2015). Towards a two-phase model of sensor and motor learning. In *Conference on Biomimetic and Biohybrid Systems* (pp. 453–460). Springer.

Acevedo Valle, J. M., Angulo, C., Agell, N., and Moulin-Frier, C. (2015). Proprioceptive Feedback and Intrinsic Motivations in Early-Vocal Development. In *Proceedings of the 18th International Conference of the Catalan Association for Artificial Intelligence*. Armengol, E., Boixader, D., Grimaldo, F.

Moulin-Frier, C. and Oudeyer, P.-Y. (2013a). Exploration strategies in developmental robotics: A unified probabilistic framework. In *International Conference on Development and Learning, ICDL/Epirob, Osaka, Japan*.

Moulin-Frier, C. and Oudeyer, P.-Y. (2013b). Learning how to reach various goals by autonomous interaction with the environment: unification and comparison of exploration strategies. In *1st Multidisciplinary Conference on Reinforcement Learning and Decision Making (RLDM2013), Princeton University, New Jersey.*

Moulin-Frier, C. and Oudeyer, P.-Y. (2013c). The role of intrinsic motivations in learning sensorimotor vocal mappings: a developmental robotics study. In *Proceedings of Interspeech, Lyon, France*, Lyon, France.

Moulin-Frier, C. and Oudeyer, P.-Y. (2012). Curiosity-driven phonetic learning. In *International Conference on Development and Learning, Epirob, San Diego, USA*. **Best paper award.**

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J., and Diard, J. (2011a). Noise and inter-speaker variability improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception : An exploratory bayesian modeling study. In *9th International Seminar on Speech Production, ISSP'11*, Montreal, Canada.

Moulin-Frier, C., Schwartz, J., Diard, J., and Bessière, P. (2010). A unified theoretical bayesian model of speech communication. In *1st conference on Applied Digital Human Modeling, Miami, USA*.

Moulin-Frier, C., Schwartz, J., Diard, J., and Bessière, P. (2008c). Emergence of a language through deictic games within a society of sensori-motor agents in interaction. In *International Workshop on "Speech and Face to Face Communication"*, Grenoble France.

Moulin-Frier, C., Schwartz, J., Diard, J., and Bessière, P. (2008b). Emergence of a language through deictic games within a society of sensori-motor agents in interaction. In *8th International Seminar on Speech Production, ISSP'08*, Strasbourg, France.

Moulin-Frier, C., Schwartz, J., Diard, J., and Bessière, P. (2008a). Emergence du langage par jeux déictiques dans une société d'agents sensori-moteurs en interaction. In *27e Journées d'Etudes sur la Parole, JEP'2008*, Avignon France.

International Conferences – Short papers

Villecroze, V., & Moulin-Frier, C. (2020). Studying the joint role of partial observability and channel reliability in emergent communication. In *1st SMILES (Sensorimotor Interaction, Language and Embodiment of Symbols) workshop, ICDL 2020.*

Moulin-Frier, C., & Oudeyer, P.-Y. (2020). Multi-Agent Reinforcement Learning as a Computational Tool for Language Evolution Research: Historical Context and Future Challenges. *Challenges and Opportunities for Multi-Agent Reinforcement Learning (COMARL), AAAI Spring Symposium Series, Stanford University, Palo Alto, California, USA*.

Puigbò, J.-Y., Vouloutsi, V., Moulin-Frier, C., & Verschure, P. F. M. J. (2015). Reactive and adaptive control loops for social learning in human--robot interaction. Workshop "Mechanisms of learning in social contexts", *IEEE International Conference on Development and Learning, ICDL/Epirob*, Providence (RI), USA.

Moulin-Frier, C., Rouanet, P., and Oudeyer, P.-Y. (2014). Explauto: an open-source Python library to study autonomous exploration in developmental robotics. In *International Conference on Development and Learning, ICDL/Epirob, Genova, Italy*.

Arbib, M. A. and Moulin-Frier, C. (2010). Recognizing speech in a novel accent: The motor theory of speech perception reframed. In *Neurobiology of Language Conference, San Diego, USA*.

Schwartz, J., Rochet-Capellan, A., and Moulin-Frier, C. (2007). Speech at reach of hand and mouth: Theoretical arguments, experimental facts and computational advances. In *Workshop " Vocoid – Vocalization, COmmmunication, Imitation and Deixis in adult and infant human and non human primates "*, Grenoble France.

Blog posts

Hamon, G., Etcheverry, M., Chan, B. W. C., Moulin-Frier, C., & Oudeyer, P. Y. (2022). Learning Sensorimotor Agency in Cellular Automata. Blog Post: https://developmentalsystems.org/sensorimotor-lenia/

Colas, C., Karch, T., Moulin-Frier, C., & Oudeyer, P. Y. (2021). Language as a Cognitive Tool: Dall-E, Humans and Vygotskian RL Agents. http://developmentalsystems.org/language_as_cognitive_tool_vygotskian_rl

Preprints

Lemesle, Y., Karch, T., Laroche, R., Moulin-Frier, C., & Oudeyer, P. Y. (2022, submitted). Emergence of Shared Sensory-motor Graphical Language from Visual Input. *arXiv preprint arXiv:2210.06468*.

Nisioti, E., Mahaut, M., Oudeyer, P. Y., Momennejad, I., & Moulin-Frier, C. (2022, submitted). Social Network Structure Shapes Innovation: Experience-sharing in RL with SAPIENS. *arXiv preprint arXiv:2206.05060*.

Arsiwalla, X.D., Moulin-Frier, C., Herreros, I., Sanchez-Fibla, M., Verschure, P.F.M.J (2017). The Morphospace of Consciousness. *arXiv preprint arXiv:1705.11190*

THESES

Moulin-Frier, C. (2011). Rôle des relations perception-action dans la communication parlée et l'émergence des systèmes phonologiques : étude, modélisation computationnelle et simulations. PhD thesis, Université de Grenoble.

Moulin-Frier, C. (2007). Jeux déictiques dans une société d'agents sensori-moteurs en interaction. Master's thesis, Grenoble-INP.

Moulin-Frier, C. (2006). Objets communicants : la traçabilité. Master's thesis, Université Joseph Fourier, Grenoble.

OUTREACH ACTIVITIES & INNOVATION

September 2017

CO-ORGANIZER OF THE RE-FLUX PERFORMANCE

Barcelona Cognition Brain and Technology summer school (BCBT 2016)

Multimodal Performance with AI, Robots, VR and Humans
http://bcbt.upf.edu/bcbt16/node/330

June 2015

CO-ORGANIZER OF A ROBOTIC ARTISTIC PERFORMANCE

Music Hack Day @ Sonar Festival, Barcelona

Audio synthesis using robotic bodies
http://new.musichackday.org/

2013 – 2017

INITIATOR AND MAIN CONTRIBUTOR OF THE OPEN-SOURCE *EXPLAUTO* LIBRARY

A library to study, model and simulate intrinsically motivated multitask learning and exploration in virtual and robotic agents
https://github.com/flowersteam/explauto

October 2014

ORGANISATION OF A 3-DAY HACKATHON

Universciences, Paris, France

Conception and programming of the Poppy robot. 25 participants
Video of the event: https://vimeo.com/109145300

September 2014

INTERVIEW FOR THE FRENCH JOURNAL BIOFUTUR

On robotic approaches to language evolution modelling

2013 – 2015

Member of the Poppy-Project

Open-source robotics for teacher, makers, artists and researchers
Realisation of a various robotic demonstrations, workshops and dissemination events
https://www.poppy-project.org

Languages

English (C1), French (native), Spanish (B1)

Recommendations

Pierre-Yves Oudeyer

INRIA Research Director, Head of the Flowers research group, Bordeaux, France
PhD thesis reviewer and post-doc advisor (2012-2014)
Specialized in developmental robotics.
pierre-yves.oudeyer@inria.fr

Peter Stone

Professor at *The University of Texas at Austin.*
COO of *Cogitai Inc.* where I worked as a Research Scientist (2017-2019)
Specialized in reinforcement learning
pstone@cs.utexas.edu

Pierre Bessiere

CNRS Research Director, Sorbonne Universités – UPMC -ISIR, Paris, France
PhD thesis advisor
Specialized in computer and cognitive sciences
pierre.bessiere@isir.upmc.fr

Yiannis Demiris

Head of the Personal Robotics Laboratory, Imperial College, London, UK
Collaborator in the WYSIWYD European project (2015-2017)
Specialized in human-robot interaction and machine learning
y.demiris@imperial.ac.uk

Paul Verschure

Head of the SPECS research group, Universitat Pompeu Fabra, Barcelona, Spain
Post-doc advisor (2015-present)
Specialized in computational neuroscience, psychology and robotics
paul.verschure@upf.edu

Michael A. Arbib

Professor, USC Brain Project, University of Southern California, Los Angeles, USA
Collaborator and PhD thesis reviewer
Specialized in computational neuroscience and language evolution
arbib@usc.edu

MATTHIEU LAPEYRE

CEO and co-founder of Pollen Robotics, Bordeaux, France. http://pollen-robotics.com/en
Designer of the *Poppy* humanoid robot for which I have realized a number of applications
Specialized in open-source robotics
matthieu.lapeyre@pollen-robotics.com

JEAN-LUC SCHWARTZ

CNRS Research Director, GIPSA-Lab, Speech and Cognition Dpt, Grenoble, France
PhD thesis advisor
Specialized in speech science
jean-luc.schwartz@gipsa-lab.grenoble-inp.fr

JACQUES DROULEZ

CNRS Research Director, Sorbonne Universités – UPMC -ISIR, Paris, France
Post-doc advisor
Specialized in computer and cognitive sciences
jacques.droulez@isir.upmc.fr

ANNE WARLAUMONT

Head of the Emergence of Communication Lab, UC Merced, USA
Collaborator and co-chair of the 2015 ICDL-Epirob conference
Specialized in computational models of speech acquisition.
awarlaumont2@ucmerced.edu

Chapter 3

Research program: Grounding Artificial Intelligence in the Origins of Human Behavior

Contents

Context

This chapter has been written for the purpose of this thesis in the form of a self-contained position paper, which I then plan to submit to a journal. It describes in detail a conceptual framework resulting from the Exploratory Action[1] I am leading at Inria: the so-called ORIGINS project[2] (2020-2022). This chapter further develops previous short position papers (Nisioti & Moulin-Frier, 2020; Nisioti et al., 2021) written with my colleague Eleni Nisioti, the post-doc researcher recruited on the ORIGINS project. Eleni Nisioti has had a major role in co-elaborating this research program.

I first present different approaches to artificial intelligence (AI) and argue that a central feature of human-like intelligence is our tendency to autonomously invent new problems and to proactively learn how to solve them: what is called *open-ended skill acquisition*

[1] https://www.inria.fr/en/inrias-exploratory-actions-taking-risks

[2] https://www.inria.fr/en/origins-grounding-artificial-intelligence-origins-human-behaviour

(OESA) (Section 3.1). Then, I highlight strong relationships between recent contributions in AI and hypotheses in human behavioral ecology (HBE), a research field studying how ecological constraints and opportunities could have shaped human behavior throughout its evolutionary history (Section 3.2). This leads to the propositon of a conceptual framework emphasizing that OESA is not only implemented as a cognitive mechanism but is the product of complex interactions between environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms (Section 3.3). I provide a detailed analysis of the relationships between these two fields (AI and HBE) developing the original research questions it raises (Sections 3.4 to 3.6) and proposing a roadmap to address them (Section 3.7).

3.1 Introduction

What does *intelligence* means in the context of artificial intelligence (AI)? There is no clear consensus about this question in the AI community, which is reflected in the diversity of terms used to characterize the end-goal of AI, e.g. artificial general intelligence (Goertzel & Pennachin, 2007), strong AI (Searle, 1980), human-level AI (McCarthy, 2007), unified theories of cognition (Newell, 1994), universal intelligence (Legg & Hutter, 2007), enactive AI (Froese & Ziemke, 2009) or open-ended intelligence (Weinbaum & Veitas, 2017).

An early proposition from Newell (1994) states that "*a system is intelligent to the degree that it approximates a knowledge-level system*", i.e. a system "*embedded in an external environment, with which it interacts by a set of possible actions [...] to attain its goals, using all the knowledge that it has*". Many other definitions have been proposed: as paraphrased by Chollet (2019), Legg et al. (2007) summarized in 2007 no fewer than 70 definitions from the literature into a single statement: "*Intelligence measures an agent's ability to achieve goals in a wide range of environments*". Goertzel & Pennachin (2007) define artificial general intelligence as "*AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation*". More recently, Chollet (2019) proposes that "*the intelligence of a system is a measure of its skill-acquisition efficiency over a scope of tasks, with respect to priors, experience, and generalization difficulty*".

The definitions mentioned above are however purely mechanistic, in the sense that they focus on the final state of an intelligent agent and charaterize it through measurable properties. This comes in contrast with the enactive view of cognition and AI, which instead emphasizes the processes driving the emergence of autonomy and intelligence (Varela et al., 1991; Brooks, 1991; Froese & Ziemke, 2009; Weinbaum & Veitas, 2017; De Jaegher & Di Paolo, 2007). I describe below the main hypotheses and methods driving these two approaches to artificial intelligence.

3.1.1 Current approaches in the quest for artificial intelligence

This quest for artificial intelligence has been revived by recent advances in machine learning (ML), leading many renowned researchers in the field to elaborate on prospective paths toward human-level AI. We can roughly categorize them into two main categories, mirroring the mechanistic vs. enactive views mentioned above.

3.1.1.1 Cognition-centric approach: Focus on architectures and optimization

.

This approach considers achieving human-level AI mostly as an engineering problem. The main assumption is that intelligence must be implemented in a structured cognitive architecture (integrating e.g. control, learning and memory mechanisms) which is optimized (using ML methods) through pre-defined objective functions. The recent rise of deep neural networks as powerful function approximators has strongly revived this approach by allowing key advances in e.g. representation learning and reinforcement learning in high-dimensional spaces (Mnih et al., 2015).

Recent position papers and debates among influential researchers in the field illustrate well the principles behind this dominant approach. A very recent and prototypical example is a position paper from Yann Lecun where he sketches an integrated cognitive architecture for achieving human-level intelligence (LeCun, 2022). This extract from the paper's abstract illustrates well the usual workflow of the cognition-centric approach: "*This position paper proposes an architecture and training paradigms with which to construct autonomous intelligent agents. It combines concepts such as configurable predictive world model, behavior driven through intrinsic motivation, and hierarchical joint embedding architectures trained with self-supervised learning*" (LeCun (2022)). As we can see, the approach clearly focuses on how to engineer a cognitive architecture in terms of interacting modules as well as how to optimize it using machine learning algorithms.

Within this cognition-centric approach, a debate has been initiated between proponents of a data-driven view (e.g. Silver et al., 2021), with the provocative claim that achieving human-level AI is nowadays mostly a matter of scaling up existing architectures, datasets and environments; vs. proponents of an architecture-driven view emphasizing the central role of priors and symbolic reasoning, that are central in human intelligence but poorly addressed by DNN models (e.g. Marcus & Davis, 2019). In both views, the proposed methods are evaluated in benchmarks designed to capture various aspects of intelligence. For example, Chollet (2019) defines intelligence as "*a measure of its skill-acquisition efficiency over a scope of tasks, with respect to priors, experience, and generalization difficulty*" and proposes a benchmark to evaluate it inspired by psychometric intelligence tests and called the *The Abstraction and Reasoning Corpus (ARC)*.

In summary, the cognition-centric approach proposes to formalize a given definition of intelligence as an objective function, optimize the parameters of a cognitive architecture against this objective and evaluate progress through adequate benchmarks testing their ability to generalize on test environments. In this sense it mirrors the standard methodology adopted in cognitive psychology and neuroscience, where researchers measure the behavioral and neural response of participants attempting to achieve pre-defined tasks.

3.1.1.2 Emergentist approach: Focus on environmental and multi-agent dynamics

.

In contrast to the previous approach considering intelligence as the result of a well-optimized cognitive architecture, an alternative approach is to view it as the emergent product of adaptive systems interacting with complex environmental dynamics. In other words, if the cognition-centric approach attempts to reverse-engineer the brain (or at least its main functions), the emergentist approach instead attempts to reverse-engineer the environmental conditions that lead to intelligence. There are two main propositions here. Some contributions study how competition and cooperation pressures in populations of co-adapting agents can result in a behavioral arms race where each agent has to continuously improve its skills against those of other agents, an approach called *multi-agent autocurriculum* (Leibo et al., 2019). Other contributions study how learning algorithms themselves can be meta-learned for operating in a diversity of environments. Clune (2020) calls this approach *AI-Generating Algorithms (AI-GA)*, with three main pillars: "*(1) meta-learning architectures, (2) meta-learning the learning algorithms themselves, and (3) generating effective learning environments*". In both propositions (autocurriculum and AI-GA), it is the complexity of the environment (either through the presence of other co-adapting agents or through its intrinsic diversity) that drives the ability to continuously acquire new skills and generalize them in novel environments. This approach is usually less concerned with human cognition than the cognition-centric approach. Instead, it takes its inspiration from the fields of complex systems (Ha & Tang, 2021) and evolutionary biology. This limited grounding in human intelligence is well illustrated in this quote from Clune (2020):

> *A major open question that remains is how we can constrain the generation of environments to be those we find interesting and/or that produce intelligence that helps us solve real-world problems. [...] For example, one might argue that such a system could produce intelligence that is alien to us and that we cannot communicate with. However, if it is truly general intelligence, presumably through its learning efforts and our own we could learn to communicate with it. Additionally, creating alien forms of intelligence would be fascinating as it would teach us about the limits and possibilities within the space of intelligent beings.*

In conclusion, the cognition-centric and the emergentist approaches are dual. On the one hand, the cognition-centric approach is focused on human intelligence as a measurable property and attempts at simulating, at different levels of abstraction, the cognitive processes underlying it. Its search space is the space of cognitive architectures and processes, while environmental properties belong to the experimental benchmarks designed for evaluating the proposed architectures. In this view, AI is strongly connected to human cognition through the modeling of mechanisms such as learning and memory for solving complex tasks. The cognition-centric approach however shows major limitations: its search space is incredibly vast (including all possible combinations of all possible implementations of diverse cognitive modules, Clune, 2020); it is highly subject to deception (focusing on improving performance measures can prevent to explore alternative promising paths to AI, Stanley & Lehman, 2015) and it mostly ignores non-cognitive mechanisms that could have played a major role in the emergence of intelligence (e.g. environmental complexity or multi-agent dynamics). On the other hand, the emergentist approach instead considers intelligence as the emergent product of a larger dynamical system involving complex environmental and multi-agent dynamics. Its search space is the space of dynamical systems potentially driving the emergence of intelligent behavior, i.e. environmental properties are considered as experimental variables. While the emergentist approach addresses most of the limitations of the cognition-centric approach, it usually considers intelligence in all its generality and rarely cares about the specificities of human intelligence. For this reason, it has rarely resulted in real-world applications and will unlikely lead to human-like intelligence (although it could, in theory, discover other interesting types of intelligence).

The main objective of this chapter is to propose a conceptual framework for grounding the emergentist approach in the origins of human behavior.

3.1.2 Human-like intelligence as open-ended skill acquisition

As mentioned above, the AI community has not yet converged on a clear consensus about what is meant by *intelligence*, nor to a consensus on how to achieve artificial intelligence. Which phenomena is considered as the most relevant to study mostly depends on the considered approach, e.g. benchmarked learning and generalization abilities in the cognition-centric approach vs. co-adaptation patterns between adaptive agents and dynamic environments in the emergentist approach. In this thesis, I propose to focus on what I consider to be the most interesting distinctive feature of human intelligence: *open-ended skill acquisition* ; and I propose a path to study its emergence in artificial systems.

Open-endedness is the property of a mechanism to continuously generate increasingly diverse and complex structures (Stanley & Lehman, 2015; Banzhaf et al., 2016). A typical example is biological evolution, where the relatively simple mechanism of natural selection with variation has produced the open-ended diversity of biological organisms that have lived on Earth. But many other mechanisms are open-ended in the natural world. The dynamics of the environment is to some extent open-ended, generating cyclic patterns at multiple spatio-temporal scales (e.g. seasonal cycles and climate variations) as well as complex chemical reactions (some of these reactions being at the origins of life itself, Walker et al., 2017). Developmental and cognitive mechanisms in the human species are also open-ended (e.g. skill and knowledge acquisition), as well as socio-cultural mechanisms (e.g. technological innovation, social organization, cultural evolution, Stanley, 2019; Solé et al., 2013).

In fact, a distinctive feature of the human species is our tendency to continuously invent new problems and to proactively acquire new skills in order to solve them, in an open-ended fashion: what is called *open-ended skill acquisition (OESA)* (Figure 3.1). This process is open-ended both at the individual scale (each of us is able to continuously acquire new knowledge and skills) and at the socio-cultural scale (our societies generate an open-ended repertoire of cultural innovations). OESA is part of what makes us truly humans, in the sense that it has enabled our most impressive achievements through science, art and technological innovation. It is however also resulting in the over-exploitation of natural resources and the misuse of technology, putting at risk our own survival as a species through the collapse of biodiversity and global warming (e.g. Cook et al., 2016).

Figure 3.1: An intriguing feature of the human species is our ability to continuously invent new problems and to proactively acquire new skills in order to solve them: what is called open-ended skill acquisition (OESA). This ability results from heterogeneous mechanisms operating at multiple spatiotemporal scales: environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms.

Understanding and modeling OESA has been already recognized as an important challenge in both cognitive science and in AI. In cognitive science, pioneer work from the 40s attempted to theorize the manifest drive of human individuals to perform activities not only for obtaining external rewards (e.g. related to feeding or mating) but also because those activities are intrinsically enjoyable in and of themselves. They coined the term *intrinsic motivation* to express this human drive to continuously invent their own challenges and try to solve them, what we call in everyday language *curiosity* (Deci & Ryan, 1985; Berlyne, 1954; Csikszentmihalyi, 1997). More recently, some research groups have been seeking the neural correlates of curiosity in primates (Gottlieb et al., 2013). In AI, the computational modeling of OESA has first attracted interest in the 90s (Schmidhuber, 1991; Schmidhuber et al., 1997). In the 2000s, *developmental robotics* emerged as a scientific field attempting to implement mechanisms inspired by infant development into algorithms that can operate on simulated or physical robots (Barto et al., 2004; Baranes & Oudeyer, 2013). More recently, principles first proposed in developmental robotics have been integrated to deep reinforcement learning algorithm (Pathak et al., 2017; Colas et al., 2019), what is sometimes called *developmental AI* (Colas et al., 2022b).

As we can see, the cognition-centric approach is here again dominating the field: research in both cognitive science and AI most often considers that the mechanisms underlying OESA are mostly cognitive. It is still rarely acknowledged that open-ended skill acquisition, an ability that is fundamentally related to the characteristics of human intelligence (Banzhaf et al., 2016), has necessarily been shaped by our environmental, evolutionary, social and cultural history, through complex feedforward and feedback effects operating at multiple spatiotemporal scales.

3.1.3 What are the mechanisms promoting open-ended skill acquisition in biological and artifical agents?

In this thesis, I instead propose to zoom out from this cognition-centric view, by considering OESA within its whole substrate. This requires acknowledging that OESA is not only implemented as a cognitive mechanism but is the product of complex interactions between environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms. For this aim, we will turn our attention towards human behavioral ecology (HBE), a research field studying how ecological constraints and opportunities could have shaped human behavior throughout its evolutionary history (Borgerhoff Mulder & Schacht, 2001; Nettle et al., 2013). Research in this field have studied, among others: how speciation, extinction and dispersal arose in the human history depending on climate variation (Maslin et al., 2015), the formation of cooperative groups (Chapman & Chapman, 2000), the rise of resource management (Gowlett, 2016), of tool use (deBeaune et al., 2004), the evolution of human language (Freeberg et al., 2012) and the emergence of cultural norms and institutions (Tomasello, 2009). **What principles can we extract from this literature that could help us to achieve OESA in artificial agents?**

In other words, our objective is not to directly implement agents capable of OESA – this is the role of the cognition-centric approach – but instead (1) to understand the origins of OESA in the human species in the light of the HBE literature, (2) to extract computational principles from this literature and to implement them in computer simulations, (3) to study how these principles can results in novel AI algorithms and, cherry on the cake, (4) to propose or reframe hypotheses in HBE.

Before going further we must acknowledge that, admittedly, there are many paths to the acquisition of open-ended skills in AI. Grounding our study in human behavioral ecology seems to be but one of the options, but several reasons may persuade us to explore it.

1. Examining *all* possible paths to OESA is infeasible considering our modern and foreseeable computational power (Clune, 2020).

2. Species that are more familiar to ours make it easier to define evaluation criteria. For example, human-inspired metrics such as equality, sustainability and social welfare have been employed to evaluate agents on their ability to forage (Pérolat et al., 2017), find optimal taxation strategies (Zheng et al., 2020) and cooperate (Baker et al., 2020).

3. Darwinian evolution offers an existence proof for human-like OESA (Clune, 2020), as well as empirical data and testable hypotheses.

4. Similar attempts at grounding AI research in a non-computational field have already proven to be a fruitful approach. For example, concepts from Development Science such as intrinsic motivation (Oudeyer et al., 2007; Pathak et al., 2017) and embodied language acquisition (Cangelosi et al., 2010) have had a significant impact on modern AI research.

5. The potential of knowledge transfer between human behavioral ecology and artificial intelligence has already been recognized (Frankenhuis et al., 2019), with the transfer of ideas having the opposite direction from the one proposed here. A proposal to study major evolutionary transitions in order to understand the general laws that underlie innovation and transfer insights to artificial evolution is presented in Solé (2016). Our proposal follows a similar direction but focuses on highlighting the overlap between concepts in AI and in HBE. In addition, a number of works in AI have recently resorted to theories from ecology, psychology and economics for inspiration (Wang et al., 2019a; Hughes et al., 2018; Jaques et al., 2019; Pérolat et al., 2017; Köster et al., 2022).

The rest of this chapter is organized as follows. In Section 3.2, I provide a review of existing contributions studying open-ended skill acquisition in both AI and in the human species. In Section 3.3, I present the core contribution of this thesis: a novel conceptual framework for understanding, modeling and simulating the emergence of OESA in artificial systems. Then, in Sections 3.4, 3.5 and 3.6, I detail the different parts of this framework and their underlying dynamics. A concluding discussion is provided in Section 3.7.

3.2 Existing approaches to open-ended skill acquisition in AI and the human species

In this section, I present existing approaches studying OESA in both AI and in the human species. I propose to organize these approaches into three categories according to their main focus: behavioral, environmental or social diversity.

3.2.1 Approaches focusing on behavioral diversity

3.2.1.1 In the human species

Open-ended skill acquisition in humans is most often considered as a product of an intrinsic motivation to acquire new knowledge and skills, or in everyday language: curiosity. The term *intrinsic motivation* refers to a mechanism driving individuals to select and engage in activities for their own sake because they are inherently interesting (in opposition to extrinsic motivation, which refers to doing something because it leads to a separable outcome, e.g. food) (Deci & Ryan, 1985).

In particular, children exploration seems to be driven by intrinsically motivated brain processes that trigger spontaneous exploration for the mere purpose of experiencing novelty, surprise or learning progress (Gopnik et al., 1999; Kaplan & Oudeyer, 2007; Kidd & Hayden, 2015). During exploratory play, children can also invent and pursue their own problems (Chu & Schulz, 2020).

Contemporary theory in cognitive psychology maintains that curiosity is a crucial developmental element that allows humans to acquire useful knowledge (Oudeyer & Smith, 2016; Gopnik, 2020). Curiosity lets us seek out certain experiences for the inherent value of learning from them, not because they are rewarding in a more tangible way. The feeling of curiosity itself is borne out by a cognitive mechanism that prioritizes potential learning experiences in order to prepare the learner for future challenges that are as yet unknown (Oudeyer, 2018; Gopnik, 2020).

3.2.1.2 In artificial intelligence

Intrinsically motivated learning. Precisely how the curiosity-inducing mechanism operates in biological organisms is poorly understood (Gottlieb & Oudeyer, 2018). However, the importance of curiosity-driven learning is increasingly recognized in AI, as a potential solution for building autonomous machines that can discover and learn an open-ended repertoire of skills.

Algorithmic models of intrinsic motivation were successfully used in developmental robotics (Oudeyer et al., 2007; Baldassarre & Mirolli, 2013), in reinforcement learning (Singh et al., 2004; Schmidhuber, 2010) and more recently in deep reinforcement learning (DRL) (Bellemare et al., 2016; Pathak et al., 2017). Intrinsically motivated goal exploration processs (IMGEPs), in particular, enable agents to sample and pursue their own goals without external rewards (Baranes & Oudeyer, 2013; Forestier & Oudeyer, 2016; Forestier et al., 2017) and can be formulated within the DRL framework (Held et al., 2017; Nair et al., 2018; Colas et al., 2019; Pong et al., 2020; Venkattaramanujam et al., 2019; Racaniere et al., 2019). Intrinsically motivated agents that learn new skills by generating and pursuing their own goals are referred as *autotelic* agents, from the Greek *auto* (self) and *telos* (goal) (Steels, 2004).

Quality-Diversity algorithms. In a parallel line of research, the recent introduction of quality-diversity algorithms (Lehman & Stanley, 2011; Cully et al., 2015; Pugh et al., 2016) has signified a departure from a purely performance-based view of AI and has renewed interest in mechanisms related to the preservation of diversity arising in natural evolution. Quality-Diversity algorithms employ populations of agents that find a set of diverse and well-performing solutions, rather than a single one. This has been proven vital in real-world settings, such as robots that need to recover from damages (Cully et al., 2015).

3.2.2 Approaches focusing on environmental diversity

3.2.2.1 In the human species

Human behavioral ecology. Human behavioral ecology (HBE) is a field studying how ecological constraints and opportunities could have shaped human behavior throughout its evolutionary history (Maslin et al., 2015; Brown et al., 2011; Sear et al., 2007). The spotlight is on the Rift Valley at East Africa during a period that approximately lasted from 7 to 2 million years ago and is hypothesized to have constituted a turning point in our evolutionary trajectory. This is the time when hominin diversity reached its highest level (with the first specimens attributed to genus Homo), with evidence of episodic migration of hominins out of the Rift Valley and into Eurasia, as well as the most dramatic increases in hominin brain size (Maslin et al., 2015). Data from paleoclimatology suggests that a main feature of that period is extreme climatic variability, characterized by the presence of lakes that could expand quickly and contract in a highly chaotic way. The interesting end result from a behavioral perspective, is that the presence of these lakes creates niches of wide diversity that vary quickly with time, presumably within a few dozen generations. This extreme climate variability could have had a strong influence on early human evolution, in particular regarding our ability to quickly generalize and extend our behavioral repertoire to changing environments. Due to partial observability of the environmental conditions and an even more partial understanding of how they affect the dynamics of populations, research in HBE have multiple hypotheses on how the environmental dynamics affected the species at that time and place. This evolutionary leap was originally studied under theories that lay emphasis on specific environmental changes. The Savannah hypothesis, for example, suggests that the change in fauna favored bipedal walking, which enabled migration and the creation of new niches for humans. Later hypotheses under the Pulsed Climate Variability (PCV) framework , however, suggest that the key change was instead the general environmental complexity characterizing that period (Potts, 2013; Maslin et al., 2015; Collard et al., 2005). The PCV framework was recently introduced in an attempt to juxtapose different mechanisms under which environmental conditions may have affected speciation and extinction rates, which further influence the population's diversity. This framework constitutes an important step towards a unified model of the ecological and evolutionary dynamics of that period.

Niche construction theories. Niche construction, the process with which species shape their environment to increase their control over it or reduce its uncertainty (Krakauer et al., 2009; Constant et al., 2018), has been termed "*the neglected process in evolution*" (Odling-Smee et al. (2003)), as natural life research previously disregarded it as a minor mechanism driving evolution, compared to natural selection. The contemporary opinion has however shifted towards viewing niche construction as an important evolutionary mechanism that accompanies natural selection. In the introduction of their book, Odling-Smee et al. (2003) explain:

> *Organisms play two roles in evolution. The first consists of carrying genes; organisms survive and reproduce according to chance and natural selection pressures in their environments. This role is the basis for most evolutionary theory, it has been subject to intense qualitative and quantitative investigation, and it is reasonably well understood. However, organisms also interact with environments, take energy and resources from environments, make micro- and macrohabitat choices with respect to environments, construct artifacts, emit detritus and die in environments, and by doing all these things, modify at least some of the natural selection pressures present in their own, and in each other's, local environments. This second role for phenotypes in evolution is not well described or well understood by evolutionary biologists and has not been subject to a great deal of investigation. We call it "niche construction" (Odling-Smee 1988) and it is the subject of this book. (Odling-Smee et al. (2003))*

This original concept of ecological niche construction has then been extended to describe more specific mechanisms, including cognitive (Clark, 2005), social (Bergmüller & Taborsky, 2010; Saltz et al., 2016) and cultural niche construction (Laland et al., 2001; Laland & O'Brien, 2011).

Cognitive niche construction, which is defined "*as the process by which animals build physical structures that transform problem spaces in ways that aid (or sometimes impede) thinking and reasoning about some target domain or domains*" (Clark (2005)).

Social niche construction is the mechanism in which individuals, singly or collectively, influence the composition and dynamics of their social environments (Flack et al., 2006; Saltz et al., 2016). Saltz et al. (2016) propose that "*social niche construction has occurred whenever an individual, through its behavior or other traits, modifies its social environment, and the modified social environment is within the focal individual's social niche.*"

Finally, another type of niche construction, which is particularly important in the human species, is cultural niche construction. According to Laland et al. (2001):

> *Human innovation and technology has had an enormous impact on the environment: it has made many new resources available via both agriculture and industry; it has influenced human population size and structure via hygiene, medicine and birth control; it is drastically reducing bio-diversity; and it may already have resulted in the degradation of large areas of our global environment. These are all potential sources of modified natural selection pressures. Thus cultural processes that precipitate niche construction might be expected to have played a critical role in human evolution for many thousands, perhaps millions of years. (Laland et al. (2001))*

3.2.2.2 In artificial intelligence

Environment design. Progressing in the design of efficient learning architectures is often associated with the proposition of more complex environments to evaluate them. Reinforcement learning (RL) environments consisting of a large collection of retro video games (Arcade Learning Environment, Bellemare et al., 2013) have been widely used since the rise of deep reinforcement learning (DRL) algorithms (Mnih et al., 2015). Challenging board games with very large search spaces and requiring long-horizon planning have served as a key challenge driving the integration of DRL, Monte-Carlo tree search and self-play (Silver et al., 2016, 2017). Surpassing human level in multi-player strategy video games has driven the development of large scale multi-agent reinforcement learning (MARL) algorithms (Vinyals et al., 2019; OpenAI, 2018). Procedurally generated environments (Risi & Togelius, 2020) are used to benchmark the ability of (meta-)DRL algorithms to generalize their learned policy to novel tasks that were not experienced during training. Among the available simulation environments, the Minecraft video game is considered has one of the most challenging for the AI community (Johnson et al., 2016; Fan et al., 2022). It consists in a procedurally generated 3D world with virtually infinite terrain where the player may discover and extract raw materials, craft tools and items, and build structures, earthworks, and simple machines (Wikipedia, 2022). I will refer to the possibility to produce new elements by composing other elements as a an *environment with compositional dynamics* (as in Minecraft but also others, see e.g. Wang et al., 2021; Platanios et al., 2020; Garcia Ortiz et al., 2021), offering opportunities for e.g. tool crafting (Figure 5.13). Interestingly, Minecraft is relevant to evaluate both cognition-centric and emergentist approaches (Section 3.1.1). On the one hand, it allows defining a large repertoire of challenging tasks on which to benchmark learning architectures (in the spirit of the cognition-centric approach). On the other hand, it is perhaps the most open-ended environment available to the community, in the sense that the player can explore these large and complex worlds without any externally provided goals (i.e. in a fully autotelic fashion, see Section 3.2.1.2). Minecraft therefore appears as a particularly relevant environment for the emergentist approach, which considers intelligence as inextricably linked to environmental dynamics.

Meta reinforcement learning. Meta reinforcement learning (META-RL) aims at equipping agents with the ability to generalize to tasks or environments that have not been encountered during training. Two nested processes of adaptation are traditionally considered: the inner level is a standard DRL algorithm operating on a given environment. The outer level is tuning the (hyper-)parameters of the inner loop such that it performs well on a distribution of environments. The end result of this nested training process is an algorithm that *learns how to learn*, i.e. learns how to adapt to unseen tasks. Works differ in the optimization technique and adaptation mechanism they consider in the two loops. The outer loop often optimizes higher-level structures such as the architecture of a neural network (Baker et al., 2017), the morphology (Gupta et al., 2021), a curiosity module providing intrinsic rewards (Alet et al., 2020) or plasticity rules (Najarro & Risi, 2021), employing either evolutionary or gradient-based optimization. Impressively, the meta-learning paradigm has managed to learn algorithms from scratch (Oh et al., 2021; Kirsch & Schmidhuber, 2021), showcasing that human-engineered solutions often stumble upon solutions found by meta-learning.

Curriculum learning. Curriculum learning refers to a family of methods aiming at improving learning agent performance by carefully choosing the task sequence the agent is exposed to. It is inspired by curriculum learning in humans and in particular in education, where children are first exposed to easy task (e.g. learning how to count integers) then progressively to more and more difficult tasks (e.g. learning additions, then multiplication, then division). The task sequence is supposed to follow a logical order taking into account dependences across the different tasks, e.g. learning how to multiply integers first requires mastering how to perform additions. This task sequence can be either predefined or adapted dynamically according to the agent's learning dynamics (see Portelas et al., 2020, for a recent review focusing on DRL agents). While meta-learning approaches sample environments from a distribution and therefore have no control over the order in which different environments are presented to the agent, curriculum learning instead focuses on the ordering in which tasks are presented to the agent. Note however that curriculum learning does not necessarily act on environments, they can also control goals or reward functions (as in intrinsically motivated approaches presented above) or the selection of opponents in multi-agent settings (see below). See again Portelas et al. (2020) for a complete taxonomy of curriculum learning approaches.

3.2.3 Approaches focusing on social diversity

3.2.3.1 In the human species

If cooperation requires that an agent pays a reproductive cost for someone else's benefit, how can cooperation emerge in a population of agents evolving selfishly? Under the *big mistake hypothesis* (Burnham & Johnson, 2005), altruism emerged in small-scale groups due to kin selection or reciprocity. In contrast, the interdependence hypothesis (Tomasello et al., 2012) proposes a theory for the emergence of cooperation that replaces altruism with mutualistic collaboration. According to it, the need for foraging led to the selective helping of those who were needed as collaborative partners in the future. In sufficiently small groups, social selection was performed based on reputation. The size and structure of groups was dynamically shaped by their need to maintain stability and defend themselves against other groups.

Competition pressures can result in arms race dynamics where different groups or species continuously co-adapt to each other, potentially leading to increased behavioral complexity (the Red Queen hypothesis, Van Valen, 1973, see also Maslin et al. (2015) for its interpretation in the context of hominin evolution). Competition between co-existing groups can also result in an increase in socio-cultural diversity through the mechanism of schismogenesis (Bateson, 1935), i.e. the tendency to define oneself by opposition to one another in a self-amplifying process of divergence (see also Graeber & Wengrow, 2021, for the relevance of this concept in anthropology).

Several theories of the origins of human language postulate a joint evolution of cooperative and communicative behaviors (Smith, 2010; Gärdenfors, 2002; Ghazanfar & Takahashi, 2014; Tomasello et al., 2012). It is in particular the central thesis of the theory developed by Michael Tomasello, who proposes that "*humans' species-unique forms of cooperation –as well as their species-unique forms of cognition, communication, and social life—all derive from mutualistic collaboration (with social selection against cheaters)"* (Tomasello et al. (2012)). In this view, it is the constraints imposed by the social niche occupied by human beings that has driven them to jointly develop complex collaborative and communicative behaviors, in a context of interdependence requiring the sharing of intentions. We also find compatible arguments in the mirror system hypothesis developed by Michael Arbib (Arbib, 2005) proposing that language evolution is grounded in the sensory-motor integration required for the execution and the observation of transitive actions towards objects, enabling other's intention recognition and providing the bases of a syntactic structure (Roy & Arbib, 2005) (see also Iriki & Taoka, 2012, for theoretical propositions on the coevolution of tool use and language in humans). Finally, the social complexity hypothesis suggests that groups with complex social structures require more complex communication systems to regulate interactions between group members (Freeberg et al., 2012).

Social behaviors are also shaped through the mechanism of cultural transmission, where behaviors of an individual arises through induction on the basis of observations of behavior in another individual. The compositionality of language could have emerged not only for its expressive power, but also as a way to facilitate its learning by new infants in a population: a mechanism called iterated learning (Kirby et al., 2014) which has been studied both in human experiments (Tessler et al., 2021) and AI (Ren et al., 2020; Li & Bowling, 2019).

In his seminal work, the developmentalist Led Vygotsky promotes a social conception of intelligence. He first proposed that linguistic social interactions such as descriptions, explanations, corrections, or play start as interpersonal mechanisms before they are turned into *intrapersonal* cognitive mechanisms through the process of *internalization* (Vygotsky, 1930, 1933, 1934). Following his vision, many psychologists (Berk, 1994; Lupyan, 2012; Gentner & Hoyos, 2017), linguists (Whorf, 1956; Rumelhart et al., 1986; Lakoff & Johnson, 2008) and philosophers (Hesse, 1988; Dennett, 1993; Carruthers, 2002) argued for the importance of socio-cultural interactions in the development of human intelligence (see Colas et al., 2022a, for our recent roadmap on how to introduce Vygotskian mechanisms in AI agents).

3.2.3.2 In artificial intelligence

Works in multi-agent reinforcement learning (MARL) have shown that the mere presence of multiple co-existing agents can give rise to open-ended learning dynamics (Leibo et al., 2019). In addition to self-play originally used in two-player problem settings (Silver et al., 2016), the presence of multiple agents can give rise to an arms race, where competing agents find solutions that act as new challenges for their opponents (a phenomenon called *autocurricula* in multi-agent AI, Leibo et al., 2019; Bansal et al., 2018; Baker et al., 2020). Notable examples of emergent social behavior in MARL include the emergence of cooperation in social dilemmas (Leibo et al., 2017; Pérolat et al., 2017) communication systems for solving referential games or sequential MARL tasks (Lazaridou & Baroni, 2020; Mordatch & Abbeel, 2017), tool use in competing groups of DRL agents playing Hide and Seek (Baker et al., 2020), social norms such as road traffic rules (Pal et al., 2020) and silly rules (Köster et al., 2022), reputation mechanisms (Anastassacos et al., 2021), social influence as a form of intrinsic motivation (Jaques et al., 2019) and cultural transmission (Team et al., 2022).

3.2.4 The need for an integrative framework

This short review reveals interesting relationships between existing works studying OESA in the human species and in AI, that we will further analyze below. Conceptualizing these relationships can help to initiate an interdisciplinary dialog and enable the cross-fertilization and transfer of ideas between the two fields. For this aim, I propose in the next section an integrated conceptual framework modeling important components of OESA and their interactions at multiple spatiotemporal scales. I will then explain in detail their underlying dynamics.

3.3 The ORIGINS framework

3.3.1 Motivation: Grounding artificial intelligence in the origins of human behavior

We can extract important relationships between open-ended skill acquisition in AI and the human species from the previous section.

There is a diversity of approaches studying OESA in both AI and in the human species. I have proposed above to organize these approaches in three categories: focusing on behavioral, environmental and socio-cultural diversity. These are however just focuses, as most of these contributions actually consider the interaction between several mechanisms. For instance, both META-RL and HBE studies how environmental diversity can encourage behavioral diversity by improving the generalization ability of acquired behaviors to novel environments.

Intrinsic motivation is considered as a key mechanism of OESA in both AI and in the human species (Section 3.2.1). However, the evolutionary basis of intrinsic motivation is overlooked in both fields. Interestingly, most contributions studying this questions come from the AI community. Barto (2013) proposes that the evolutionary origins of intrinsic motivation can be explained because it maximizes long-term evolutionary fitness under changing environmental conditions, "*promot[ing] behavior that is ubiquitously useful across many different environments*" (Barto (2013)). This hypothesis is supported by computational RL experiments in Singh et al. (2010). With Ten et al. (2022), we build upon this framework to extract key functional aspects of intrinsic motivation in both biological and artificial agents: procurement of extrinsic primary rewards, learning internal models, goal discovery and cultural innovation.

Adaptation occurs at different timescales, each one employing a mechanism of distinct nature. On the evolutionary time scale, populations of agents adapt through natural selection (Darwin, 1859), with periods of environmental instability resulting in speciation and extinction events (Potts, 2013). Within their lifetime, individuals employ learning mechanisms of evolutionary origins hypothesized to be favored when the environment is unpredictable between generations but predictable enough within a lifetime for exploratory behavior to offer an evolutionary advantage (Niv et al.; Stephens, 1991; Johnston, 1982) . In the AI community, learning initially acquired a broad, mechanistic definition that encompassed all processes with which an agent improves its performance on a given task based on its experience (Mitchell, 1997). However, the polyphony of mechanisms employed to achieve learning in the sub-fields of supervised, unsupervised and reinforcement learning, suggests that learning mechanisms are associated with opportunities and compromises. The recent introduction of META-RL resonated with the needs of the community for agents with stronger generalization abilities (Vanschoren, 2018). META-RL is achieved through a bi-level optimization process: in the inner loop, a standard learning algorithm is optimized on a given task, while the outer tunes the algorithm in the inner loop for improving its performance on a wide distribution of tasks. This resembles the nested adaptive mechanisms of evolution (outer loop) and development (inner loop) in biological organisms. The end result of this nested training process is an algorithm that *learns how to learn*, i.e. learns how to solve unseen tasks faster.

It is recognized that environmental variability promotes adaptability in both AI and in the human species. Contributions in META-RL study how training an agent on a distribution of environments can improve its generalization abilities on unseen test environments, while curriculum learning instead focus on the order in which tasks are presented to the agent (Section 3.2.2.2). Under the PCV framework discussed in Section 3.2.2.1, climate variability has served as a drive for the ability of humans to adapt to complex and rapidly changing environments, with a particular focus on the alternation between stable periods with either low or high resource availability and periods with more chaotic climate variability. Adaptation is achieved through mechanisms whose form partially depends on properties of the environment. If the environment is constant across time and space, natural selection may favor innate behaviors. By contrast, if the environment varies, natural selection might favor behavioral plasticity (Stephens, 1991; Johnston, 1982; Hougen & Shah, 2019; Eskridge & Hougen, 2012; Frankenhuis et al., 2019).

Socio-cultural dynamics is an important mechanism driving OESA. Many theories on the evolution and acquisition of complex social behavior in the human species –including language– emphasize a strong role of the pre-existing ecological and social conditions in the environment of our distant ancestors (Section 3.2.3.1). These conditions shape the cooperative and competitive pressures driving the acquisition of complex social behavior, which in turn modulate those pressures. These arguments resonate with results in multi-agent learning from the AI community –in particular recently in MARL– where properties of the physical and social environment are the main driver for the joint learning of cooperative, communicative and normative social behavior (Section 3.2.3.2).

Niche construction is a neglected process in both fields, and especially in AI. Niche construction has been in the past termed as the neglected process in evolution (Odling-Smee et al., 2003). It is however now admitted that agents modify their own environments and therefore their own evolutionary fitness, which are then inherited by their descendants and can have a significant impact on the course of evolution (Section 3.2.2.1). Niche construction increases the control that agents have over their environments, examples being tool use, agriculture and nest building. Whereas niche construction is recognized as a key driving force of evolution, in particular in the human species (Kendal et al., 2011), it is still widely ignored in AI research (see however Clune, 2020, for recent propositions).

This analysis shows that open-ended skill acquisition is not a mere property of some advanced cognitive architectures. Instead, I propose in this thesis to conceive OESA as an emergent property resulting from the coupling of heterogeneous mechanisms operating at multiple spatiotemporal scales, namely environmental, evolutionary, morphological, sensorimotor, developmental, cognitive, social and cultural mechanisms.

3.3.2 Overview of the proposed framework

3.3.2.1 Components

Informed by the analysis of the relationships between artificial and human OESA proposed above, we can extract key components driving open-ended skill acquisition in biological and artificial agents. These components are represented as boxes in Figure 3.2 and are explained below. In the text, I will refer to them in bold, e.g. **environmental complexity**. Each component can also be referred in abbreviated form indicated in parentheses in each paragraph below (e.g. EC for **environmental complexity**). Each component contains items, that I indicate in italics in the text, e.g. *resource availability* is an item of the **environmental complexity** component (see Figure 3.2).

Environmental complexity (EC). The environment exhibits multiscale and compositional dynamics (Section 3.2.2). *Multiscale dynamics* refers to alternating cycles at various spatiotemporal scales (day/night, seasonal cycles, climate variations). These cycles modulate *constraints and opportunities*, e.g. in terms of the *resources* available to the agents, as well as other environmental properties affecting their fitness (e.g. *exposition to predator*). *Compositional dynamics* (Section 3.2.2.2) refers to the possibility to produce new elements in the environment, or to change the properties of existing ones, by composing other elements (offering opportunities for e.g. tool crafting, see **cultural repertoire** below). The environment contains other agents with possibly converging or conflicting interests (see **multi-agent dynamics** below).

Adaptability (A). The agents adapt to their environment on two main scales: evolutionary and developmental. At the *evolutionary scale*, natural selection shapes the morphological, physiological, sensorimotor and cognitive apparatuses from generations to generations according to environmental selective pressures, driving speciation and extinction events (Section 3.2.2.1). At the *developmental scale*, morphogenesis shapes the organism and its interaction with the environment. This results in built-in self-regulatory sensorimotor loops for anticipating needs and preparing to satisfy them before they arise (e.g. allostasis Sterling, 2012), acting as behavioral priors bootstrapping higher-level cognitive mechanisms (Verschure et al., 2014) such as *learning* and *exploration* (Section 3.2.1). Adaptability is thus structured in two nested mechanisms: an outer evolutionary loop optimizing fitness at the generational scale by selecting for morphological, sensorimotor and cognitive priors shaping the dynamics of an inner developmental loop at the individual scale.

Multi-agent dynamics (MD). Multiple agents interact in a shared environment. Environment properties induce a *game-theoretic context* encouraging cooperation or competition among agent populations co-adapting at the evolutionary and developmental scales. Multi-agent *interactions* are shaped by this context, from simple *signalling* or *manipulation* to the self-organization of more complex social behavior (Section 3.2.3, see also **cultural repertoire** below). The game-theoretic context is however perturbed by environmental multiscale dynamics (see **environmental complexity**), driving the agents to continuously readapt to changing conditions.

Cultural repertoire (CR). Multi-agent dynamics in populations of adaptive agents can potentially lead to the formation of a cultural repertoire (Section 3.2.3) through pressures and opportunities to acquire more complex skills such as *technology* (e.g. tool use), *communication* (e.g. language) and *social organization* (e.g. social norms). Although there is no strong consensus on the definition of *culture*, a standard one refers to a behavior that is transmitted repeatedly through social or observational learning to become a population-level characteristic (Gruber et al., 2010). The behaviors acquired in a generation influence the experience of the next one through cultural transmission (i.e. the process by which cultural behaviors are transmitted), potentially bootstrapping cultural evolution (i.e. the cumulative change of culture over time, Richerson & Boyd, 2008).

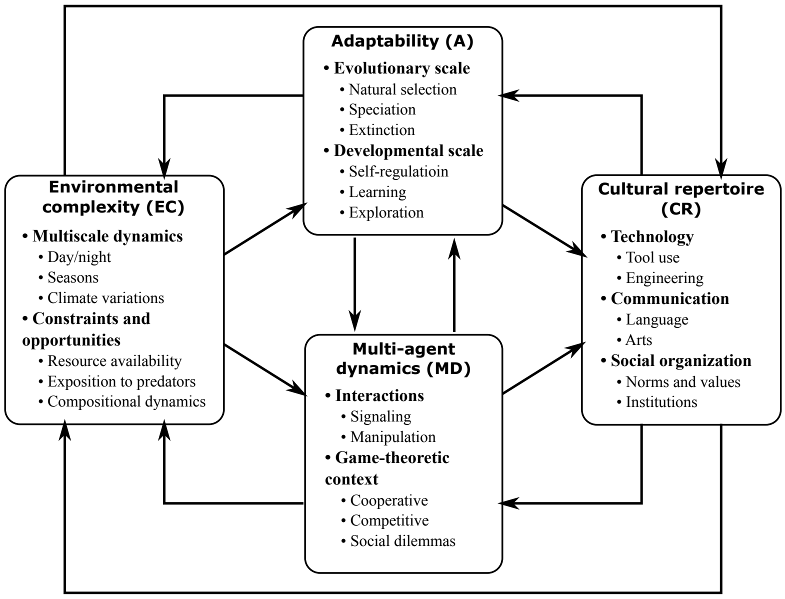


Figure 3.2: The proposed ORIGINS framework aims at *Grounding Artificial Intelligence in the Origins of Human Behavior*. It identifies central components (boxes) and their interactions (arrows) driving open-ended skill acquisition, both in terms of its evolution from environmental complexity (roughly: left to right arrows) as well its open-ended aspect through feedback mechanisms (right to left arrows). The employed terminology reflects a diversity of mechanisms considered in both AI and human behavioral ecology. See text for details.

3.3.2.2 Interactions between the components

In Figure 3.2, we introduce a conceptual framework, called ORIGINS, that recognizes important ecological components as well as the feedforward and feedback links driving the whole system dynamics. I provide below a summarized overview of the main aspects of this dynamics. In the next sections, I will explain in more detail the different subparts of the diagram. As mentioned above, I indicate the framework components (i.e. the boxes in Figure 3.2) in bold, e.g. **environmental complexity**. Links between the framework components (i.e. arrows in the diagram of Figure 3.2) are referred as X→Y, where X is the source component and Y is the target component, in abbreviated form as indicated above. For example, EC→A indicates the link from **environmental complexity** (EC) to

adaptability (A). Finally, I indicate specific items of a framework component in italics, e.g. *multiscale dynamics* is an item of the **environmental complexity** component.

Under the proposed framework, **environmental complexity** is characterized by *multi-scale dynamics*, which is driven by *daily, seasonal and climate variations* (Section 3.2.2.1). It modulates *constraints and opportunities* available to the agents, in particular through changes in *resource availability*, *exposition to predators* and the items available for *compositional dynamics* (Section 3.2.2). This complexity has a strong influence on two major phenomena. First, it drives **adaptability** (EC→A) both at the *evolutionary scale*, through *natural selection*, *speciation* and *extinction*, and at the *developmental scale* by shaping cognitive mechanisms for *learning* and *exploration* (Section 3.3.1). Extending existing theories in HBE and informed by recent advances in AI presented in previous sections, a key hypothesis of the ORIGINS framework is that there exist certain levels of **environmental complexity** able to drive the *evolution* of autotelic exploration (Section 3.2.1.2), i.e. of an intrinsic motivation to discover new goals and learn how to achieve them (see Section 3.2.1 above about intrinsic motivation, as well as Section 3.4 below for more detail). Second, varying the levels of *resource availability* and *exposition to predators* has a strong influence on **multi-agent dynamics** (EC→MD) through the modulation of the *game-theoretic context*, i.e. *cooperation* and *competition* pressures (Section 3.2.3).

The influence of **environmental complexity** on **adaptability** and **multi-agent dynamics** can then have feedback and feedforward effects on the whole system dynamics. Firstly, increased morphological and cognitive complexity resulting from **adaptability**, as well as increased complexity in the **multi-agent dynamics**, feed back to **environmental complexity** (A→EC and MD→EC) through mechanisms of niche construction modifying *resource availability* and *predation pressure* (Section 3.2.2). Competition among different groups or species is also considered as an important driver of adaptation, possibly resulting in an arms race between co-adapting species (A→MD and MD→A, see also Section 3.2.3). Secondly, evolved exploration strategies in **adaptability** and complex **multi-agent dynamics** can bootstrap in a feedforward manner the formation of a **cultural repertoire** (A→CR and MD→CR), i.e. of socially transmitted skills. These skills, first of relatively low complexity (e.g. *tool use*), then evolve into increasingly complex *technology* (e.g. *engineering*), *communication* (e.g. *language*) and *social organization* (e.g. social *norms and values* and *institutions*) (Section 3.2.3). Such cultural evolution is made possible through feedback effects from the **cultural repertoire** to all other components of the system, e.g. *technology* modifying the structure of the environment (CR→EC), using *language* as a cognitive tool (CR→A, coresponding to the social conception of intelligence mentioned in Section 3.2.3.1), or *social organization* promoting *cooperation* or *competition* (CR→MD). Cultural evolution thus induces a positive feedback loop driving the ever-expanding complexity of the human cultural repertoire (see Section 3.2.2 above about niche construction, as well as Section 3.6 below for more detail).

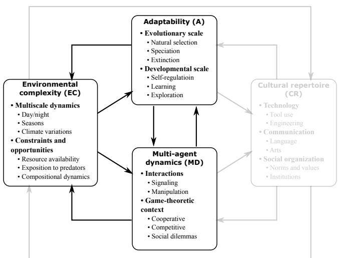
I believe this framework to be of interest for three main reasons. First, it highlights the links between two different domains that usually do not interact together: namely artificial intelligence and human behavioral ecology. Second, it can help to understand how processes operating at multiple spatiotemporal scales can support open-ended skill acquisition (both its origins expressed as feedforward links, and its open-ended aspect expressed as feedback links). Third, it provides a tentative roadmap for achieving human-like open-ended skill acquisition in artificial systems, grounded in theories and hypotheses from HBE.

As an interesting example of how the ORIGINS framework is able to express complex mechanisms driving OESA, we can analyze how it is able to represent each type of niche construction mechanism presented in Section 3.2.2.1. In the proposed framework, niche construction corresponds to feedback effects from any component of the system to **environmental complexity** (i.e. A→EC, MD→EC and CR→EC). A→EC corresponds to the original concept of ecological niche construction as well as to cognitive niche construction where adaptive agents modify the properties of their environment, potentially modifying their own fitness landscape at the *evolutionary* scale or aiding their own behavior at the *developmental* scale. MD→EC corresponds to social niche construction, where **multi-agent dynamics** can modulate *resource availability* and *exposition to predators* in the environment. As mentioned by Saltz et al. (2016): "*Simply joining a group increases its size, illustrating how individuals are closely intertwined with their social environments"*. In turn, increasing group size reduces both resource availability and exposition to predators (Brown, 1982). Finally, CR→EC corresponds to cultural niche construction, i.e. how *technology*, *communication* and *social organization* modify the structure of the environment and the fitness landscape of new generations. Cultural niche construction is considered as a central mechanism driving human-like OESA (Laland et al., 2001; Laland & O'Brien, 2011; Fogarty & Creanza, 2017; Iriki & Taoka, 2012).

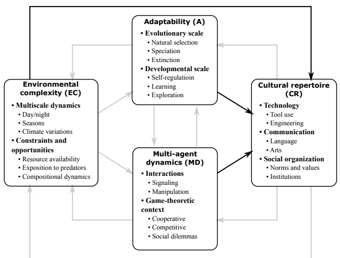
The proposed framework thus highlights the importance of niche construction mechanisms, as feedback effects on **environmental complexity**, in human-like open-ended skill acquisition (OESA). By inducing positive feedback loops that continuously increase **environmental complexity**, which in turn encourage the acquisition of increasingly complex skills, these different types of niche construction provide the driving force that makes human skill acquisition truly open-ended. I believe these mechanisms to be of central importance in order to achieve OESA in artificial systems.

The next three sections explain in more detail the underlying dynamics of subparts of the ORIGINS framework and how it relates to specific hypotheses in both HBE and AI. I propose to distinguish three levels of dynamics (Figure 3.3).

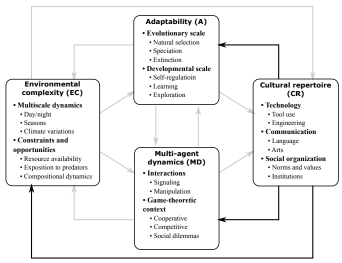
- Level-1 dynamics (Section 3.4) focuses on the interaction between **environmental complexity**, **adaptability** and **multi-agent dynamics** and is concerned with the research question:
  - What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?
- Level-2 dynamics (Section 3.5) focuses on the feedforward effects from **environmental complexity**, **adaptability** and **multi-agent dynamics** to the **cultural repertoire** and is concerned with the research question:
  - What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire?
- Level-3 dynamics (Section 3.6) focuses on feedback effects from the **cultural repertoire** to all the other components of the system and is concerned with the research question:
  - What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition?



(a) **Level-1 dynamics**: What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies? .



(b) **Level-2 dynamics**: What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire? .



(c) **Level-3 dynamics**: What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition? .

Figure 3.3: The ORIGINS framework proposes to distinguish three levels of dynamics. Each level focuses on specific interactions between the framework components and is associated with a main research question. Same conventions as in Figure 3.2.

3.4 Level-1 dynamics: the eco-evolutionary origins of autotelic agents

Level-1 dynamics concerns the interaction between **environmental complexity**, **adaptability** and **multi-agent dynamics**. These three components of the framework reflect the categorization I proposed in Section 3.2 for the existing approaches to open-ended skill acquisition (OESA), focusing on environmental, behavioral and social diversity, respectively. However, the proposed framework emphasizes the emergent effects resulting from the interaction between these three components (i.e. EC→A, EC→MD, MD→A, A→MD, A→EC and MD→EC).

3.4.1 Interactions between environmental complexity and adaptability (EC→A and A→EC)

Under the Pulsed Climate Variability (PCV) framework mentioned in Section 3.2.2, environmental factors such as *climate variability*, *resource availability* and *exposition to predators* have served as a drive for the ability of humans to adapt to complex environments:

> *A new consensus is emerging that suggests the unusual geology and climate of East Africa created a complex, environmentally very variable setting. This new understanding of East African climate has led to the pulsed climate variability hypothesis that suggests the long-term drying trend in East Africa was punctuated by episodes of short alternating periods of extreme humidity and aridity which may have driven hominin speciation, encephalization and dispersals out of Africa (Maslin et al. (2015))*

The role of such complex environmental dynamics is also suggested in Potts (2013), who presents a synthesis of paleoclimate and human evolution data over the past 7 million years suggesting that the adaptations that unpin much of human environmental interactions reflect past evolutionary responses to ecological instability and environmental novelty.

Adaptation to environmental complexity operates on two scales: the evolutionary one drives *natural selection*, *speciation* and *extinction* according to the environment conditions, shaping the developmental one which drives *learning* and *exploration* to adapt to those conditions. For example, during dry periods with strong limitations on *resource availability*, specialist species would struggle having lost their environmental niche and their competitive advantage (Brown et al., 2011) while generalist species, such as those having evolved more advanced exploration strategies, would as a consequence increase their relative fitness.

In AI, most contributions consist in the proposition of adaptive decision-making algorithms (**adaptability**) allowing sensorimotor agents to learn how to optimally behave in an environment with a predefined dynamics. In HBE, however, the interaction between **environmental complexity** and **adaptability** is recognized to also operate in the reverse direction (i.e. A→EC) through the process of ecological niche construction (see Sections 3.2.2 and 3.3). Although recent theoretical propositions in AI have identified feedback effects from an agent behavior to its environment as an important mechanism (Clune, 2020; Stanley & Lehman, 2015), computational contributions of ecological niche construction mechanisms in AI are mostly nonexistent. Notable exceptions can be found in the study of non-episodic reinforcement learning (RL) (Co-Reyes et al., 2020) and in common pool resource appropriation in multi-agent reinforcement learning (Pérolat et al., 2017) but these contributions do not interpret their results in the light of niche construction theories.

In fact, while the RL approach has recently made impressive progress, in particular due to the rise of deep reinforcement learning (DRL, Mnih et al., 2015), it only covers a very limited part of level-1 dynamics in the proposed framework (i.e. mostly EC→A). There are three main limitations in the RL approach.

First, the dynamics of the environment is mostly influenced by the agent's actions. This comes in contrast to the theories in HBE described above, where environments are supposed to display rich intrinsic dynamics independently of the presence of agents (e.g. complex African paleoclimate was mostly independent of human activity). As seen above, it is this rich environmental dynamics which is supposed to have driven the emergence of complex adaptive mechanisms in humans and in particular advanced exploration strategies such as intrinsically motivated goal exploration.

Second, the standard training paradigm in RL is episodic, i.e. the environment is regularly reset to its initial conditions. While this procedure has the benefit to facilitate training from a machine learning perspective (Pardo et al., 2018), it strongly differs from natural settings where environments are persistent, i.e. where the behavior of agents affects the environment in which the next generations will further evolve and learn (see however Co-Reyes et al., 2020, for a recent attempt at studying ecologically valid non-episodic RL settings). Episodic training in RL prevents the study of both niche construction and eco-evolutionary feedback effects, which require that populations alter their environment and that those changes in the environment influence the subsequent evolution of the population (Post & Palkovacs, 2009). The field of automatic curriculum learning (Portelas et al., 2020) studies such feedback effects but focuses on how to adaptively sample novel environments of increasing complexity, using episodic training, with the explicit objective to improve an agent's learning performance.

Third, the RL paradigm traditionally assumes that rewards are provided by the environment. From a biological perspective, however, the environment does not contain any reward whatsoever. Rewards instead result from the agent's own physiology and *self-regulation* and have emerged from evolution as a way to guide *learning* and *exploration* (see e.g. the metaphor of evolved stick/carrot mechanisms in Sterling, 2012).

Despite these major limitations of the RL paradigm, AI approaches focusing on environmental diversity described in Section 3.2.2.2 (including emergentist approaches described in Section 3.1.1) are relevant to better model the whole complexity of level-1 dynamics. Meta reinforcement learning (META-RL) approaches have shown how an outer optimization loop, resembling an evolutionary mechanism, can meta-learn learning algorithms able to adapt to a distribution of diverse environments. Within the META-RL framework, few contributions have shown how intrinsic rewards can emerge as a way to quickly adapt to changing environments (Singh et al., 2010; Alet et al., 2020). The evolution of autotelic exploration has however, to my knowledge, not yet been studied. In addition, curriculum learning methods have shown how the learning of complex behavior can be facilitated by exposing the agent to an adequate sequence of training environments with increasing difficulty. In both approaches, however, diverse environments are presented to the agent through an external mechanism (sampling from a distribution in META-RL or as a sequence in curriculum learning), with weak feedback effects from **adaptability** to **environmental complexity** (see however automatic curriculum learning as in Wang et al., 2019b, studying interesting feedback effects that are nevertheless not ecologically valid). This is radically different from a natural mechanism, where species explore diverse ecological niches with complex multi-scale dynamics which is intrinsic to a persistent environment and with recognized feedback effects from **adaptability** to **environmental complexity** shaping evolutionary mechanisms.

3.4.2 Interactions with multi-agent dynamics (EC→MD, MD→A, A→MD, MD→EC)

Another important aspect of level-1 dynamics is the presence of multiple agents that jointly adapt in a shared environment, i.e. **multi-agent dynamics**. The interactions between **adaptability** and **multi-agent dynamics** (MD→A and A→MD) depend on the game-theoretic structure of the environment: either cooperative, competitive or mixed (e.g. social dilemma, Leibo et al., 2017; Lanctot et al., 2017). For example, as seen in Section 3.2.3, cooperative environments can favor the learning of a shared communication system as a solution to solve complex cooperative tasks (as studied in emergent communication, a subfield of multi-agent AI). Mixed game-theoretic structures such as social dilemmas can lead to complex learning dynamics where agents first acquire individual skills for their own benefit, then jointly converge toward (potentially costly) cooperation strategies for their common good (Pérolat et al., 2017).

In addition, co-adaptation increases **environmental complexity** by influencing variations in resource availability and exposition to predators (A→EC and MD→EC) that in turn modulates cooperation and competition pressures in agent populations (EC→MD). In competitive environments, this feedback loop results in complex co-adaptation mechanisms potentially leading to arms race dynamics where agents continuously adapt new behaviors and morphologies in reaction to the adaptation of other agents. It has been shown that these co-adaption mechanisms can result in open-ended dynamics in both ecology (where this mechanism is called and *arms race*, Van Valen, 1973; Pearson, 2001) and in AI (where it is called *multi-agent autocurricula*, Leibo et al., 2019; Baker et al., 2020)).

Another important aspect concerns the role of **environmental complexity** in shaping the topology of social networks (EC→MD) and the role of this topology on **adaptability** (MD→A). Complex environmental dynamics results in the appearance and disappearance of ecological barriers inducing spatial structures producing population isolation and vicariance (Shultz & Maslin, 2013; Larrasoaña, 2012) (EC→MD). Human studies in the lab have shown how such a dynamic social network topology, alternating between periods of relative isolation between populations with periods of denser connectivity, can improve collective innovation in complex crafting tasks (Derex & Boyd, 2016) (MD→A, see also Sections 6.4.2.2 and 7.3 for our propositions on this topic).

3.4.3 Summary of level-1 dynamics

The main research question related to level-1 dynamics is: *What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?* The notion of *autotelic agents* refers to agents that autonomously learn new skills by generating and pursuing their own goals (Section 3.2.1.2). In the proposed framework I hypothesize that such exploration strategies evolve from certain conditions on the dynamics of **environmental complexity**. This claim is supported by both recent theories in HBE based on paleoclimatology data as well as computational experiments in AI. In HBE, chaotic paleoclimate dynamics on the African continent is hypothesized to have driven the evolution of humans as a generalist species able to quickly adapt to abrupt environmental variations (Section 3.4.1). This claim mirrors computational experiments in AI, in particular in META-RL, showing how intrinsic motivation mechanisms can emerge as a solution to adapt to changing environments (Hougen & Shah, 2019; Singh et al., 2010; Alet et al., 2020). In addition, **environmental complexity** influences **multi-agent dynamics** through the modulation of cooperation and competition pressures, potentially leading to positive feedback loops such as arms race, as well as dynamic social network topologies favoring innovation (Section 3.4.2). The resulting complex **multi-agent dynamics** increases **environmental complexity** and further drives the evolution of complex exploration strategies (**adaptability**) as a way to explore and construct new niches.

In Chapter 5 I will present my existing contributions on this topic as well as future research directions for addressing the main research question of level-1 dynamics: *What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?*

3.5 Level-2 dynamics: the formation of a cultural repertoire

As seen above, level-1 dynamics results in populations of autotelic agents having evolved an intrinsic motivation to constantly explore new goals and learn how to achieve them, as a solution to quickly adapt to rapidly changing environments. Whereas they originally emerged as a way to prepare agents to abrupt environmental changes, this intrinsic motivation has the side effect of also generating skills that have no direct relationship to evolutionary fitness but could, later on, be recruited by the evolutionary process (Oudeyer & Smith, 2016). These skills depend on the opportunities offered by the environment and are transmitted from generation to generation to form a **cultural repertoire** (EC→CR, A→CR and MD→CR).

Studying a **cultural repertoire** can be divided in three subproblems (Oudeyer, 2006; Moulin-Frier & Oudeyer, 2021). Firstly, the study of the *forms* of a cultural repertoire: in the case of language for instance, this corresponds to studying the phonemic, semantic, syntactic or pragmatic structures constituting it. Secondly, the study of the *formation* of a cultural repertoire, i.e. of the genesis of its forms through sensorimotor, cognitive, environmental, social, cultural or evolutionary mechanisms. Thirdly, the study of its *origins*, i.e. of the biological and environmental conditions that could have bootstrapped the formation process.

Due to **multi-agent dynamics**, culturally transmitted skills are strongly shaped by social *interactions* (Section 3.2.3). While non-human species often exhibit impressive behavioral repertoires (Krams et al., 2012), human ecology is however characterized by a uniquely large behavioral repertoire involving complex *technology* (e.g. engineering), *communication* (e.g. language) and *social organization* (e.g. institutions). This complex cultural ecosystem has led scientists in the search for factors that differentiated us from other species (Tomasello et al., 2012; Tomasello, 2019; Botero et al., 2014). According to the inter-dependence hypothesis, social norms and institutions emerged to counteract the fact that reputation alone could no longer alleviate the problem of free riding in large groups. In addition, the social complexity hypothesis (Krams et al., 2012) states that language worked as a bonding mechanism that replaced grooming, practiced in small-scale societies, and thus helped with maintaining group stability in larger groups (Dunbar, 1993). Tool use has also been investigated under a number of, often contesting, hypotheses. Based on the data analysis in Fogarty & Creanza (2017), environmental variability such as the risk of resource failure, mobility and climate characteristics correlate significantly with tool use in food-gathering societies. However, it is the group size and not these factors that affect tool use in food-producing societies. It is therefore conjectured that the feedback link of societies with a larger cultural repertoire has a stabilizing effect, dampening the forward impact of environmental variability (Collard et al., 2011).

Contributions in multi-agent AI, and notably in multi-agent reinforcement learning (MARL), have also studied the formation of a **cultural repertoire** in populations of adaptive agents from a computational perspective. As mentioned in Section 3.2.3.2, this includes the emergence of cooperation in social dilemmas (Leibo et al., 2017; Pérolat et al., 2017) communication systems for solving referential games or sequential MARL tasks (Lazaridou & Baroni, 2020; Mordatch & Abbeel, 2017), tool use in competing groups of DRL agents playing Hide and Seek (Baker et al., 2020), social norms such as road traffic rules (Pal et al., 2020) and silly rules (Köster et al., 2022), reputation mechanisms (Anastassacos et al., 2021), social influence as a form of intrinsic motivation (Jaques et al., 2019) and cultural transmission (Team et al., 2022).

In Chapter 6 I will present my existing contributions on this topic as well as future research directions for addressing the main research question of level-2 dynamics: *What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire?*

3.6 Level-3 dynamics: towards human-like open-ended skill acquisition through cultural feedback effects

Level-1 and level-2 dynamics results in agent populations forming a repertoires of collective skills with no necessary direct link to evolutionary fitness, called **cultural repertoire** in the ORIGINS framework. Level-3 dynamics focuses on the feedback effects from the **cultural repertoire** to all the other components of the system (i.e. CR→EC, CR→A and CR→MD), driving the open-ended dynamics of human skill acquisition.

While a large part of contributions in modern AI study how skills emerge from environmental pressures, there are surprisingly few works that study how skills affect the environment. This comes in contrast to the increasing importance of niche construction theories in ecology (Section 3.2.2 and 3.3). Adaptive agents change their environments, which are also inherited by their descendants and can have a significant impact on the course of evolution, a key mechanism which is widely ignored in AI research (see Section 3.3.2 on how the different types of niche construction theories are expressed in the ORIGINS framework, in particular cultural niche construction, i.e. CR→EC).

An example of cultural niche construction concerns the potential role of fire control in language evolution. Some contributions studying the discovery of fires by humans propose that having to maintain a fire during the night considerably increased the waking time (about 16 hours in humans, compared to 8 hours in other mammals, Gowlett, 2016). This "extended day time" could have resulted in a novel niche without strong functional pressures, where the campfire protected the group from predators and the absence of natural light prevented any hunter-gatherer activity (CR→EC). This could have played an important role in language evolution, dynamically modulating the structure of our social networks (CR→MD) by providing free time at night that could be dedicated to communication (Wiessner, 2014; Dunbar, 2014).

Another important feedback effect highlighted by the proposed framework is from the **cultural repertoire** to **adaptability** (CR→A). In a recent position paper (Colas et al., 2022a), we propose a roadmap towards *Vygotskian autotelic agents* able to interact with others and, more importantly, able to internalize socio-cultural interactions to transform them into *cognitive tools* supporting the development of new cognitive functions. This social conception of intelligence is grounded in the seminal work of the developmentalist Led Vygotsky, as described in Section 3.2.3.1. In Section 7.2, I will present a computational architecture of a Vygotskian autotelic agent showing how the compositionality of language can be leveraged to imagine creative goals and foster exploration.

In Chapter 7 I will present my existing contributions on this topic as well as future research directions for addressing the main research question of level-3 dynamics: *What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition?*

3.7 General conclusions and discussion

In this chapter I have proposed an integrated conceptual framework, called ORIGINS, highlighting the main mechanisms driving open-ended skill acquisition (OESA) in the human species and in artificial intelligence (AI). The framework is grounded in hypotheses and theories from human behavioral ecology (HBE) and motivated by the observation of strong relationships between HBE and experimental results in AI (Section 3.3.1). In this sense, the framework can be considered as an invitation to dialog between researchers in HBE (and more generally in life science, including evolutionary biology, developmental psychology and niche construction theories) and AI.

Main hypotheses. The proposed framework identifies key components driving OESA, namely **environmental complexity**, **adaptability**, **multi-agent dynamics** and **cultural repertoire** (Section 3.3.2.1). It describes the complex interactions between these components in terms of feedforward and feedback effects at different levels (Section 3.3.2.2).

Level-1 dynamics describes the interaction between **environmental complexity**, **adaptability** and **multi-agent dynamics** (Section 3.4). The main hypothesis at this level is that **environmental complexity** is the main driver of behavioral diversity, as proposed by recent theories in HBE (such as the Pulsed Climate Variability framework) as well AI experiments in meta reinforcement learning and curriculum learning. Feedback effects of co-adaptation also play an important role in behavioral diversity at this level, in line with the concepts of arms race dynamics in behavioral ecology (Red Queen hypothesis) and of multi-agent autocurriculum in AI. Environmental complexity and arms race dynamics both contribute to the evolution of autotelic agents equipped with an intrinsic motivation to discover and explore new niches and skills as a solution to adapt to abrupt environmental changes. Level-2 dynamics describes the feedforward effects from **environmental complexity**, **adaptability** and **multi-agent dynamics** to the formation of a **cultural repertoire** (Section 3.5). Intrinsically motivated goal exploration coupled with varying cooperation and competition pressures, emerged from level-1 dynamics, result in the acquisition of complex skills that are not necessarily linked to evolutionary fitness. These contribute to the formation of a cultural repertoire through level-2 dynamics, i.e. an expanding collection of behaviors that are transmitted repeatedly through social or observational learning to become a population-level characteristic. These skills include *technology*, *communication* and *social organization*, originally of relatively low complexity. Finally, level-3 dynamics describes how a **cultural repertoire** influences all the other components of the system, resulting in increased complexity of the environment, of the agents' cognitive abilities and of their multi-agent interactions. This results in positive feedback loops continuously feeding the dynamics of the three proposed levels. The main hypothesis here is that these positive feedback loops, which are especially prominent in the human species due to the complexity of our cultural repertoire, are what make human skill acquisition truly open-ended.

The ORIGINS framework as a roadmap. Thus, the framework can be considered as a roadmap for future research in AI. The underlying methodology requires AI researchers to dive into the HBE literature and to engage in an interdisciplinary dialog with this field. In AI, the objective is to extract important hypotheses and mechanisms that can potentially advance the state of the art in artificial OESA and be tested in computational experiments. This effort can in turn benefit research in HBE by proposing computational tools to test their hypotheses in simulated environments (as already proposed to some extent in Frankenhuis et al., 2019). In Part II I will present my existing contributions and will show how the ORIGINS framework allows the proposition of concrete and original future experiments, at each of the three levels.

Limitations. As with any conceptual model, the framework relies on several simplifications. In particular, each component contains different mechanisms operating at several spatiotemporal scales. **Environmental complexity** involves multiscale dynamics at the scale of days, seasons and climate variations. **Adaptability** and **multi-agent dynamics** operate at both the developmental at evolutionary scales. Finally, while behaviors in the **cultural repertoire** can be formed at the level of a generation, their effects on other components of the system are supposed to span several generations. This abstraction of spatiotemporal scales in the framework can limit the representation of important interactions, e.g. between evolutionary and developmental processes. The underlying design choice is that a component of the framework represents a certain type of functional process, e.g. **adaptability** represents mechanisms adapting phenotypes to environmental conditions at multiple spatiotemporal scales. Distinguishing these different time scales and their interactions can however be realized when implementing parts of the framework in computational architectures, as we will see in Part II.

Related conceptual frameworks. The ORIGINS framework displays strong relationships with other theoretical frameworks in both biology and in AI. In biology, the *extended evolutionary synthesis* (Pigliucci & Muller, 2010; Laland et al., 2015) considers that, in the standard evolution theory as well as its modern synthesis, "*too much causal significance is afforded to genes and selection, and not enough to the developmental processes that create novel variants, contribute to heredity, generate adaptive fit, and thereby direct the course of evolution*" (Laland et al. (2015)). It recognizes important feedback effects in terms of *constructive development*, i.e. the ability of an organism to shape its own developmental trajectory by constantly responding to, and altering, internal and external states, as well as of *reciprocal causation*, i.e. that developing organisms are not solely products, but are also causes, of evolution. Such feedback effects are also central in the proposed ORIGINS framework which can be considered as a call for an extended synthesis in AI focusing on open-ended skill acquisition. On the AI side, the ORIGINS framework shares several principles with the AI-GA framework (Clune, 2020) mentioned in Section 3.1.1, in particular regarding the the central role of bi-level optimization such as meta-learning and feedback effects from **adaptability** to **environmental complexity**. However, the AI-GA framework is entirely focused on an AI perspective and places little emphasis on biological and ecological grounding, in particular with respect to the human species. Finally, Steels (2011) also emphasizes the role of cultural evolution within constraints provided by the biology and the ecological niches in which human populations operate, recognizing strong interactions between three evolutionary processes (biological, social and cultural). Although his analysis can to some extent map to some of the components and interactions proposed in our ORIGINS framework, it is however entirely focused on language evolution.

A humanist perspective on AI. To conclude this chapter, I would like to illustrate how the ORIGINS framework can potentially help us to think about important opportunities and challenges met by our species. As mentioned in the introduction (Section 3.1.2), the framework considers open-ended skill acquisition (OESA) as the most distinctive and interesting feature in humans, in particular because it is responsible for the most beautiful and the most harmful human activities including science, arts and technological innovation, as well as the over-exploitation of natural resources leading to a major ecological issues. Interestingly we can consider science as a result of the interaction between the three main items of the **cultural repertoire** component in the framework: *technology*, *communication* and *social organization*. Indeed, modern science is made possible through a collective agreement (*communication*) for allocating a part of the population resource budget (*social organization*) to allow scientists to invest their time to explore new opportunities (e.g. *technology*) (Harari, 2014, see also Graeber & Wengrow (2021) for alternative theories). In turn, scientific outcomes have a strong influence on all other components of the system by increasing the understanding of our environment (e.g. physics), of our multi-agent dynamics (e.g. social science), of our adaptability (e.g.

cognitive science) and of our cultural repertoire (e.g. history). The resulting knowledge and skills also inevitably increase our control over all these components, augmenting our cultural repertoire with new technologies and cultural practices (e.g. medicine or agriculture). However, perhaps ironically, this strong feedback loop is currently resulting in an unprecedented loss of control, with the over-exploitation of resources leading to a major collapse in bio-diversity and to the global warming of the Earth climate (e.g. Cook et al., 2016). It is therefore essential for our species to better understand this complex dynamics in order to (1) identify and find agreements on which outcomes of OESA we want to encourage or to discourage, (2) characterize potentially harmful feedback loops that will lead to irreversible effects and (3) take advantage of our self-determination, as autotelic agents, to shape our own cultural repertoire in a way that self-regulate the entire system and prevent it to reach critical states that could put our own survival at risk. Better understanding the dynamics of open-ended skill acquisition appears as an important research direction in this context. In this sense, the ORIGINS framework proposes a humanist perspective on AI: using it as a family of computational tools that can help us to explore and study the mechanisms driving open-ended skill acquisition in both artificial and biological systems, as a way to better understand the dynamics of our own species within its whole ecological context.

Part II

Selected contributions and future work

Chapter 4

Introduction

In this Part, I present some selected contributions realized during my scientific career (see Chapter 2 for a summary of this career). Two main criteria have driven the selection of these contributions: (a) their impact and (b) their relevance to the ORIGINS framework presented in Chapter 3. I have taken the occasion to re-interpret my previous contributions in the light of this framework, even though many of them were realized prior to its proposition (i.e. prior to 2021). I found this exercise interesting, as it allows to test the expressiveness of the framework (i.e. is it able to express prior contributions which were not realized with this framework in mind?).

This part is organized in three sections mirroring the three levels proposed in Sections 3.4, 3.5 and 3.6. In consequence, it does not follow a chronogical order but instead:

- Contributions related to level-1 dynamics in Chapter 5 were all realized after having proposed the first version of the ORIGINS framework in Nisioti & Moulin-Frier (2020). It is at this time that I became strongly interested in the main research question of level-1 dynamics: *What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?*
- Contributions related to level-2 dynamics in Chapter 6 were realized during the first part of my career (PhD thesis at GIPSA-Lab and post-doc at Inria-Flowers, i.e. 2007-2015). As mentioned in the description of my scientific trajectory in Chapter 2, I was mostly interested at this time in modeling the evolution and acquisition of language. In the proposed framework, this corresponds to the main research question regarding level-2 dynamics: *What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire?*
- Contributions related to level-3 dynamics in Chapter 7 were realized after my recruitment as a permanent researcher at Inria-Flowers in 2019. In particular, the contribution in Section 7.2 corresponds to a research project which started before my recruitment and the proposition and the ORIGINS framework, but it addresses central aspects of the main research question of level-3 dynamics: *What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition?*

Thus, the contributions related to level-1 dynamics presented in Chapter 5 are the most recent (from 2021) and, in consequence, have yet had a limited impact in the community. I have chosen to present them in this thesis because they illustrate well my most recent research interests, as well as the methodology adopted for the implementation of concrete experiments related to the ORIGINS framework.

As we will see, the presented contributions rely on a diversity of computational methods including: abstract eco-evolutionary models (Section 5.2), simulation environments (Section 5.3), artificial life (Section 5.4), multi-agent representation learning (Section 5.5), intrinsically-motivated machine learning (Section 6.2), Bayesian cognitive modeling (Section 6.3), deep reinforcement learning (7.2) and multi-agent reinforcement learning (Section 7.3).

I conclude each chapter with concrete propositions for future work.

Chapter 5

The eco-evolutionary origins of autotelic agents (Level-1 dynamics)

Contents

5.1 Introduction

In this chapter, I present some of my contributions and propose future work related to the modeling of level-1 dynamics. As presented in Section 3.4, this part of the ORIGINS framework considers pre-existing **environmental complexity** operating at multiple spatiotemporal scales (e.g. seasonal cycles and climate variation), modulating constraints and opportunities (e.g. in terms of resource availability and exposition to predators) exposed to the agents. The objective is to study how this environmental complexity interacts with **adaptability** and **multi-agent dynamics** toward the evolution of learning and exploration mechanisms. The main research question at this level is: *What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?*

In Section 5.2, I first present a recent contribution on a minimal evo-eco-devo simulation model where a population of agents adapts to environmental changes both at the developmental scale (through plasticity) and at the evolutionary scale (through fitness-based selection and niche-limited competition). In Section 5.3, I present our first attempt at designing an ecologically valid simulation environment for reinforcement learning which is grounded in hypotheses from human behavioral ecology. Then, in Section 5.4, we study how low-level sensorimotor control can emerge from a self-organizing system where there is initially no distinction between agents and the environment. Finally, in Section 5.5, I present another recent contribution showing how multi-agent environments, where agents do not have access to the observations of others but can communicate within a limited range, guarantees a common context that can be leveraged in individual representation learning. Finally, in Section 5.6, I mention my other contributions on this topic and propose concrete future work.

5.2 Plasticity and Evolvability Under Environmental Variability: the Joint role of Fitness-based Selection and Niche-limited Competition

5.2.1 Context

This work was realized in 2021-2022 in the Flowers team at Inria (France), in collaboration with Eleni Nisioti. It was partially funded by the Inria Exploratory action ORIGINS (https://www.inria.fr/en/origins) as well as the French National Research Agency (https://anr.fr/, project ECOCURL, Grant ANR-20-CE23-0006). This work also benefited from access to the HPC resources of IDRIS under the allocation 2020-[A0091011996] made by GENCI, using the Jean Zay supercomputer.

This contribution has been published at the GECCO 2022 conference:

Nisioti, E. and Moulin-Frier, C.
Plasticity and evolvability under environmental variability: The joint role of fitness-based selection and niche-limited competition.
In *Proceedings of the 2022 Genetic and Evolutionary Computation Conference (GECCO 2022)*, 2022

Open access version of this publication is available at https://arxiv.org/abs/2202.08834. The rest of this section is a summary of this paper.

5.2.2 Relevance to the proposed ORIGINS framework

Figure 5.1 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. Here we study the interactions between **environmental complexity**, **adaptability** and **multi-agent dynamics** within a minimal evo-eco-devo simulation model. We consider a population of agents that can adapt both at the developmental scale (through plasticity) and at the evolutionary scale (through fitness-based selection and niche-limited competition). Agents evolve in environments that varies from generation to generation and that are structured in multiple ecological niches with different capacities. We study how different types of environmental variability influences plasticity and evolvability under different selection mechanisms.

5.2.3 Abstract

The diversity and quality of natural systems have been a puzzle and inspiration for communities studying artificial life. It is now widely admitted that the adaptation mechanisms enabling these properties are largely influenced by the environments they inhabit. Organisms facing environmental variability have two alternative adaptation mechanisms operating at different timescales: *plasticity*, the ability of a phenotype to survive in diverse environments and *evolvability*, the ability to adapt through mutations. Although vital under environmental variability, both mechanisms are associated with fitness costs hypothesized to render them unnecessary in stable environments. In this work, we study the interplay between environmental dynamics and adaptation in a minimal

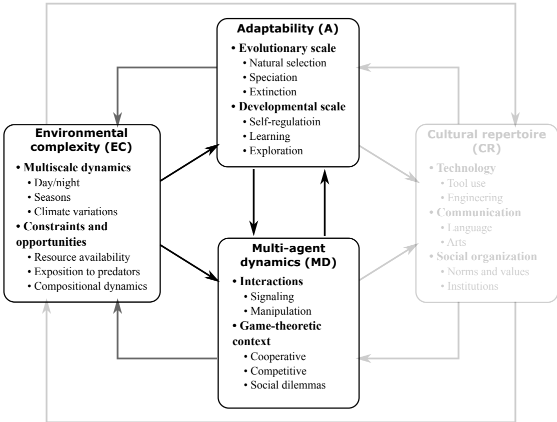


Figure 5.1: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

model of the evolution of plasticity and evolvability. We experiment with different types of environments characterized by the presence of niches and a climate function that determines the fitness landscape. We empirically show that environmental dynamics affect plasticity and evolvability differently and that the presence of diverse ecological niches favors adaptability even in stable environments. We perform ablation studies of the selection mechanisms to separate the role of fitness-based selection and niche-limited competition. Results obtained from our minimal model allow us to propose promising research directions in the study of open-endedness in biological and artificial systems.

5.2.4 Methods

We model environmental variability through a climate function L (e.g. a sinusoid) from time (in number of generations) to a 1D environmental state in $\mathbb{R}$ (Figure 5.2, Left). We consider N niches differing from an offset in their respective climate functions. The environmental state of a niche at a given time specifies its current capacity, i.e. the maximal number of agents that can occupy it.

Each individual agent is characterized by a tolerance curve specifying its probability of surviving depending on the current environmental state in a given niche (Figure 5.2, Right). A tolerance curve is a Gaussian pdf, with its mean corresponding to the preferred environmental state of the agent (highest fitness) and its standard deviation corresponding to the plasticity of the agent (how much the fitness is reduced when moving away from the preferred state).

The genome of an individual agent contains both the mean and standard deviation of its tolerance curve, as well as its mutation rate. Therefore, both plasticity (through the tolerance curve standard deviation) and evolvability (through the mutation rate) can evolve in this model. Our proposed selection mechanism entails two independent assumptions inspired from natural evolution:

1. *niche-limited competition (N-selection)*: when deciding which individuals will repro-

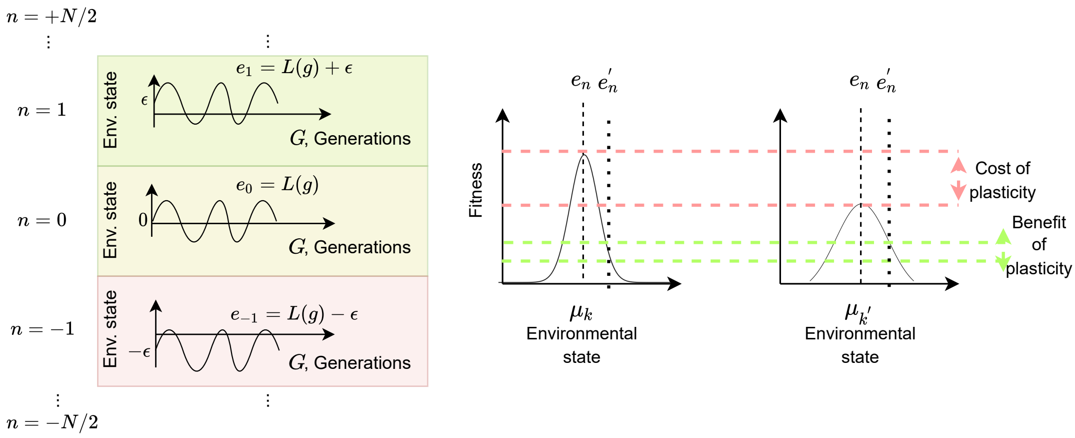


Figure 5.2: (Left) The latitudinal model we employ to describe how the environmental state varies across niches: a single climate function L (illustrated here as a sinusoidal curve) specifies the global environmental dynamics across niches, each niche n differing from its vertical offset equal to $\epsilon \cdot n$. Thus niches with higher index n have higher states, and therefore, higher capacity. (Right) Modeling plasticity as a normal distribution $\mathcal{N}(\mu_k, \sigma_k)$. A non-plastic individual (k) has small σ_k and a high peak at their preferred niche, while a plastic individual ($k^{'}$) has large $\sigma_{k'}$ and a lower peak at their preferred niche. Fitness in a given niche n is computed as the probability density function of the distribution at the environmental state e_n. This figure also illustrates the cost and benefit of plasticity, assuming that $\mu_k = \mu_{k'}$. If $e_n = \mu_k$ (the actual environmental state is identical to the preferred niche of both individuals) the plastic individual has lower fitness (cost of plasticity). If the environmental state is different from the preferred state, e.g. $e_n' > \mu_k$, the plastic individual has higher fitness (benefit of plasticity).

duce, we study each niche independently;

2. *fitness-based selection (F-selection)*: within a niche, individuals produce offspring until its capacity is filled, with fitter individuals being chosen with higher probability.

Only individuals that can survive in a niche are considered for reproduction within it. We refer to this mechanism as NF-selection.

To evaluate the population we compute the following metrics at the end of each generation g:

1. $\bar{\mu}^g$, the value of the preferred environmental state, averaged over the population. This metric indicates that the population is well-adapted when $\bar{\mu}^g$ tracks the form of the climate function L.

2. $\bar{\sigma}^g$, the value of the standard deviation σ^g for preferred environmental states, averaged over the population. We refer to this metrics as the population-average plasticity.

3. $\bar{r}^g$, the mutation rate r component averaged over the population, which denotes the population-average evolvability.

4. $X^g = \sum_k X_k^g$, the number of extinctions. We denote the survival of individual k in niche n at generation g as a binary variable:

$$s_{k,n}^g = (e_{n,g} \in [\mu_k^g - 2\sigma_k^g, \mu_k^g + 2\sigma_k^g]) \tag{5.1}$$

Thus, an individual goes extinct ($X_k^g = 1$) if $\sum_n s_{k,n}^g$ is zero and survives ($X_k^g = 0$) if $\sum_n s_{k,n}^g$ is positive.

5. Population survival A^g, the percentage of generations that a run of our algorithm survived for. Values smaller than 1 are indicative of a mass extinction.

6. V^g, the diversity of the population defined as the standard deviation of the population's genes, formally:

$$V = \sigma_{\mu^g} + \sigma_{\sigma^g} + \sigma_{r^g} \tag{5.2}$$

This metric captures the genetic diversity of the population.

7. D^g, the dispersal of the population, computed as the number of niches over which at least one individual survives for a temporal window of at least w generations. As this metric arises from the interaction of the genome and environment, we can view it as a measure of phenotypic diversity.

5.2.5 Results

5.2.5.1 Evolving in a stable environment

We define a stable environment as one where the climate function, and hence, the reference environmental state is constant, i.e. $e_0^g = e_0^0, \quad \forall g \in \mathcal{G}$. The environmental state of the different niches is, therefore, equal to $e_n^g = e_0^0 + n \cdot \epsilon$. We experiment with high-quality environments ($e_0^0 > 4$) which can support a large population, low-quality environments ($e_0^0 \leq 0.5$) where the capacity is low with the majority of niches being unable to support any individuals and medium-quality environments ($0.5 < e_0^0 < 4$) where some of the niches become uninhabitable for large enough N.

Low-quality environments with multiple niches favor plasticity. Figure 5.3 presents the population-average plasticity after convergence, $\bar{\sigma}^*$, under NF-selection, various environmental conditions and number of niches. We observe that when there is a single niche ($N = 1$) plasticity converges to a very low value regardless of the state. This is intuitive as the cost of plasticity captured by our genome model renders individuals with the smallest σ_k^* the fittest. This agrees with previous studies in constant environments Grove (2014) (note that F-selection is identical with NF-selection when there is a single niche). However, the picture differs significantly when there are multiple

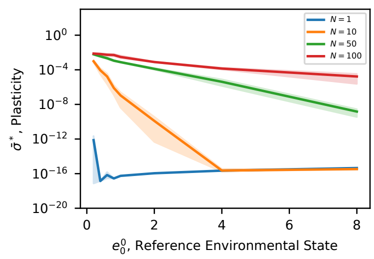


Figure 5.3: Population-average plasticity after convergence ($\bar{\sigma}^*$) in a constant environment under NF-selection.

niches and low-quality environments: as an individual can reproduce in any of the niches it can survive in, higher plasticity means higher chances of reproduction, which counteracts the cost of plasticity. As the quality of environments increases the benefit of plasticity disappears: non-plastic individuals dominate the available niches even though some individuals choose to disperse. In contrast to these interesting dynamics of plasticity, we observed that the population-average evolvability remained very low ($\bar{r}^* < 10^{-10}$) in all conditions. This observation is inline with the intuition that mutations disappear in stable environments as they incur fitness costs (Lynch et al., 2016; Giraud et al., 2001).

In the full paper (Nisioti & Moulin-Frier, 2022), we also show that:

- niche-limited competition is necessary for plasticity to persist,
- diversity is highest under NF-selection,
- F-selection leads to more early extinction events.

5.2.5.2 Evolving under periodic variability

We model periodic variability as a sinusoid with period T_e and amplitude A_e that dictates the evolution of the reference environmental state $e_{0,g}$.

Plastic and evolvable individuals emerge under NF-selection only when the number of niches is sufficient. In Figure 5.4 (Left and Middle) we observe the ability of the population to survive when we vary the amplitude of oscillations (A_e) and number of niches (N) respectively, for different values of the oscillation period T_e. We can draw various conclusions from these results:

1. increasing the number of niches enables the population to survive longer in environments changing more frequently
2. survival is guaranteed for small-amplitude variations ($A_e = 0.2$) regardless of their frequency
3. in the case of large-amplitude variations ($A_e = 8$) high frequency does not allow the population to adapt at all
4. In the case of medium-amplitude ($A_e = 1$) survival is possible only under low-frequency variation

In Figure 5.4 (Right), we further analyze the ability of the population to survive by looking at how the evaluation metrics evolve with increasing generations. We observe that the population manages to track the environmental variability by keeping both plasticity and evolvability high, with oscillations in plasticity and evolvability occuring at twice the period of variability T_e, as they increase at both transition points. Diversity is slightly lower during the low peaks of e_0. We should note that we did not find high-frequency cases where the population reacted solely through plasticity or low-frequency cases where the

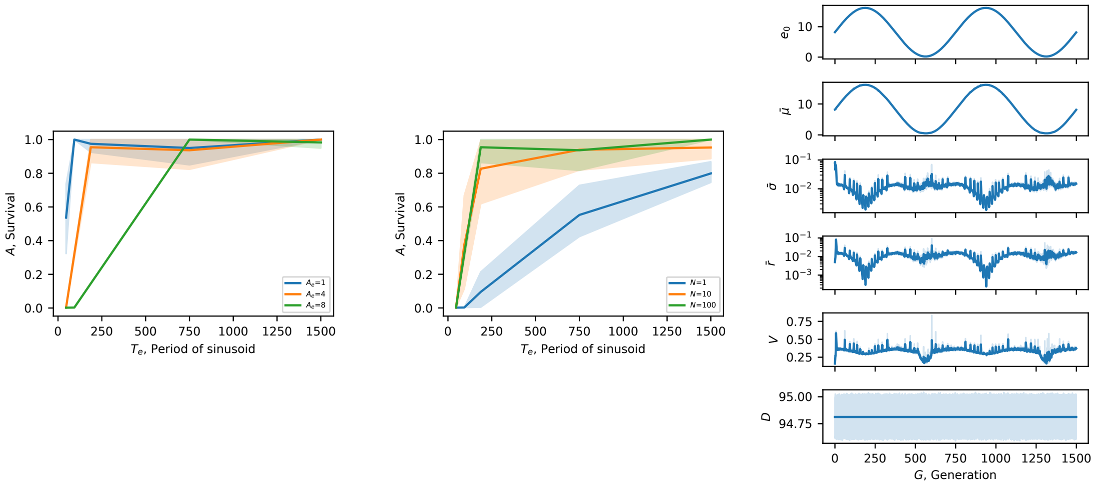


Figure 5.4: Results in a periodic environment. Left: Survival (A) as the percentage of generations without a mass extinction under NF-selection with genome R_{evolve}, N =100 niches and varying period T_e and amplitude A_e. Middle: Survival (A) as the percentage of generations without a mass extinction under NF-selection with genome R_{evolve}, A_e = 4 and varying period T_e and Number of niches. Right: Evolution under NF-selection with genome R_{evolve}, N =100 niches, T_e = 750 and A_e=8.

population adapted solely through evolvability, as indicated by previous studies (Cuypers et al., 2017).

In the full paper (Nisioti & Moulin-Frier, 2022), we also show that F-selection and N-selection in isolation favor the evolution maladapted plastic individuals in periodic environment. In addition, we also analyze the dynamics of the system in noisy environments.

5.2.6 Discussion

We have designed a simple model of the evolution of plasticity and evolvability and studied the complex interactions between environmental and population dynamics. Experiments have revealed many insights into the evolution of adaptation mechanisms. Taking into account the effect of niche-limited competition gives rise to qualitatively different solutions to the plasticity-evolvability trade-off, in turn affecting population properties such as diversity and dispersal. We hope that our work sheds light into the plethora of related works and will prove useful in future studies of both artificial and natural systems.

We believe that Quality-Diversity algorithms (Lehman & Stanley, 2011; Cully et al., 2015; Pugh et al., 2016) can be a particularly promising application area for such studies. Similarly to our proposal, this community lays emphasis on the benefits of combining niche-limited competition and fitness-based selection; however, as our empirical results indicate, parameters such as the number and quality of niches, as well as the form and presence of environmental variability show great qualitative impact and can potentially act as a curriculum for the emergence of adaptation.

From the perspective of human behavioral ecology our results offer insights into existing hypotheses:

1. the observation that low-quality environments favor plastic individuals, while high-quality environments favor non-plastic individuals (see Section 5.2.5.1) hints to the turnover pulse hypothesis (Vrba, 1985);

2. the observation that adaptability is favored by abrupt transitions (see Section 5.2.5.2) and high variability (see published paper Nisioti & Moulin-Frier, 2022) hint to the variability selection hypothesis (Potts, 2013)

We should note that these hypotheses share a common model of environmental variability characterized by a large diversity of niches (Maslin et al., 2015).

It is important to also note the assumptions made by our study and how future work in artificial intelligence and artificial life can help overcome its limitations. First, tolerance curves assume that plasticity comes at a cost, an assumption that is often questioned in natural systems; we believe that studies with artificial systems can reveal whether such costs indeed arise. Second, our model does not capture the mechanism of species co-adaptation and can therefore not offer insights on how the dynamics of arms races are influenced by resource availability, as proposed by the Red Queen hypothesis (Van Valen, 1973; Solé, 2022; Pearson, 2001). Then, our model assumes that there are no constraints on plasticity; we believe that studies with a more complex genotype to phenotype mapping that employ different adaptation mechanisms to ensure plasticity can reveal mechanism-specific limits that will extent the conclusions reached in this work. From an evaluation perspective, we have limited ourselves to measuring easily quantifiable properties of the genome and behavior space, such as adaptability and diversity, that have been linked to open-endedness (Pugh et al., 2016). As a next step we plan to investigate direct measures of open-endedness (Dolson et al., 2019). Finally, we believe that progress in open-endedness requires a better understanding of niching in recent simulation environments employed by the deep reinforcement learning community (Suarez et al., 2019; Cobbe et al., 2019; Nisioti et al., 2021).

Finally, as proposed by the ORIGINS framework presented in Chapter 3, the effect of environmental variability on adaptation mechanisms is only part of the overall picture. To understand the complexity of the human ecological niche, we need to take into account multi-agent dynamics and culture, which modulate the processes of selection and niche construction (Eppe & Oudeyer, 2021; Nisioti et al., 2021). We believe that exploring these links in social-centric multi-agent reinforcement learning environments (Kovac et al., 2021) will further improve our understanding of adaptability.

5.3 Towards an ecologically valid simulation environment

5.3.1 Context

This work was realized in 2021-2022 in the Flowers team at Inria (France), in collaboration with Eleni Nisioti. It was partially funded by the Inria Exploratory action ORIGINS (https://www.inria.fr/en/origins) as well as the French National Research Agency (https://anr.fr/, project ECOCURL, Grant ANR-20-CE23-0006). This work also benefited from access to the HPC resources of IDRIS under the allocation 2020-[A0091011996] made by GENCI, using the Jean Zay supercomputer.

This contribution has been published at the *Ecological Theory of Reinforcement Learning* workshop of the *Conference on Neural Information Processing Systems* (NeurIPS 2021):

Nisioti, E., Jodogne-del Litto, K., and Moulin-Frier, C. Grounding an Ecological Theory of Artificial Intelligence in Human Evolution. In *NeurIPS 2021 - Conference on Neural Information Processing Systems / Workshop: Ecological Theory of Reinforcement Learning*, virtual event, France, 2021

Open access version of this publication is available at https://hal.archives-ouvertes.fr/hal-03446961/.

This paper presents both an earlier version of the framework presented in Chapter 3 as well as a novel simulation environment grounded in hypotheses from human behavioral ecology (HBE).

The rest of this section only focuses on the simulation environment.

5.3.2 Relevance to the proposed ORIGINS framework

Figure 5.5 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. Here we entirely focus on **environmental complexity**, leaving the interaction with all the other components of the framework for future work. The presented environment aims at simulating some aspects of climate dynamics supposed to have played a key role in human behavioral ecology (Sections 3.2.2.1 and 3.4).

5.3.3 Proposed simulation environment

Many exciting environments have been recently proposed in the AI community (Hafner, 2021; Johnson et al., 2016). However, most existing environments do not display rich intrinsic dynamics independently of the agents' actions and, to the best of our knowledge, none implements climate dynamics. The Jelly Bean World (JBW) is a two-dimensional grid-world where agents navigate and collect items (Platanios et al., 2020). Originally introduced as a benchmark for continuous learning, this environment automatically expands the world when the agent approaches its boundaries. The necessity for generating new parts of the world on demand in JBW led to the adoption of a low-complexity yet powerful mechanism for creating new items (e.g. new resources or environmental

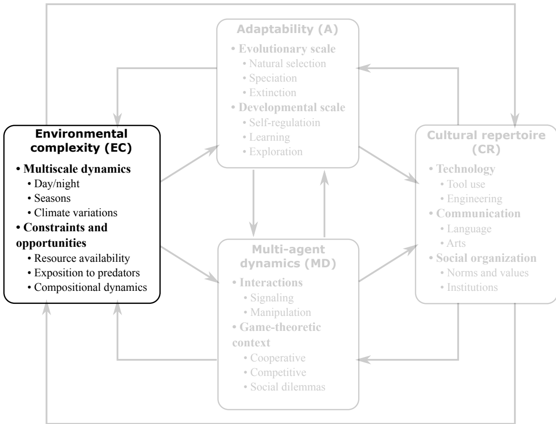


Figure 5.5: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

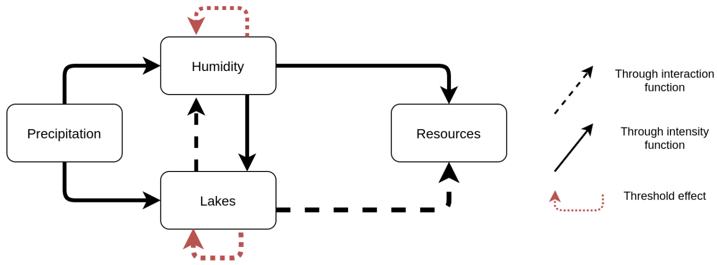
elements). Specifically, the creation and deletion of items are controlled by a probability distribution that can be configured through an intensity and interaction function, the former determining the probability of existence of an item independently of others, and the latter in relation to them. Using this mechanism, one can form a variety of item generation patterns, such as clusters and custom, spatially non-stationary distributions.

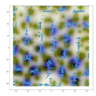
To enable the empirical experimentation of hypotheses proposed under the Pulsed Climate Variability (PCV) framework (see Section 3.2.2.1) we have equipped an existing simulation, The Jelly Bean World (JBW), with climate dynamics. Our objective is to observe the appearance of lakes with interesting dynamics, such as quick expansion and chaotic contraction (Trauth et al., 2010), which in its turn will modulate the presence of resources available to the agents. To achieve this, we have extended the existing item generation mechanism in JBW with context-dependent resource generation. From an ecological perspective, the intensity function can be used to model an external climate-related parameter, which in our case is the level of precipitation. Then, the interaction function can be used to model climate-related constraints, such as "resources grow only near water" and "lakes change their size based on humidity". To implement this functionality, we defined new types of items, i.e. , water cells that can be used to form lakes, resources (called "jelly-beans" and "bananas" as in the original JBW environment) that grow near lakes and trees.

A simplified model of our proposed climate dynamics is depicted in Figure 5.6(a), while Figure 5.6(b) presents the item presence patterns that arise from it. In this example, precipitation has a pulse form, which allows us to compare item patterns between periods of low (in Figure 5.6(c)) and high precipitation (5.6(d)).

5.3.4 Discussion

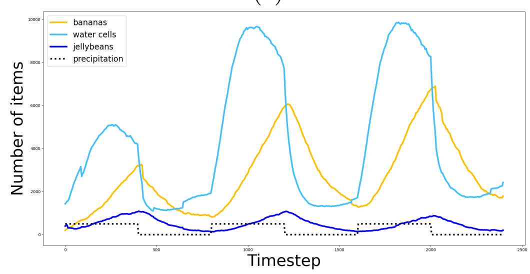
This first step towards an ecologically-valid environment has obvious limitations: it only consider a single component of the ORIGINS framework (**environmental complexity**) the introduction of adaptive agents has been left for future work (see Section 5.6.2). Our

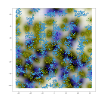




(c)

(a)





(d)

(b)

Figure 5.6: Climate dynamics in our proposed environment: (a) simplified model of the climate dynamics (b) temporal patterns of lake and item presence during simulations with a precipitation function having a pulse form (c) a top-view of a gridworld where an agent navigates in a grid-world populated by lakes (green), jelly-beans (purple), bananas (yellow) and trees (green), whose presence is influenced by a user-designed precipitation function during a low-precipitation period (c) and a high-precipitation period (d)

prediction is that our proposed environment will pose a challenge for standard methods in reinforcement learning that will struggle with non-stationarity and the sequential nature of the climate dynamics. We believe that bi-level optimization (Hutchison et al., 2010; Najarro & Risi, 2021) is an interesting direction as it can model the interaction between evolutionary and developmental processes (see Section 5.6.2 for more detail on future work).

5.4 Learning Sensorimotor Agency in Cellular Automata

5.4.1 Context

This work was realized in 2020-2022 as a collaboration between the Flowers group at Inria (France) and Bert Chan[1], now at Google Brain Tokyo (Japan). It builds upon recent exploration algorithms for automated discovery in complex systems developed with Mayalen Etcheverry (PhD student at Flowers) and Chris Reinke (former post-doc at Flowers) (Reinke et al., 2020-02-17, 2020-2-17; Etcheverry et al., 2020). This work benefited from access to the HPC resources of IDRIS under the allocation 2020-[A0091011996] made by GENCI, using the Jean Zay supercomputer.

The main publication presenting this contribution takes the form of an extended online blog post, including videos and an online demo:

Hamon, G., Etcheverry, M., Chan, B. W.-C., Moulin-Frier, C., and Oudeyer, P.-Y. Learning Sensorimotor Agency in Cellular Automata (blog post: https://developmentalsystems.org/sensorimotor-lenia/), 2022

Open access version of this publication is available at `https://developmentalsystems.org/sensorimotor-lenia/`. The rest of this section is a summary of this blog post. We are currently working on a journal article, extending it with a quantitative analysis.

5.4.2 Relevance to the proposed ORIGINS framework

Figure 5.7 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. This contribution adopts a radical enactive view of cognition where, contrarily to all the other contributions presented in this thesis, there is originally no distinction between agents and the environment. Instead, we use a multi-channel continuous cellular automaton where the local rules governing some channels are fixed, modeling an imposed environmental dynamics, while the rules governing other channels can be adapted according to some objective. We study how such a self-organizing system, under some conditions of **environmental complexity**, can evolve low-level sensorimotor skills from scratch (**adaptability**). Interestingly we observe that the evolved proto-agents (called *creatures* in the paper), while trained in isolation for spatial localization and navigation skills, show impressive generalization abilities to unseen environmental conditions. In particular, rich **multi-agent dynamics** emerges when multiple creatures are simultaneously placed in a shared environment at test time.

5.4.3 Abstract

Novel classes of Cellular Automata (CA) have been recently introduced in the Artificial Life (ALife) community, able to generate a high diversity of complex self-organized patterns from local update rules. These patterns can display certain properties of biological systems such as a spatially localized organization, directional or rotational movements, etc. In

[1] `https://chakazul.github.io/`

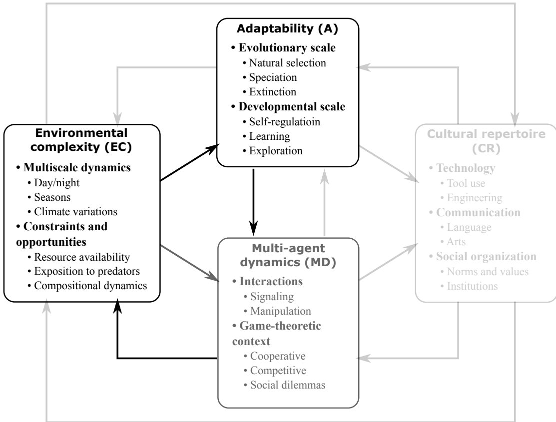


Figure 5.7: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

fact, CA have a long relationship with biology and especially the origins of life/cognition as it is a self-organizing system that can serve as a computational testbed and toy model for such theories (Beer, 2004) but also as a source of inspiration on what are the basic building block of "life". However, while the notions of embodiment within an environment, individuality and self-maintenance are central in theoretical biology and in particular in the definition of agency (e.g. Maturana & Varela, 1980; Varela, 1997), it remains unclear how such mechanisms and properties can emerge from a set of local update rules in a CA. In this contribution, we propose an approach enabling to learn self-organizing agents capable of reacting to the perturbations induced by the environment, i.e. robust agents with sensorimotor capabilities. We provide a method based on curriculum learning, on diversity search and on gradient descent over a differentiable CA able to discover the rules leading to the emergence of such creatures (Figure 5.8 A and B). The creatures obtained, using only local update rules, are able to regenerate and preserve their integrity and structure while dealing with the obstacles or other creatures in their way (Figure 5.8 C). They also show great generalization, with robustness to changes of scale, random updates or perturbations from the environment not seen during training. We believe that the field of artificial intelligence could benefit from these capabilities of self-organizing systems to make robust intelligent systems that can quickly adapt to new environments and to perturbations.

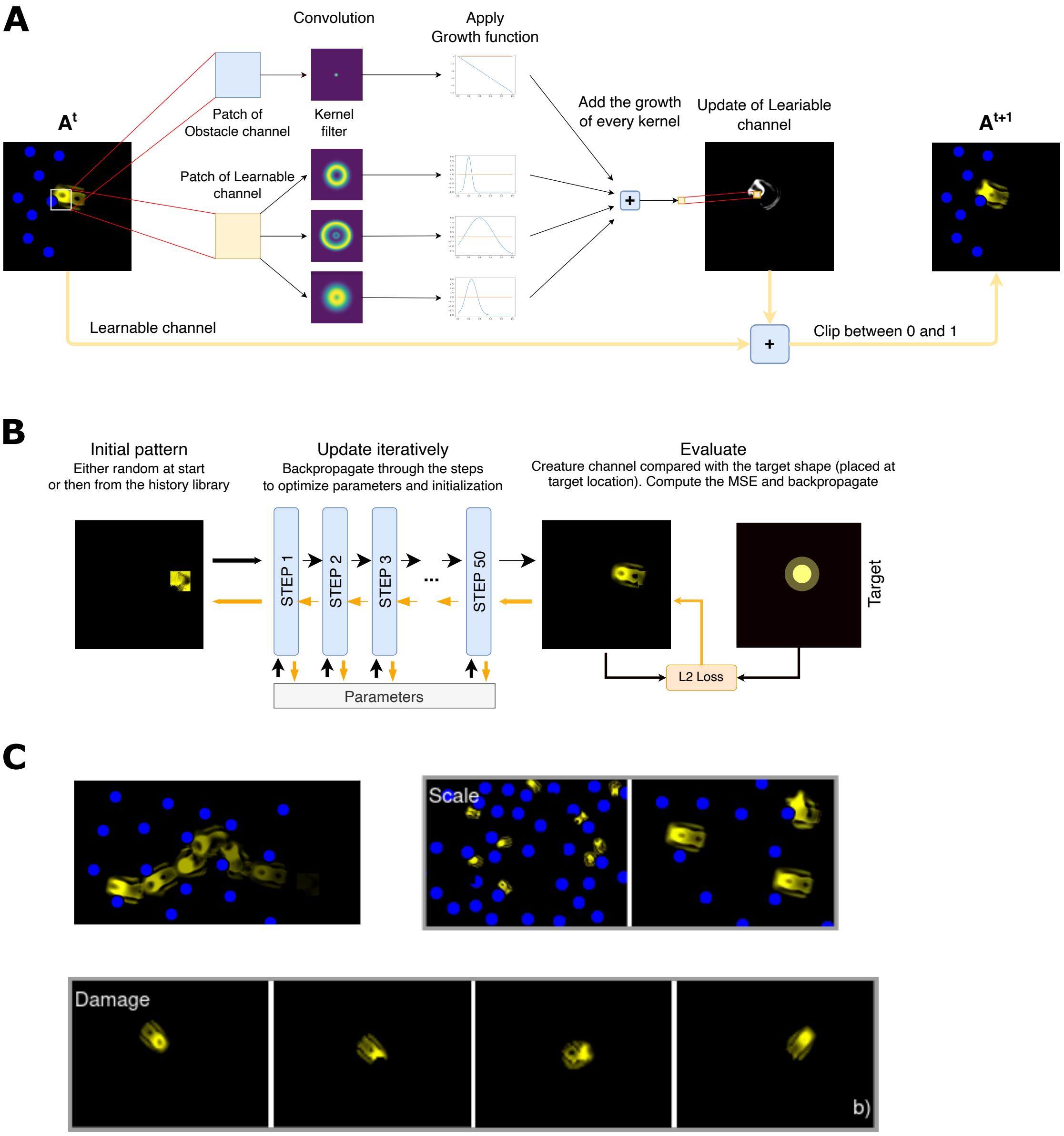


Figure 5.8: A. We use a multi-channel continuous cellular automaton (Lenia). The state A^t at time t is a square grid with two colored channels: an Obstacle channel in blue and a Learnable channel in yellow. Each channel is associated with a set of convolution kernel filters and growth functions, whose outputs are summed up to compute of next step A^{t+1}. B) Parameters of kernels and growth functions are optimized using backpropagation through time, minimizing a L2 loss between the final state of the learnable channel and a target shape at a specific location. Target locations and obstacle positions are randomized across episodes. C) Optimized creatures in the Learnable channel are able to maintain their integrity while navigating in a field of obstacles (top left). They also generalize their skills in environments not encountered during training, e.g. when placed in multiple-creature environments or when damaged (top-right and bottom).

5.5 Socially Supervised Representation Learning: the Role of Subjectivity in Learning Efficient Representations

5.5.1 Context

This work was realized in 2021-2022 in the Flowers team at Inria (France), in collaboration with Julius Taylor and Eleni Nisioti. It was partially funded by the Inria Exploratory action ORIGINS (https://www.inria.fr/en/origins) as well as the French National Research Agency (https://anr.fr/, project ECOCURL, Grant ANR-20-CE23-0006). This work also benefited from access to the HPC resources of IDRIS under the allocation 2020-[A0091011996] made by GENCI, using the Jean Zay supercomputer.

This contribution has been published at the AAMAS 2022 conference:

Taylor, J., Nisioti, E., and Moulin-Frier, C.
Socially Supervised Representation Learning: The Role of Subjectivity in Learning Efficient Representations.
In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2022)*, 2022

Open access version of this publication is available at https://arxiv.org/abs/2109.09390. The rest of this section is a summary of this paper.

5.5.2 Relevance to the proposed ORIGINS framework

Figure 5.9 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. Here we study how particular **multi-agent dynamics**, where agents do not have access to the observations of others but can communicate within a limited range, guarantees a common context that can be leveraged in individual representation learning to improve **adaptability**. In the presented experiments, **environmental complexity** is highly simplified.

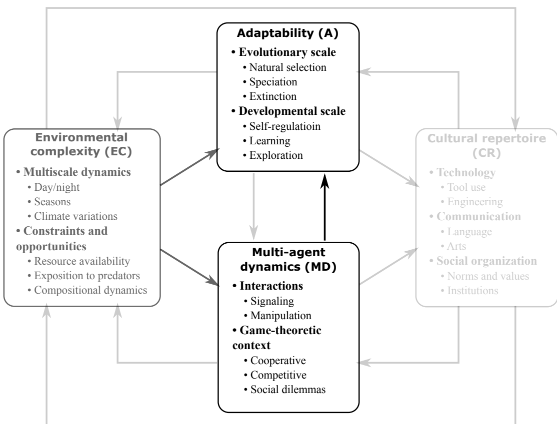


Figure 5.9: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

5.5.3 Introduction

In this work, we propose that aligning internal subjective representations, which naturally arise in a multi-agent setup where agents receive partial observations of the same underlying environmental state, can lead to more data-efficient representations. We propose that multi-agent environments, where agents do not have access to the observations of others but can communicate within a limited range, guarantees a common context that can be leveraged in individual representation learning. The reason is that subjective observations necessarily refer to the same subset of the underlying environmental states and that communication about these states can freely offer a supervised signal. To highlight the importance of communication, we refer to our setting as *socially supervised representation learning*. We present a minimal architecture comprised of a population of autoencoders, where we define loss functions, capturing different aspects of effective communication, and examine their effect on the learned representations.

Contributions. We summarize our contributions as follows:

1. We highlight an interesting link between data-augmentation traditionally used in single-agent self-supervised setting and a group of agents interacting in a shared environment (Figure 5.10).
2. We introduce *Socially Supervised Representation Learning*, a new learning paradigm for unsupervised learning of efficient representations in a multi-agent setup (Figure 5.11).
3. We present a detailed analysis of the conditions ensuring both the learning of efficient individual representations and the alignment of those representations across the agent population (Section 5.5.4).

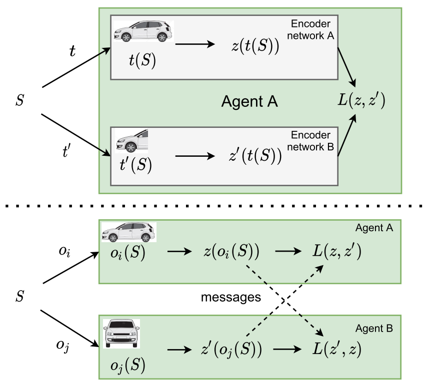


Figure 5.10: Data augmentation in self-supervised representation learning using stochastic image augmentations $t \sim T$ (top) vs. Socially Supervised Representation Learning that substitutes engineered augmentations for perspectives $(o_i, o_j, \dots)$ that arise naturally in multi-agent systems (bottom). Green boxes indicate conceptual *agents*, while we assume that a singular representation learning method may be interpreted as a single agent.

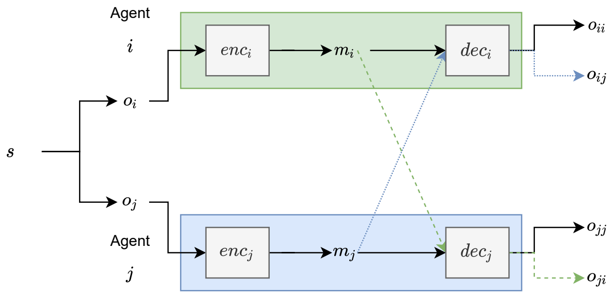


Figure 5.11: Proposed architecture. Two agents i and j (top and bottom coloured boxes) are presented with different observations (o_i and o_j) of the same underlying environment state (s). Each agent (say i) implements a standard autoencoder architecture with an encoder enc_i mapping the input observations to latent representations and a decoder dec_i mapping latent representations to reconstructed observations. Each agent i communicates its latent representation m_i (also called a message) of its own observation input o_i to the other agents. This way, each agent i is able to reconstruct the observation from its own latent representation (o_{ii}: reconstruction by agent i from its own message m_i) as well as from the latent representation of the other (o_{ij}: reconstruction by agent i from the other's message m_j). The architecture can be trivially extended to a larger population, where each agent communicates their latent representations to with each other.

5.5.4 Results

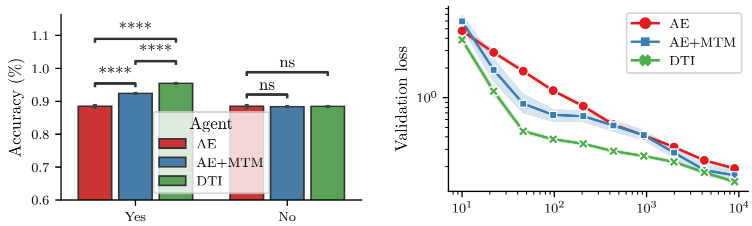


Figure 5.12: Representation quality in terms of standard linear probing and data efficiency. *AE+MTM* and *DTI* are our proposed methods whereas *AE* is an autoencoding baseline which does not benefit from perspectives. **Left**: Classification accuracy using linear probing on top of the learned representations, comparing a condition with different perspectives (*Yes*) and without (*No*). **Right**: Linear probing using validation datasets of varying sizes to assess the data efficiency of representations.

We show that our proposed architecture allows the emergence of aligned representations. This means that different agents find similar encodings for the same sensory inputs. The subjectivity introduced by presenting agents with distinct perspectives of the environment state contributes to learning abstract representations that outperform those learned by a single autoencoder and a population of autoencoders, presented with identical perspectives of the environment state, which is shown in the left column of Fig. 5.12. Furthermore, in Fig. 5.12 (right) we show that the learned representations are data-efficient, i.e. they enjoy the most benefit when evaluated on small testing splits. This is important, because good representations should allow agents to adapt to downstream tasks quickly and with few samples. Altogether, our results demonstrate how communication from subjective perspectives can lead to the acquisition of more abstract representations in multi-agent systems, opening promising perspectives for future research at the intersection of representation learning and emergent communication.

5.6 Other contributions and future work

5.6.1 Other contributions

I have presented in this chapter most of my contributions related to level-1 dynamics, which is the most recent topic I have been interested in (from 2020). On the methodological side, we are currently developing software tools for automated discovery in science and art. These tools aim at providing easy-to-use interactive interfaces for assisting the exploration of complex systems. An automated discovery algorithm was published at the NeurIPS conference (Etcheverry et al., 2020), with its application in the context of the *Minecraft Open-Endedness Challenge*[2] in a blog post (Etcheverry et al., 2021). We plan to release a beta version of an open-source software for automated discovery by the end of 2022. I believe that such automated discovery algorithms can help to search for interesting levels of environmental complexity in the context of studying level-1 dynamics.

5.6.2 Future work

Level-1 dynamics considers pre-existing **environmental complexity** operating at multiple spatiotemporal scales (e.g. seasonal cycles and climate variation), modulating constraints and opportunities (e.g. in terms of resource availability and exposition to predators) exposed to the agents. The objective is to study how this environmental complexity interacts with **adaptability** and **multi-agent dynamics** toward the evolution of learning and exploration mechanisms. The main research question at this level is: *What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies?*

Based on the theoretical analysis of level-1 dynamics in Section 3.4 as well as my existing contributions presented in the current chapter, future work will focus on the following directions.

5.6.2.1 Simulation environments grounded in human behavioral ecology

A central aspect of the ORIGINS framework is that the evolution of OESA is originally driven by certain characteristics of **environmental complexity**. As we have seen in Section 3.4, a variety of hypotheses in HBE highlight the importance of climate dynamics in providing a wide diversity of environmental constraints and opportunities for human evolution. We have also highlighted the importance of diverse ecological niches promoting behavioral diversity. In consequence, simulation environments for studying the evolution of OESA in AI should respect the following desiderata:

- **Unbounded and realistic dynamics at multiple spatiotemporal scales**. Rather than requiring explicit design, patterns of resource availability and exposition to predators should emerge naturally and exhibit complex dynamics (e.g. through

[2] https://evocraft.life/

resource regrowth as in Pérolat et al., 2017). This feature will enable the study of how environmental variability at different scales influence the interaction between evolution, development and learning (see Section 5.6.2.3 below for more detail).

- **Spatial open-endedness, a requirement for the appearance of diverse niches**. The environment should simulate the co-existence of diverse ecological niches resulting in diverse constraints and opportunities for the agents and potentially allowing their speciation and dispersal (Lehman & Stanley, 2013).
- **A variety of tasks relevant to human evolution, such as navigation, harvesting, hunting and crafting through tool use**. In particular, the environment should display *compositional dynamics*, i.e. offering opportunities for agents to produce new elements (e.g. a hammer) by composing other existing elements (e.g. wood and stone), as illustrated in Figure 5.13. This is crucial for enabling behavioral diversity and innovation, an important property of OESA (Pugh et al., 2016; Hintze, 2019).

We have recently achieved preliminary steps in this direction including spatial open-endedness in a grid world with complex climate dynamics (see Section 5.3), diverse niches with various types of abstract environmental dynamics (see Section 5.2), as well as compositional dynamics defined as large "recipe books" defining how elements of the environment can be combined together to create new ones (see Section 7.3 and 7.2).

5.6.2.2 Design of bi-level optimization algorithms mimicking the interplay between evolution and development

For this we want to integrate mechanisms inspired by meta reinforcement learning (META-RL) (Finn et al., 2017), curriculum learning (Portelas et al., 2020) and neuro-evolution (Papavasileiou et al., 2021) (see Section 3.2.2.2). A central conceptual limitation of the META-RL paradigm, in particular, is the fact that environments are sampled randomly, ignoring the curriculum potential of continuous interactions with dynamic environments. In contrast, theories from HBE suggest a major role of spatiotemporal dynamics in human evolution (e.g. the Pulsed Climate Variability (PCV) framework, see Section 3.2.2.1). On the other hand, curriculum learning takes into account the order in which environments are presented to the agents, but existing contributions in this field do not integrate with bi-level optimization as in META-RL. We instead propose to study algorithms with the following desiderata: a) the outer-loop optimization (analogous to evolution) is influenced by the order in which different tasks are experienced b) the inner-loop (analogous to development) can generate a wide range of behavioral patterns, from hard-wired behavior to adaptive learning dynamics to autotelic exploration.

5.6.2.3 Study the role of environmental variability at multiple spatiotemporal scales on the evolution of autotelic agents

As we have seen in Section 3.4, ecological (Johnston, 1982; Stephens, 1991) and computational (Singh et al., 2010; Niv et al.; Lange & Sprekeler, 2021) studies point to the following hypotheses: a) low environmental variability between generations favors innate behaviors b) high environmental variability between generations favors the evolution of learning c) advanced exploration strategies, e.g. in the form of intrinsic motivation, emerge when environmental variability within a lifetime is high enough for exploration to be vital d) prohibitively high variability during a lifetime renders adaptation futile. In a recent contribution, we evaluated the evolution of both developmental plasticity and evolvability under various spatiotemporal environmental dynamics (Nisioti & Moulin-Frier, 2022, see Section 5.2). We now plan to benchmark the performance of bilevel optimization algorithms (as described in 5.6.2.2 above) to gauge their ability to appropriately respond to various types of variability in richer environments (as described in 5.6.2.1). We are particularly interested in environments containing multiple niches, each with diverse sets of available elements, and environmental variability modulating how these elements can be combined through compositional dynamics. We predict that such environments will encourage the evolution of autotelic agents continuously exploring new crafting opportunities even if they don't necessarily influence their evolutionary fitness over their own lifetime, as a solution to be better prepared for upcoming environmental changes.

5.6.2.4 Empirically evaluate hypotheses featured under the Pulsed Climate Variability (PCV) framework

To achieve this, we will model environmental dynamics following the climate curves suggested by paleoclimatology data (as described in 5.6.2.1) and evaluate populations of agents in foraging tasks where resource availability patterns are modulated by these climate curves in our proposed simulation environment (Nisioti et al., 2021, see Section 5.3). Our objective will measure the proposed metrics of speciation, extinction and diversity, which will enable the evaluation and comparison of the hypotheses proposed under the PCV framework. We believe that this type of interdisciplinary work will prove fruitful for both communities: HBE is in need of advanced computational methods (Marean et al., 2015; Frankenhuis et al., 2019) and research in AI will go beyond studying ungrounded types of environmental variability. We have recently proposed a first step in this direction in an evo-eco-devo model (Nisioti & Moulin-Frier, 2022, see Section 5.2) and in a simulation environment (Nisioti et al., 2021, see Section 5.3).

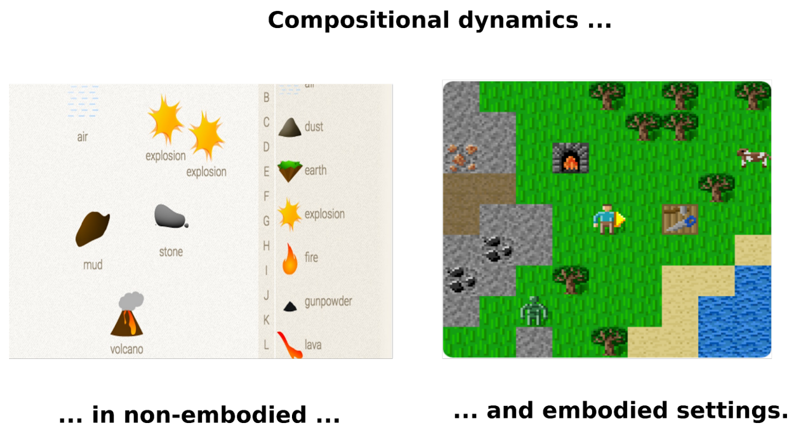


Figure 5.13: **Compositional dynamics as a key feature of simulation environments for OESA**. An environment is said to support compositional dynamics when the agent can produce new elements (e.g. an axe) by composing other existing elements (e.g. wood and stone). On the left, Little Alchemy (https://littlealchemy.com/) is a game with compositional dynamics in a non-embodied setting. On the right, Krafter (Hafner, 2021), a simulation environment inspired from Minecraft where agents encounter elements upon navigating.

Chapter 6

The formation of a cultural repertoire (Level-2 dynamics)

Contents

6.1 Introduction

In this chapter, I present some of my contributions and propose future work in modeling the formation of a cultural repertoire. As presented in Section 3.5, this part of the proposed framework considers pre-existing mechanisms for **adaptability** (emerged from level-1 dynamics, in particular intrinsically motivated goal exploration) as well as for **multi-agent dynamics** (e.g. cooperation and competition pressures with other agents in a shared environment). The objective is to study how the interaction between these mechanisms can bootstrap the formation of a **cultural repertoire**, from its most primitive forms (e.g. structured vocalizations) to more advanced ones (e.g. phonological systems for vocal communication). The main research question at this level is: *What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire?*

In Section 6.2, I present a computational contribution showing how some aspects of early vocal development, required for articulated language to emerge, can self-organize out of general mechanisms of intrinsically motivated goal exploration. Then, in Section 6.3, I present a contribution showing how shared speech sound systems can emerge at the population level through the interaction of vocal agents. Finally, in Section 6.4, I mention my other contributions on this topic and propose concrete future work.

6.2 Self-organization of early vocal development in infants and machines: the role of intrinsic motivation

6.2.1 Context

This work was realized in 2012-2014 in the Flowers team at Inria (France), in collaboration with Sao Mai Nguyen and Pierre-Yves Oudeyer. It was partially funded by the ERC Starting Grant EXPLORERS 240 007 of Pierre-Yves Oudeyer.

The main publication presenting this contribution has been published in *Frontiers in Psychology (Cognitive Science)*:

Moulin-Frier, C., Nguyen, S. M., and Oudeyer, P.-Y.
Self-Organization of Early Vocal Development in Infants and Machines: The Role of Intrinsic Motivation.
Frontiers in Psychology (Cognitive Science), 4(1006), 2014a.
ISSN 1664-1078.
doi: 10.3389/fpsyg.2013.01006

Open access version of this publication is available at `https://www.frontiersin.org/articles/10.3389/fpsyg.2013.01006/full`. The rest of this section is a summary of this paper.

6.2.2 Relevance to the proposed ORIGINS framework

Figure 6.1 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. Here we consider an adaptive agent equipped with a general mechanism of intrinsically-motivated goal exploration (i.e. **adaptability** at the developmental scale). This agent is equipped with a model of the human vocal tract and of the auditory system. In a first experiment, there is **no** environmental complexity (except the agent's own morphology, i.e. its vocal tract producing sounds) and **no** multi-agent dynamics. We study how a general mechanism of intrinsically motivated spontaneous exploration can self-organize developmental stages during early vocal learning, without any imposed constraint to communicate. We observe that the agent spontaneously produces vocalizations of increasing complexity (from no phonation, to unarticulated sounds, to articulated sounds resembling proto-syllables in infants). Then, in a second experiment, the agent also perceives mature vocalizations from surrounding "adult" agents (hence introducing very basic **multi-agent dynamics**) and we study how it influences the agent's exploration toward the ambient sounds. We argue that such developmental structures, self-organizing from general mechanisms of intrinsically-motivated learning, can provide the building blocks in the emergence of a **cultural repertoire** (here providing a repertoire of diverse vocalizations that can potentially bootstrap language evolution and acquisition, see Oudeyer & Smith, 2016, for a complete explanation of the theoretical argument).

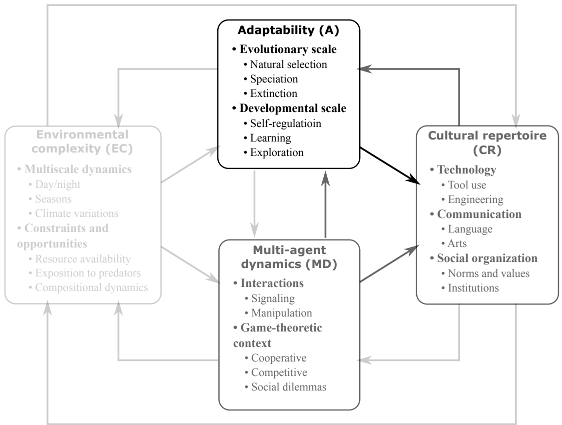


Figure 6.1: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

6.2.3 Abstract

We propose and experimentally test the hypothesis that general mechanisms of intrinsically motivated spontaneous exploration, also called curiosity-driven learning, can self-organize developmental stages during early vocal learning. We introduce a computational model of intrinsically motivated vocal exploration, which allows the learner to autonomously structure its own vocal experiments, and thus its own learning schedule, through a drive to maximize competence progress (Figure 6.2.B). This model relies on a physical model of the vocal tract, the auditory system and the agent's motor control (Figure 6.2.A), as well as vocalizations of social peers. Computational experiments show how such a mechanism can explain the adaptive transition from vocal self-exploration with little influence from the ambient speech environment, to a later stage where vocal exploration becomes influenced by vocalizations of peers (Figure 6.2.D). Within the initial self-exploration phase, a sequence of vocal production stages self-organizes (Figure 6.2.C), and shares properties with data from infant developmental psychology (Figure 6.2.E): the vocal learner first discovers how to control phonation, then focuses on vocal variations of unarticulated sounds, and finally automatically discovers and focuses on babbling with articulated proto-syllables. As the vocal learner becomes more proficient at producing complex sounds, imitating vocalizations of peers starts to provide high learning progress explaining an automatic shift from self-exploration to vocal imitation.

6.2.4 Discussion

Our main contribution with respect to previous computational models of speech acquisition is that we do not presuppose the existence of successive developmental stages, but rather they can emerge from an intrinsic drive to maximize competence progress in continuous sensorimotor spaces. We showed that vocal developmental stages can self-organize autonomously, from simple sensorimotor activities to more complex ones. The agent starts producing no phonation and unarticulated vocalizations, which are easy to produce because limited in the range of their auditory effects. This can be related

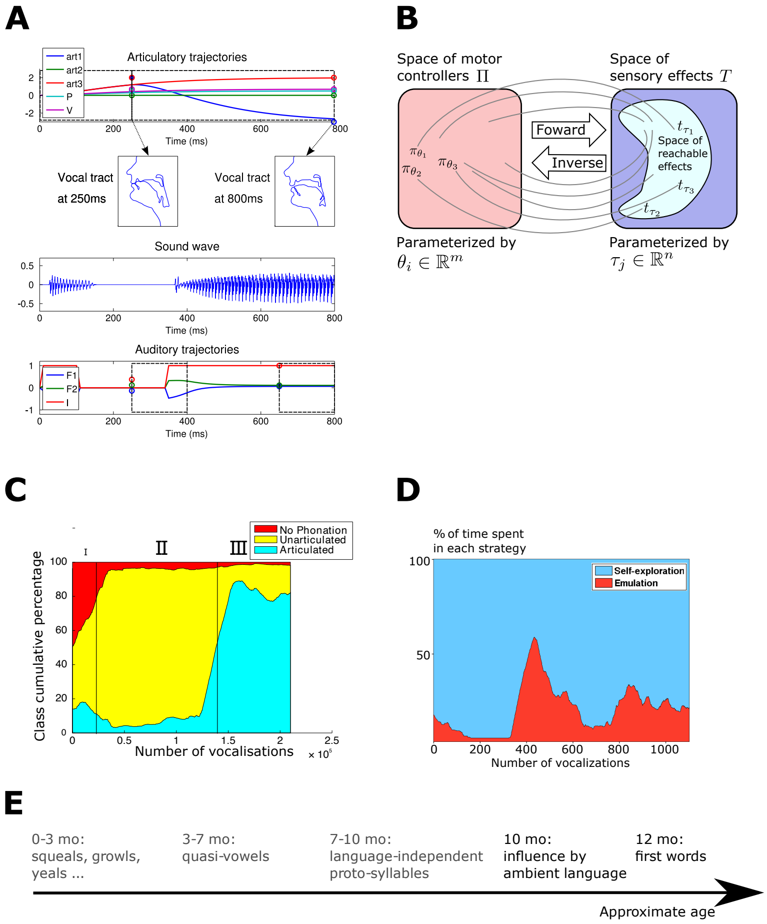


Figure 6.2: The model is able to reproduce the main stages of the developmental sequence observed in infant vocal development. A) The agent produces vocalizations through a realistic model of the human vocal tract. B) The agent learns an inverse model from a goal space T of desired auditory trajectories, to the motor trajectories in Π required to achieve those goals. For this aim it uses a generic intrinsically motivated goal exploration process (IMGEP, see Section 3.2.1.2) where it self-generates its own goals in T and iteratively learns a mapping $T \rightarrow \Pi$. C) When exploring autonomously how articulatory commands produce auditory effects, the model first produces non-speech sounds (no phonation), then unarticulated sounds and finally articulated sounds on the form of proto-syllables. D) When the agent has the choice to either self-explore its vocal abilities or imitate sounds from an ambient language, it chooses to imitate only once it has acquired sufficient motor control over its vocal tract. This choice is made by sampling the strategy (self-exploration or imitation) where it observes maximal learning progress. E) These results are coherent with the main developmental stages observed in infant vocal development (Kuhl, 2004).

to the first stage in infant vocal development (Figure 6.2.E), where the agent produces non speech-sounds (e.g., growls, squeals...) before learning phonation and then produces not well-articulated quasi-vowels. Later on, once the agent does not progress much in producing unarticulated vocalizations, it focuses on more complex vocalizations of the articulated class. The reason is that, due to the properties of the sensorimotor system and internal model, the mastering of complex tasks require first the mastering of simpler tasks in order to yield significant competence progress, so that these complex tasks are selected as interesting goals.

We also showed that intrinsically motivated exploration can lead to a progressive interest toward the sounds of the ambient language. Whereas the first vocalizations are mainly the result of self-exploration, they progressively lead to mastering necessary speech principles (e.g., phonation). This progressive mastering allows in turn to make significant progress in adult-speech imitation, which explains why the vocal learner starts to choose more often as targets the sound of its environment. Competence-progress based curiosity-driven exploration could thus explain a progressive influence of the ambient language on infant vocalizations.

We therefore showed that intrinsically motivated active exploration can self-organize a coherent developmental sequence, without any external clock or preset specification of this sequence. This possible role of intrinsic motivation, providing a mechanism to discover autonomously necessary developmental stages to structure the learning process, is here validated computationally. We believe that it could be of major interest for understanding the structuration of early vocal development in infants (see Oller, 2000, for theoretical arguments on the role of intrinsic motivation in vocal development). Moreover, active exploration can spontaneously generate diverse behaviors from modality-independent and task-independent internal drives. Such spontaneous behavior can result in vocal activity that may have bootstrapped the evolution of an articulated vocal language in the human species (Oudeyer & Smith, 2016).

6.3 COSMO: A Bayesian modeling framework for studying speech communication and the emergence of phonological systems

6.3.1 Context

This work was realized in 2007-2012 during my PhD thesis at GIPSA-Lab at Grenoble University (France), in collaboration with Jean-Luc Schwartz, Julien Diard and Pierre Bessière. It was partially funded by the French ministry of research.

The main publication presenting this contribution has been published as the target article of a Special Issue in *Journal of Phonetics*:

Moulin-Frier, C., Diard, J., Schwartz, J.-L., and Bessière, P.
COSMO ('Communicating about Objects using Sensory-Motor Operations'): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems.
Journal of Phonetics, 53:5–41, 2015a.
ISSN 00954470.
doi: 10.1016/j.wocn.2015.06.001

Open access version of this publication is available at https://www.sciencedirect.com/science/article/pii/S0095447015000352. The rest of this section is a summary of this paper.

6.3.2 Relevance to the proposed ORIGINS framework

Figure 6.3 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. Here we consider a population of interacting agents, each equipped with a model of the human vocal tract and the ear. In terms of **adaptability**, each agent is equipped with a Bayesian cognitive architecture allowing to learn a joint probability distribution of motor commands to its vocal tract, produced vocal sounds and perceived objects in the environment. This distribution is learned through repeated interactions with other agents and objects, with minimal **environmental complexity**. **Multi-agent dynamics** is modeled as *deictic games*, where agents share their attention on objects of the environment and produce and perceive their vocalizations. We study how a shared sound system for communicating about the surrounding objects can self-organize out of the interactions between adaptive agents, thus forming a **cultural repertoire** in the form of a proto-language.

6.3.3 Abstract

While the origin of language remains a somewhat mysterious process, understanding how human language takes specific forms appears to be accessible by the experimental method. Languages, despite their wide variety, display obvious regularities. This contribution attempts to derive some properties of phonological systems (the sound systems for human languages) from computer simulations of interacting learning agents. It in-

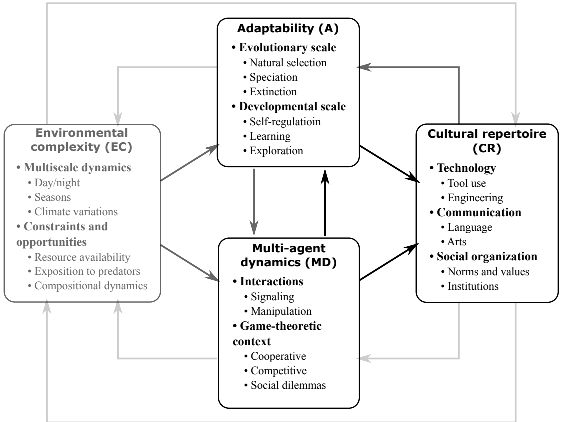


Figure 6.3: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

troduces a model of the cognitive architecture of a communicating agent, called COSMO (for "Communicating about Objects using Sensory-Motor Operations") that allows a probabilistic expression of the main theoretical trends found in the speech production and perception literature. This enables a computational comparison of these theoretical trends, which helps to identify the conditions that favor the emergence of linguistic codes. It provides realistic simulations of phonological system emergence where sensory-motor agents equipped with a computer model of the human vocal tract and the ears interact together and learn to associate speech sounds with the objects they perceive in the environment. COSMO can predict the main regularities in vowel, consonant and syllable systems in human languages.

6.3.4 Methods

6.3.4.1 Cognitive architecture

The scenario that we propose for the emergence of speech communication starts from the inherent structure of a communication situation. In COSMO, speech communication (C) is a success when an object O_S in the speaker's mind is transferred, via sensory and motor mechanisms S and M, to the listener's mind, where it is correctly recovered as O_L (Figure 6.4.A).

The central hypothesis of the COSMO model is that a communicating agent, which is both a speaker and a listener, is able to internalize fully the structure of the communication situation inside an internal model (Figure 6.4.B).

This "internalization" hypothesis therefore results in an agent cognitive architecture (COSMO agent box in Figure 6.4) combining:

- a motor system able to associate communication objects O_S with motor gestures M,

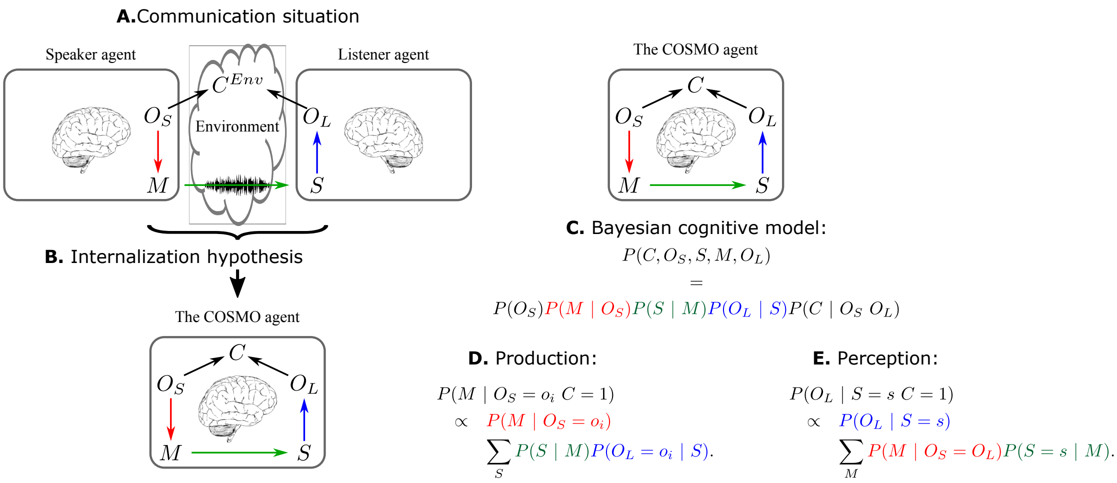


Figure 6.4: The COSMO model. **A.** Schema of the speech communication situation. **B.** The cognitive architecture of a COSMO agent is based on the internalization hypothesis. **C.** Bayesian model of a COSMO agent as a joint probability distribution. **D** and **E.** Production and perception are expressed as Bayesian inferences in the joint distribution.

- an auditory system able to associate communication objects O_L with auditory stimuli S,
- a sensory-motor link able to associate motor gestures M with auditory stimuli S, (providing an internal model of the articulatory-to-acoustic transformation), and
- a fusion system able to associate the communication objects in both the motor (O_S) and auditory (O_L) branches through C.

This internalization hypothesis could be discussed within the framework of general cognitive theories of social communication and human evolution (Baron-Cohen, 1997; Tomasello et al., 2005) (see also (Moore, 2007) for similar views about internalization, expressed in a control theory framework).

The Bayesian model of COSMO is a direct translation of the conceptual cognitive architecture into the joint distribution $P(C, O_S, S, M, O_L)$ (Figure 6.4.C), which can be interpreted as the structure of a Bayesian Network model. (The COSMO acronym, for *"Communicating about Objects using Sensory-Motor Operations"*, also refers the five variables of the model: C, O_S, S, M, O_L.)

An agent with this Bayesian architecture possesses a model of the entire communication situation and is therefore able to perform both production (Figure 6.4.D) and perception (Figure 6.4.E) tasks as Bayesian inferences ($P(M \mid O_S\ C = 1)$ and ($P(O_L \mid S\ C = 1)$, respectively) on the joint distribution.

6.3.4.2 Emergence of phonological systems in populations of deictic agents

Following the language-game paradigm introduced by Steels (Steels, 1997) and applied to phonological systems (Berrah et al., 1996; Berrah, 1998; de Boer, 2000; Oudeyer, 2005), we simulate societies of evolving communicating agents by exploiting language games, based on deixis, that we call "deictic games" and that may be considered as a variant of Steels' "naming games" (Steels, 1994, 1997).

In a deictic game, two agents communicate in front of a given "object" on which their attention is jointly focused via a deictic process of some kind, such as pointing (Figure 6.5.A).

Each agent embeds it own copy of the COSMO Bayesian model, with no prior knowledge in the motor and auditory subsystems ($P(M \mid O_S)$ and $P(O_L \mid S)$ distributions respectively). Each agent will iteratively update the distributions of its own subsystems through it interaction with other agents and various objects (see Moulin-Frier et al. (2015a) for computational detail).

6.3.4.3 Results

We ran large-scale simulations of COSMO agents interacting with objects through deitic games. We studied how their coupled learning processes can self-organize speech sound systems shared among the population to name the objects they interact with. We found that the structure of the emerged speech sound systems share the main statistical tendency the structure of phonological systems in human languages (Figure 6.5.B). In the paper (Moulin-Frier et al., 2015a), we analyze in detail the emergence of vowel, consonant and syllable systems.

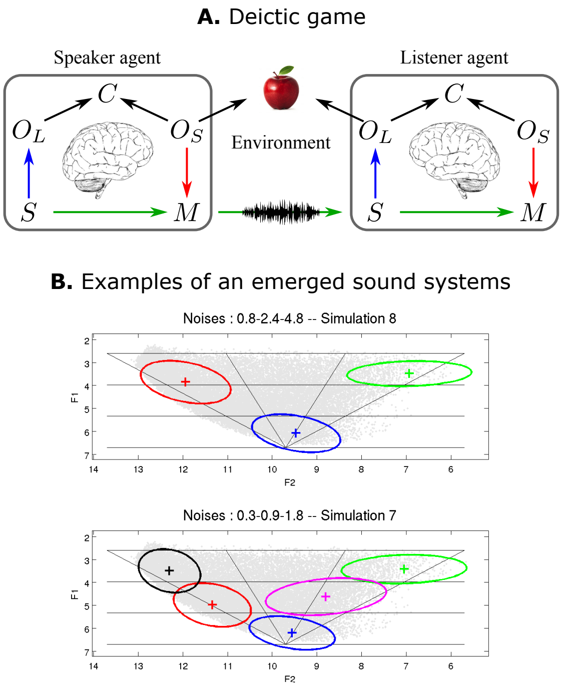


Figure 6.5: **A.** Illustration of a deitic game, where two agents share their attention on the same object. The speaker agent produces a motor gesture M for that object using its production model (Figure 6.4.D). This produces in a sound wave S perceived by a listener agent, which infers the corresponding object using its perception model (Figure 6.4.E). **B.** Examples of two speech sound systems emerging from repeating many deictic games in an agent population with several objects, where each agent alternates between the speaker and the listener role in a random fashion. These sound systems are structurally coherent with the two main vowel systems used in human world language. For example, /i,a,u/ (top) and /i,e,a,o,u/ (bottom) are very common in human languages as well as in COSMO simulations (see Moulin-Frier et al., 2015a, for the full statistics of emerged phonological systems in COSMO).

6.4 Other contributions and future work

6.4.1 Other contributions

Studying the formation of a cultural repertoire (i.e. level-2 dynamics) has occupied a large part of my scientific carrier, especially between 2007 and 2015. In addition to the two contributions presented in this chapter, my other contributions on the topic include:

- Emergence of shared sensory-motor graphical language from visual input (Lemesle et al., 2022),
- Emergent communication as a solution to the architect-builder problem (Barde et al., 2022),
- Modeling the formation of social conventions from embodied real-time interactions (Freire et al., 2020),
- Self-organization of turn-taking behavior in RL agent populations (Moulin-Frier et al., 2015b),
- A unified probabilistic framework (Moulin-Frier & Oudeyer, 2013a) and an open-source library (Moulin-Frier et al., 2014b) for exploration strategies in developmental robotics,
- Cognitive architectures for speech production and perception (Moulin-Frier et al., 2012; Moulin-Frier & Arbib, 2013),
- Distinguishing self, other, and autonomy from visual feedback (Demirel et al., 2021),
- Theoretical propositions on the origins of communication (Moulin-Frier & Oudeyer, 2021; Moulin-Frier & Verschure, 2016) and consciousness (Arsiwalla et al., 2018) in animals and machines.

6.4.2 Future work

Level-2 dynamics considers pre-existing mechanisms for **adaptability** (emerged from level-1 dynamics, for example learning and exploration) as well as for **multi-agent dynamics** (e.g. cooperation and competition pressures with other agents in a shared environment). The objective is to study how the interaction between these mechanisms can bootstrap the formation of a **cultural repertoire**. The main research question at this level is: *What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire?*

Based on the theoretical analysis of level-2 dynamics in Section 3.5 as well as my existing contributions presented in the current chapter, future work will focus on the following directions.

6.4.2.1 Emergent language as a solution to align goals in agent populations

The IMGEP framework introduced by our research group (see Section 3.2.1.2) encompasses a family of learning mechanisms with which an agent autonomously defines its own goals and self-supervises their achievement. It allows to implement learning agents autonomously discovering a large diversity of goals and associated behaviors to achieve those goals (i.e. autotelic agents). Extending this framework to multi-agent settings would enable the study of emergent language as a solution to align the goals of multiple autotelic agents within a group. The MARL community have recently studied multi-goal multi-agent RL, but with externally provided goals that are already aligned (Yang et al., 2019; Wang et al., 2020). We believe that by combining the benefits of intrinsically motivated goal exploration and multi-agent learning, we will be able to improve the generalization abilities of groups of agents in diverse *game-theoretic contexts*. Our proposal is grounded in theories from human studies suggesting that shared intentionality is a key element in achieving human-like cooperation (Tomasello & Carpenter, 2007).

6.4.2.2 Open-ended collective innovation

In recent work (Nisioti et al., 2022, see Section 7.3) we have proposed the concept of *innovation tasks*, where groups of RL agents explore how to combine existing elements to create new ones through compositional dynamics in the environment. Innovation tasks are often characterized by a deceptive optimization landscape with strong local optima and usually require collective exploration to be solved (Derex & Boyd, 2016). Recent human studies of the innovation abilities of groups (Mason et al., 2008; Muthukrishna & Henrich, 2016; Derex & Boyd, 2016) have revealed that properties of the task, such as the existence of compositional dynamics and path-dependence, as well as the form of social connectivity, significantly influence the group's performance. An intriguing observation is that fully-connected structures can lead to suboptimal performance due to the over-exploitation of learned behaviors, while partial connectivity can help achieve the right balance between exploration and exploitation. Based on these observations, we have recently tested this hypothesis in a non-embodied simulation environment, the wordcraft playground (Jiang et al., 2020), a simplified version of the Little Alchemy game (Figure 5.13, left) where DRL agents continuously crafts elements out of existing ones (Nisioti et al., 2022, see Section 7.3). We now plan to extend these experiments in an embodied environment with compositional dynamics as described in Section 5.6.2.1 and will explore the possibility to meta-learn the social connectivity structure itself using bilevel optimization algorithms as proposed in Section 5.6.2.2, potentially in the context of the multi-agent IMGEP framework proposed in Section 6.4.2.1.

Chapter 7

Towards human-like open-ended skill acquisition through cultural feedback effects (Level-3 dynamics)

Contents

7.1 Introduction

In this chapter, I present some of my contributions and propose future work in the modeling of level-3 dynamics. As presented in Section 3.6, this part of the proposed framework is interested in the feedback effects from a **cultural repertoire** to all the other components of the system. The main research question at this level is: *What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition?*

In Section 7.2, I first present a contribution showing how the compositionality of language can be used as a cognitive tool for imagining novel creative goals, thus augmenting the learning and exploration abilities of reinforcement learning agents. Then, in Section 7.3, I present a recent contribution studying the dynamics of open-ended technological innovation in groups of adaptive agents sharing experiences with each other. Finally, in Section 7.4, I mention my other contributions on this topic and propose concrete future work.

7.2 Language as a Cognitive Tool to Imagine Goals in Curiosity-Driven Exploration

7.2.1 Context

This work was realized in 2020-2021 in the Flowers group at Inria (France), in collaboration with Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Peter Dominey, Pierre-Yves Oudeyer. Cédric Colas and Tristan Karch were partly funded by the French Ministère des Armées - Direction Générale de l'Armement. Nicolas Lair was supported by ANRT/CIFRE contract No. 151575A20 from Cloud Temple.

The main publication presenting this contribution has been published at the NeurIPS 2020 conference:

Colas, C., Karch, T., Lair, N., Moulin-Frier, C., Dussoux, J.-M., Dominey, P. F., and Oudeyer, P.-Y.
Language as a Cognitive Tool to Imagine Goals in Curiosity Driven Exploration.
In *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 2020

Open access version of this publication is available at `https://proceedings.neurips.cc/paper/2020/hash/274e6fcf4a583de4a81c6376f17673e7-Abstract.html`. The rest of this section is a summary of this paper.

Demonstration videos are available at `https://sites.google.com/view/imagine-drl`. The source code of playground environment can be found at `https://github.com/flowersteam/playground_env` and the source code of the IMAGINE architecture `https://github.com/flowersteam/Imagine`.

7.2.2 Relevance to the proposed ORIGINS framework

This contribution proposes a novel intrinsically motivated deep reinforcement learning architecture, called IMAGINE, where an agent learn how to achieve a wide variety of goals through its interaction with a social partner (Figure 7.2). We design a novel environment, called *Playground*, displaying complex dynamics, where various objects can interact together in a compositional way. The learning agent can navigate in the environment and move the objects around, observing their potential interactions. For instance, animals grow when water or food is placed on them, while plants only grow with water. At the end of each episode, a simulated social partner provides linguistic descriptions of the interaction between the agent and the environment (e.g. *you grow the green cat*). The agent interprets these descriptions as potential future goals and learn a reward function and a goal-conditioned policy for achieving them. In a second phase where the social peer is absent, we then study how the agent can leverage the compositionality of language to invent new creative goals. We show how this mechanism of goal imagination improves generalization and exploration over agents lacking this capacity.

Figure 7.1 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. **Environmental complexity** corresponds here to the compositional dynamics of the *Playground* environment. **Multi-agent dynamics** is limited to the learning agent perceiving the utterances from the social partner. We assume a pre-existing **cultural repertoire**, where a simulated social partner provides descriptions of the interaction between the agent and the environment in language form. The objective of this contribution is to show how the agent can internalize this cultural knowledge, augmenting its exploration and generalization abilities (**adaptability**) through a compositional goal imagination mechanism. In this sense, IMAGINE demonstrates an important influence from the cultural repertoire to the adaptability of a learning agent, mediated through environmental complexity and multi-agent dynamics.

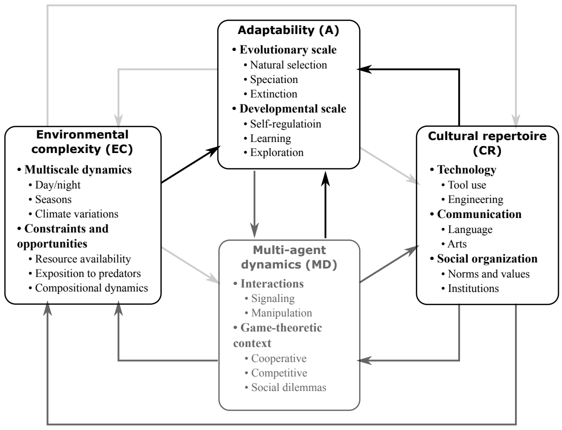


Figure 7.1: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

7.2.3 Abstract

Developmental machine learning studies how artificial agents can model the way children learn open-ended repertoires of skills. Such agents need to create and represent goals, select which ones to pursue and learn to achieve them. Recent approaches have considered goal spaces that were either fixed and hand-defined or learned using generative models of states (Held et al., 2017; Nair et al., 2018; Colas et al., 2019; Pong et al., 2020; Venkattaramanujam et al., 2019; Racaniere et al., 2019). This limited agents to sample goals within the distribution of known effects. We argue that the ability to imagine out-of-distribution goals is key to enable creative discoveries and open-ended learning. Children do so by leveraging the compositionality of language as a tool to imagine descriptions of outcomes they never experienced before, targeting them as goals during play (Tomasello, 2009; Bruner, 1991; Piaget, 1926; Vygotsky, 1978) . We introduce IMAGINE, an intrinsically motivated deep reinforcement learning architecture that models this ability. Such imaginative agents, like children, benefit from the guidance of a social peer who provides language descriptions. To take advantage of goal imagination, agents must be able to leverage these descriptions to interpret their imagined out-of-distribution goals. This generalization is made possible by modularity: a decomposition between learned goal-achievement reward function and policy relying on deep sets, gated attention and object-centered representations. We introduce the Playground environment and study how this form of goal imagination improves generalization and exploration over agents

lacking this capacity. In addition, we identify the properties of goal imagination that enable these results and study the impacts of modularity and social interactions.

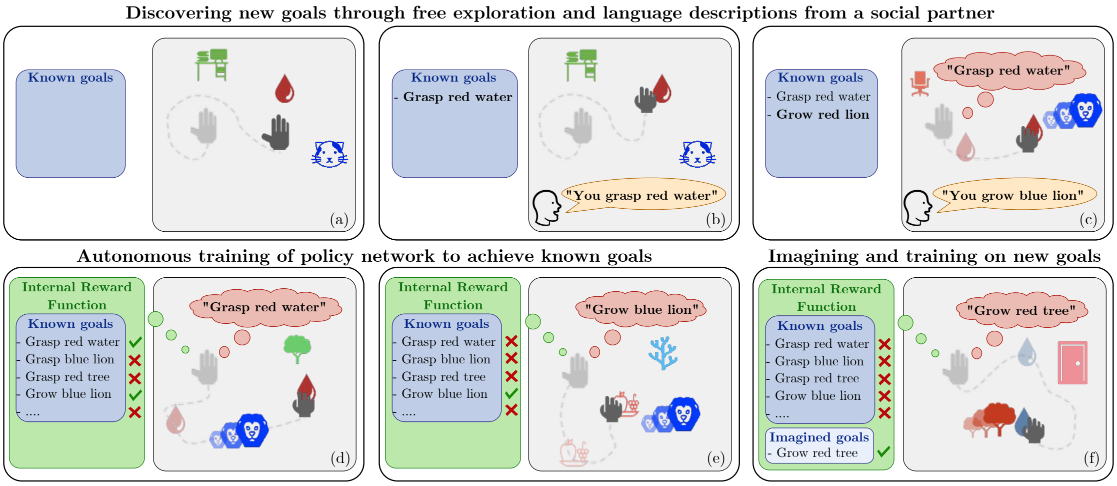


Figure 7.2: **IMAGINE overview**. In the *Playground* environment, the agent (hand) can move, grasp objects and grow some of them. Scenes are generated procedurally with objects of different types, colors and sizes. A social partner provides descriptive feedback (orange), that the agent converts into targetable goals (red bubbles).

7.2.4 Methods

7.2.4.1 General overview

This contribution proposes **I**ntrinsic **M**otivations **A**nd **G**oal **IN**vention for **E**xploration (IMAGINE): a learning architecture which leverages natural language (NL) interactions with a descriptive social partner (SP) to explore procedurally-generated scenes and interact with objects. IMAGINE discovers meaningful environment interactions through its own exploration (Figure 7.2a) and episode-level NL descriptions provided by SP (7.2b). These descriptions are turned into targetable goals by the agent (7.2c). The agent learns to represent goals by jointly training a language encoder mapping NL to goal embeddings and a goal-achievement reward function (7.2d). The latter evaluates whether the current scene satisfies any given goal. These signals (ticks in Figure 7.2d-e) are then used as training signals for policy learning. More importantly, IMAGINE can invent new goals by composing known ones (7.2f). Its internal goal-achievement function allows it to train autonomously on these imagined goals.

7.2.4.2 The IMAGINE Architecture

IMAGINE agents build a repertoire of goals and train two internal models: 1) a goal-achievement reward function $\mathcal{R}$ to predict whether a given description matches a behavioral trajectory; 2) a policy π to achieve behavioral trajectories matching descriptions. The architecture is presented in Figure 7.3 and follows this logic:

1. The *Goal Generator* samples a target goal g_{target} from known and imagined goals $(\mathcal{G}_{\text{known}} \cup \mathcal{G}_{\text{im}})$.
2. The agent (*RL Agent*) interacts with the environment using its policy π conditioned on g_{target}.
3. State-action trajectories are stored in a replay buffer *mem*(π).
4. SP's descriptions of the last state are considered as potential goals $\mathcal{G}_{\text{SP}}(\mathbf{s}_T) = \mathcal{D}_{\text{SP}}(\mathbf{s}_T)$.
5. *mem*($\mathcal{R}$) stores positive pairs $(\mathbf{s}_T,\ \mathcal{G}_{\text{SP}}(\mathbf{s}_T))$ and infers negative pairs $(\mathbf{s}_T,\ \mathcal{G}_{\text{known}} \setminus \mathcal{G}_{\text{SP}}(\mathbf{s}_T))$.
6. The agent then updates:
   - *Goal Gen.*: $\mathcal{G}_{\text{known}} \leftarrow \mathcal{G}_{\text{known}} \cup \mathcal{G}_{\text{SP}}(\mathbf{s}_T)$ and $\mathcal{G}_{im} \leftarrow \text{Imagination}(\mathcal{G}_{\text{known}})$.
   - *Language Encoder* (L_e) and *Reward Function* ($\mathcal{R}$) are updated using data from *mem*($\mathcal{R}$).
   - *RL agent*: We sample a batch of state-action transitions $(\mathbf{s},\ \mathbf{a},\ \mathbf{s}')$ from *mem*(π). Then, we use *Hindsight Replay* and $\mathcal{R}$ to bias the selection of substitute goals to train on (g_{s}) and compute the associated rewards $(\mathbf{s},\ \mathbf{a},\ \mathbf{s}',\ g_{\text{s}},\ r)$. Substituted goals g_s can be known or imagined goals. Finally, the policy and critic are trained via RL.

All algorithmic details are provided in the paper (Colas et al., 2020) and the source code is available at `https://github.com/flowersteam/Imagine`.

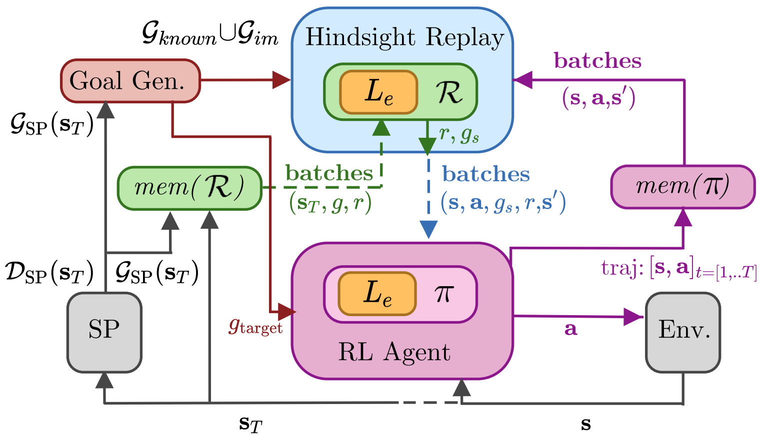


Figure 7.3: **IMAGINE architecture.** Colored boxes show the different modules of IMAGINE. Lines represent update signals (dashed) and function outputs (plain). The language encoder L_e is shared.

7.2.5 Results

We carried out experiments in order to evaluate the benefits from goal imagination in intrinsically motivated learning. Experiments are split into two phases. In the first one, the agent interacts with the social partners, collects descriptions of goals and stores them in a set of known goal descriptions. The agent uses these descriptions paired with its observations in order to learn an internal reward function that detects when the goal represented by the descriptions are achieved in a given scene. Once this internal reward function is obtained, the agent uses its output (the reward signal) in order to train a goal-conditioned policy enabling it to reach any goal. In the second phase, the social partner disappears and the agent starts imagining new goals by composing the descriptions stored in the set of known goals. The agent then targets these new goals and by doing so, discovers new interactions.

We measured the success rate of agents on a wide set of different skills and observed that agents that do not imagine goals (that stop at phase 1) master a smaller set of skills than agents that do imagine goals (Figure 7.4.A).

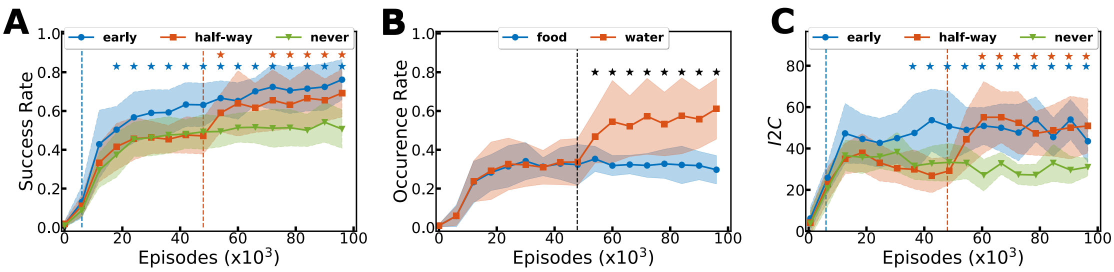


Figure 7.4: **Goal imagination drives exploration and generalization.** Vertical dashed lines mark the onset of goal imagination. **A)** Success rate on testing set depending on when phase 2 starts (*early, half-way* or *never*. **B)** Behavioral adaptation, empirical probabilities that the agent brings supplies to a plant when trying to grow it. **C)** Evaluation metrics (I2C) computed on the testing set. Stars indicate significance (A and C are tested against *never*).

A particular generalization: growing plants. Agents learn to grow animals from SP's descriptions, but are never told they could grow plants. When evaluated offline on the *growing-plants* goals before goal imagination, agents' policies perform a sensible zero-shot generalization and bring them water or food with equal probability, as they would do for animals (Figure 7.4.B, left). As they start to imagine and target these goals, their behavior adapts (Figure 7.4.B, right). If the reward function shows good zero-shot abilities, it only provides positive rewards when the agent brings water. The policy slowly adapts to this internal reward signal and pushes agents to bring more water. We call this phenomenon *behavioral adaptation*. The full paper (Colas et al., 2020) details the generalization abilities of IMAGINE for 5 different types of generalizations involving predicates, attributes and categories.

Exploration. Figure 7.4.C presents the I2C metric (*interesting interaction count*, see Colas et al. (2020) for detail) computed on the set of interactions related to $\mathcal{G}^{\text{test}}$ and demonstrates the exploration boost triggered by goal imagination.

In the full paper (Colas et al., 2020), we provide additional results answering the following questions:

- What if we used other goal imagination mechanisms?
- How does modularity interact with goal imagination?
- Can we use more realistic feedbacks?

7.3 Social Network Structure Shapes Innovation: Experience-sharing in RL with SAPIENS

7.3.1 Context

This work was realized in 2020-2022 as a collaboration between the Flowers group at Inria (France) and Microsoft Research NYC (USA), in collaboration with Eleni Nisioti, Mateo Mahaut, Pierre-Yves Oudeyer and Ida Momennejad. This research was partially funded by the Inria Exploratory action ORIGINS (https://www.inria.fr/en/origins) as well as the French National Research Agency (https://anr.fr/, project ECOCURL, Grant ANR-20-CE23-0006). This work also benefited from access to the HPC resources of IDRIS under the allocation 2020-[A0091011996] made by GENCI, using the Jean Zay supercomputer.

The main publication presenting this contribution has been recently submitted to a major AI conference and is available as a preprint:

Nisioti, E., Mahaut, M., Oudeyer, P.-Y., Momennejad, I., and Moulin-Frier, C. Social Network Structure Shapes Innovation: Experience-sharing in RL with SAPIENS, 2022

Open access version of this publication is available at https://arxiv.org/abs/2206.05060. The rest of this section is a summary of this paper, part of it being adapted from a Twitter thread by Ida Momennejad: https://twitter.com/criticalneuro/status/1540004222858657792.

The source code of the experiments can be found at https://github.com/eleninisioti/SAPIENS.

7.3.2 Relevance to the proposed ORIGINS framework

Figure 7.5 provides a visual representation of the relevance of this contribution to the ORIGINS framework presented in Chapter 3. **Environmental complexity** is modeled as a compositional environment with various elements that can be combined together to create new elements. Regarding **adaptability** we consider vanilla DQN agents, each learning their own action policy by interacting with the environment in order to maximize their cumulative reward. In terms of **multi-agent dynamics**, agents are placed in several groups and can regularly share their experience (by sharing transitions from their replay buffer) according to different social network topologies. We study how these topologies influence the ability of the agents to solve complex hierarchical innovation tasks, forming a **cultural repertoire**. This cultural repertoire in turn influences both adaptability and multi-agent dynamics through the sharing of experiences and innovations to other groups.

7.3.3 Abstract

The human cultural repertoire relies on innovation: our ability to continuously and hierarchically explore how existing elements can be combined to create new ones. Innovation is not solitary, it relies on collective accumulation and merging of previous

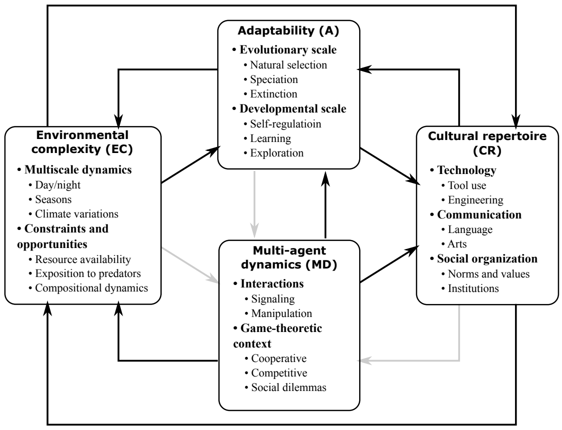


Figure 7.5: Relevance of the contribution to the ORIGINS framework presented in Chapter 3 . Low, medium or high opacity indicates the respective importance of the components (boxes) and their interactions (arrows) in this contribution.

solutions. Machine learning approaches commonly assume that fully connected multi-agent networks are best suited for innovation. However, human laboratory and field studies have shown that hierarchical innovation is more robustly achieved by dynamic communication topologies Derex & Boyd (2016). In dynamic topologies, humans oscillate between innovating individually or in small clusters, and then sharing outcomes with others. To our knowledge, the role of multi-agent topology on innovation has not been systematically studied in machine learning. It remains unclear a) which communication topologies are optimal for which innovation tasks, and b) which properties of experience sharing improve multi-level innovation. Here we use a multi-level hierarchical problem setting (WordCraft), with three different innovation tasks. We systematically design networks of DQNs sharing experiences from their replay buffers in varying topologies (fully connected, small world, dynamic, ring). Comparing the level of innovation achieved by different experience-sharing topologies across different tasks shows that, first, consistent with human findings, experience sharing within a dynamic topology achieves the highest level of innovation across tasks. Second, experience sharing is not as helpful when there is a single clear path to innovation. Third, two metrics we propose, conformity and diversity of shared experience, can explain the success of different topologies on different tasks. These contributions can advance our understanding of optimal AI-AI, human-human, and human-AI collaborative networks, inspiring future tools for fostering collective innovation in large organizations.

7.3.4 Methods

We used a multi-level hierarchical innovation environment (WordCraft, Jiang et al. (2020), illustrated Figure 7.6) with 3 innovation tasks. Each task had an initial set of elements (e.g. Earth, Water). Some elements can be combined to make new elements. Creating a new element moves player one innovation level higher and provides a reward.

We defined 3 innovation tasks (Figure 7.7):

Single innovation path. The recipe book consists of an initial set of 3 base elements ($\mathcal{X}_{valid} = \{a_1, a_2, a_3\}$) and 8 innovation levels. To create the first element an agent needs to combine two of them. To progress further the agent needs to combine the most recently created element with an appropriate one from the initial set. This optimization problem contains a single global optimum.

Merging paths. There are two independent paths, A and B, and at level 2 there is a cross-road that presents the player with three options: moving forward on path A, moving forward on path B, or combining elements from path A and B to progress on path C. The latter path is more rewarding and is, thus, the optimal choice. This innovation task is particularly challenging: to explore path C the player needs to first go to lower innovation levels instead of just progressing on a single path. This task thus exhibits two local optimum (8 elements on A, 8 elements on B) and one global optimum (2 elements on path A + 2 elements on path B + 4 elements on path C).

Best-of-ten paths. Here, there are ten paths, one of which is the most rewarding. The optimal strategy is to move only on the most rewarding path but, to do so, the player must first explore and reject the other paths. This optimization task is characterized by a single global optimum and aims at evaluating the ability of agents to explore effectively in large spaces.

We systematically designed networks of DQN agents sharing experiences from their replay buffers in varying communication topologies: fully connected, small world, ring and dynamic (Figure 7.8). We call this architecture SAPIENS; for Structuring multi-Agent toPology for Innovation through ExperieNce Sharing.

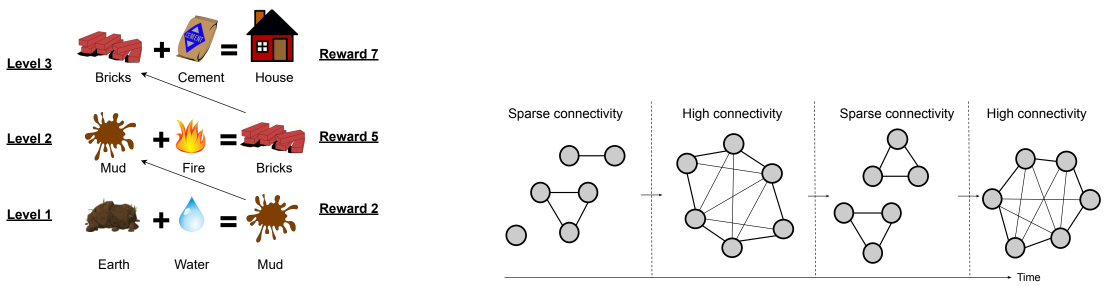


Figure 7.6: (Left) Illustration of an innovation task: the task consists of an initial set of elements (Earth, Water) and a recipe book that determines which element combinations create new elements. Some elements, such as Earth + Mud, cannot be combined. Upon creating a new element the player moves one innovation level higher and receives a reward that increases monotonically with levels. (Right) Dynamic social network structures oscillate between phases of low connectivity, where experience sharing takes place within clusters, and high connectivity, where experiences spread between clusters.

7.3.5 Results

Comparing the level of innovation achieved by different experience-sharing topologies across different tasks shows that (Figure 7.9):

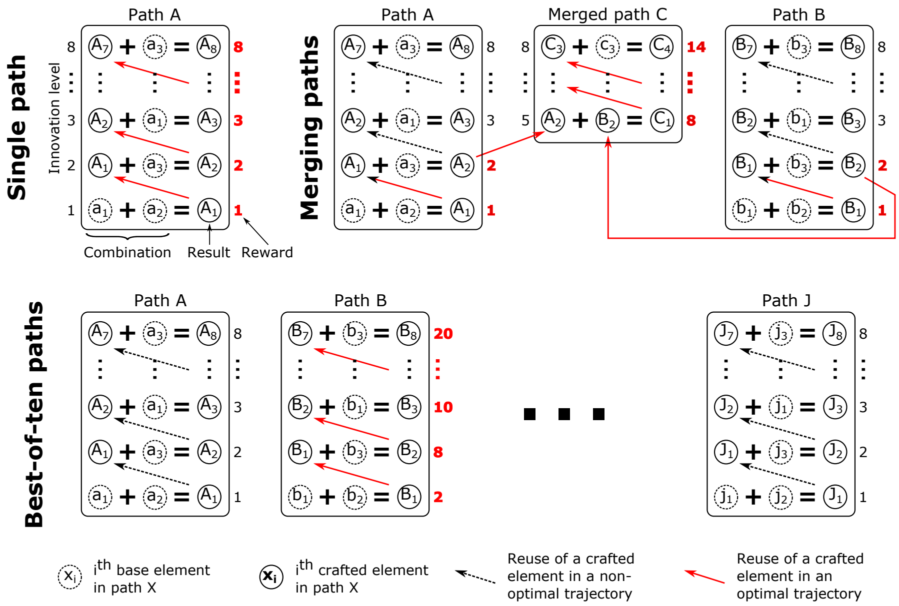


Figure 7.7: We evaluate our algorithm on three innovation tasks called single path, merging paths and best-of-ten paths. Each task contains one or more paths, labeled by an uppercase letter (A to J). Each path X has its own initial set of three base elements $\{x_1, x_2, x_3\}$, which are represented in dashed circles. Crafted elements in path X are represented in upper case (X_i) in solid circles. Optimal trajectories for each tasks are represented by solid red arrows, with their corresponding reward in bold red.

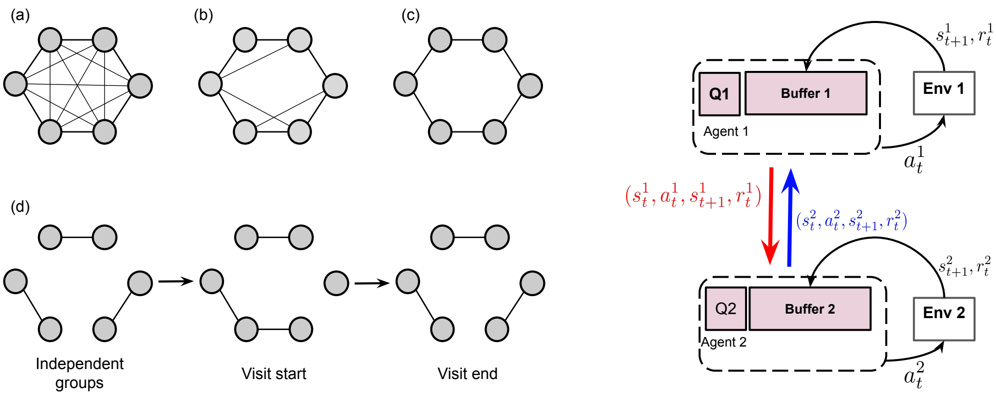


Figure 7.8: (Left) Visualizing communication graphs (a) fully-connected (b) small-world (c) ring (d) dynamic. (Right) Schematic of two neighboring DQNs sharing experiences: agent 1 shares experiences from its own replay buffer to that of agent 2 (red arrow) and vice versa (blue arrow) while both agents are independently collecting experiences by interacting with their own copy of the environment.

- Consistent with human findings a dynamic topology of experience sharing achieves highest level of innovation across tasks
- Experience sharing is not as helpful when there's a single clear path to innovation

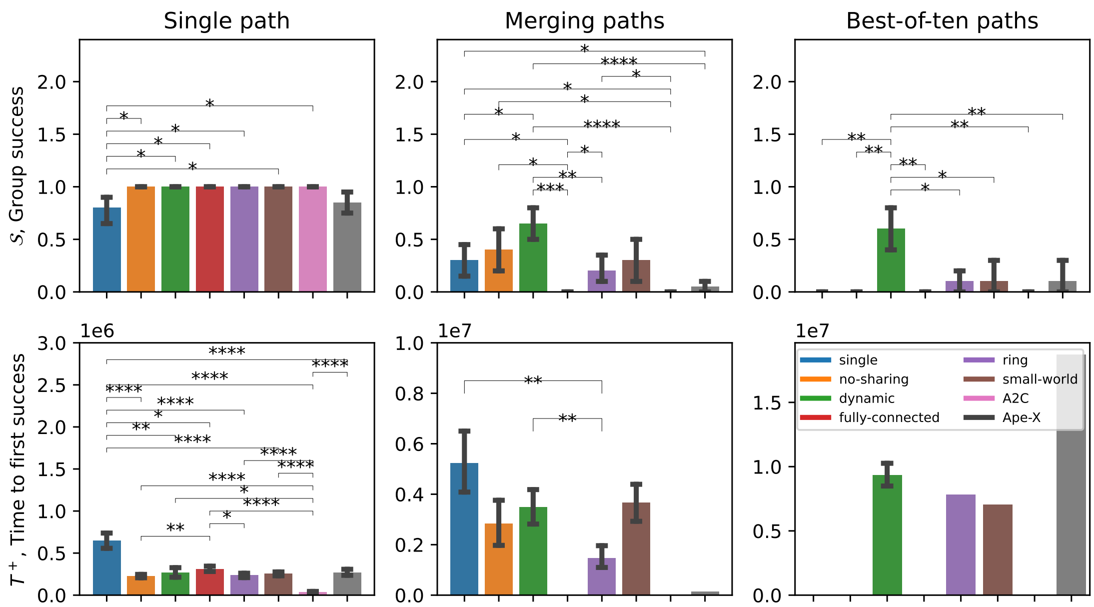


Figure 7.9: Overall comparison of performances for the single path task (first column), merging paths task (second column) and best-of-ten paths task (third column). We present two metrics: $S^{\mathcal{G}}\%$, of group success denotes the percentage of trials in which at least one agent in the group found the optimal solution (top row) and T^{+}, Time to first success, which is the number of training time steps required (second for group success (bottom row). We compare the performance of our proposed social network topologies and baseline distributed RL algorithms (A2C and Ape-X).

Moreover, we propose metrics of shared experience:

- *Conformity* C_t is a group-level metric that denotes the percentage of agents in a group that followed the same trajectory in a give evaluation trial
- *Average Volatility* V_t is an agent-level metric that denotes the cumulative number of changes in the trajectory followed by an agent
- *Average Diversity* D_t (mnemonic metric): number of unique experiences in the replay buffer of an agent, averaged over all agents.

These metrics can explain the success of different topologies on different tasks (Figures 7.10 and 7.11).

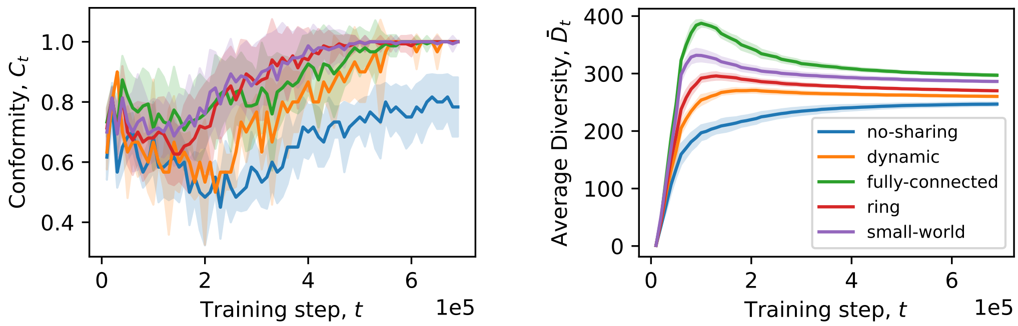


Figure 7.10: Analyzing group behavior in the single path task:: (left) Conformity C_t is a behavioral metric that denotes the percentage of agents in a group that followed the same trajectory in a given evaluation trial (right) Average Diversity $\bar{D}_t$ is a mnemonic metric that denotes the number of unique experiences in the replay buffer of an agent, averaged over all agents

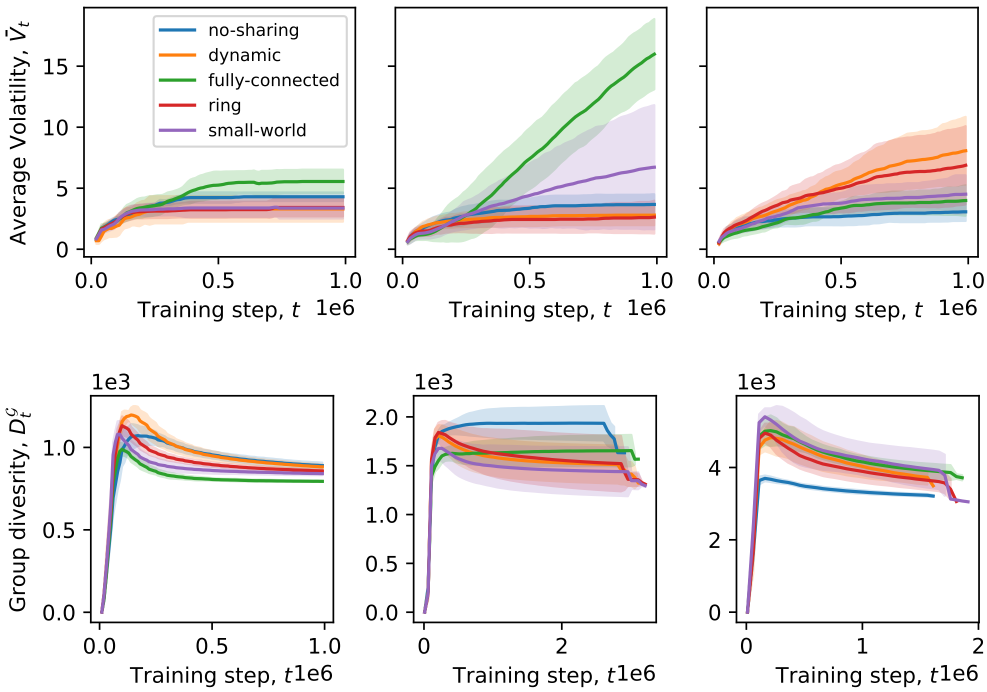


Figure 7.11: Analyzing group behavior in the single path task (first column), merging paths task (second column) and best-of-10 paths task (third column). On the top row, the Average volatility (V_t) is a behavioral metric indicating the cumulative number of changes in the trajectory followed by an agent, averaged by all agents. On the bottom row, Group Diversity $D_t^{\mathcal{G}}$ is a mnemonic metric that captures the diversity of the aggregated group buffer.

7.3.6 Discussion

From tool use and language to music and mathematics, human innovation is characterized by cumulative solutions. Human studies have shown that different communication networks are best suited to solve different innovation tasks. However, most decentralized RL approaches assume static, often fully-connected networks. Here we test the hypothesis that the topology of experience sharing in a group of deep RL agents can shape its performance using our proposed algorithm SAPIENS. In line with human experiments, SAPIENS experiments show that dynamic topologies of experience sharing are best suited to solve complex innovation tasks. Our work extends upon the classical decentralized RL paradigm by explicitly measuring diversity, which has been associated with improved exploration (Horgan et al., 2018; Christianos et al., 2020), and demonstrating that certain multi-agent topologies are good at maintaining it, without requiring agents with different hyper-parameters.

Our study shows that both multi-agent network topology and task structure affect the performance of our proposed SAPIENS architecture. Based on our experimental results, we can provide general recommendations on which topology to use for which task class. The single-path task is an instance of a class of tasks with no strong local optima (similarly to long-horizon tasks as in Gupta et al. (2019)). In this case, our experimental results show no benefit of experience sharing. The merging-path task exhibits strong local optima that have to be explored up to a certain point in order to discover the global optimum (in the spirit of hard exploration tasks as in Baker et al. (2022); Ecoffet et al. (2021)). Our results show that topologies with low initial connectivity (such as no-sharing, small world and dynamic) performs best here by improving the exploration of different innovation paths. The dynamic topology shows the highest performance, allowing different groups to reach the merging innovation level in non-optimal paths before sharing their experience during visits to other groups to find the optimal one. Finally, the best-of-ten task is an instance of a class of tasks with a large search space, many local optima and a few global ones (in the spirit of combinatorial optimization tasks as in Mazyavkina et al. (2021)). In this case, our results show that the dynamic topology performs best, allowing different groups to first explore different paths, then spread the optimal solution to other groups once discovered.

Limitations and future works We currently limit our experiments to symbolic innovation tasks implemented in Wordcraft. We believe that scaling up our study by applying SAPIENS in environments commonly employed by the multi-agent reinforcement learning community to study innovation (Leibo et al., 2019) is an important next step. In the appendix of the paper (Nisioti et al., 2022), we therefore provide results on a deceptive game task implemented in a grid world. While the present implementation considers groups of DQN agents, SAPIENS is in theory agnostic to the choice of off-policy algorithm. In future studies we plan to investigate SAPIENS networks with different types of RL agents. For instance, we could identify what network characteristics (e.g. centrality, brokering, bridging, etc) is best suited for which learning agents across different tasks. We also envision a meta-learning version of SAPIENS where an outer-loop optimizes the social network structure.

Societal implications. Such future studies can have wide societal implications, paving the way toward predicting optimal social network structures for human-AI, AI-AI human-human cooperation networks. In an era of pending climate catastrophes and global pandemics, better understanding collaboration networks appears relevant. We believe SAPIENS can shed light on how to solve a wide range of innovation tasks and lead to future tools for fostering human and machine innovation in large organizations. Particular attention will however be required to prevent potential negative impact, where controlling the dynamics of social network topologies could result in unequal access to information and privacy issues.

7.4 Other contributions and future work

7.4.1 Other contributions

Some of my existing contributions on level-2 dynamics presented in the last chapter (Chapter 6) also consider, to some extent, aspects of level-3 dynamics, i.e. of feedback effects from a cultural repertoire on other components of the system. It is in particular the case of most contributions studying the emergence of proto-cultural structures such as communication systems (Lemesle et al., 2022; Barde et al., 2022; Moulin-Frier et al., 2015b) and social norms (Freire et al., 2020) in agent populations. In these contributions, emerging proto-cultural structures necessarily influence adaptability and multi-agent dynamics. This illustrates the strong coupling between the three levels of dynamics proposed in the ORIGINS framework.

In addition to these contributions and the two main ones presented in this chapter, my other contributions on the topic also include:

- A recent position paper on Vygotskian Autotelic Artificial Intelligence: Language and Culture Internalization for Human-Like AI (Colas et al., 2022a),
- Grounding spatio-temporal language with transformers (Karch et al., 2021),
- Language acquisition in human-robot interaction (Moulin-Frier et al., 2018),
- Epidemioptim: A toolbox for the optimization of control policies in epidemiological models (Colas et al., 2021).

7.4.2 Future work

Level-3 dynamics is interested in the feedback effects from a **cultural repertoire** to all the other components of the system. The main research question at this level is: *What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition?* .

Based on the theoretical analysis of level-3 dynamics in Section 3.6 as well as my existing contributions presented in the current chapter, future work will focus on the following directions.

7.4.2.1 Non-episodic persistent environments

Most existing environments used in RL relies on an episodic training paradigm, where the state of the environment is regularly reset to an initial configuration, either when the agent has solved the task or after a predefined timeout (Section 3.4.1). This prevents the study of niche construction effects as proposed in Section 3.6, where **adaptability**, **multi-agent dynamics** and the **cultural repertoire** all feed back to **environmental complexity**, bootstrapping a positive feedback loop driving the open-ended aspect of human-like skill acquisition. Studying the role of niche construction mechanisms in OESA therefore requires the design of simulation environments that enable niche construction by avoiding resets and ensuring environmental persistence across generations, allowing agents' behavior to modify their own niche and fitness landscape.

7.4.2.2 Iterated learning as a driver of niche construction

Non-episodic persistent environments proposed above, combined with multiscale dynamics, spatial open-endedness and compositional dynamics as proposed in Section 5.6.2.1, will enable the study of typical niche-constructive behaviors encountered in nature. For example, agents will be able to build some sort of shelters for the winter (keeping themselves warm and protecting resources), which then become part of the environment and create new constraints and opportunities for the next generations. It will enable the empirical evaluation of the recently proposed computational hypothesis that niche construction can be viewed as an iterated meta-learning process, where previous generations increase the saliency of the environment so that future agents interacting with it will encounter a more informative niche (Constant et al., 2018; Kirby et al., 2014) with potential benefits for exploration and survival.

Chapter 8

Conclusion

In this thesis, I have first detailed my scientific trajectory in Chapter 2, from the start of my PhD thesis in 2007 to my recruitment as a permanent researcher in the Flowers group at Inria in 2019. Then in Chapter 3, I have developed in detail the current research program I have initiated since my recruitment as a permanent researcher. It relies on the proposition of a novel conceptual framework, called ORIGINS. This framework aims at *Grounding Artificial Intelligence in the Origins of Human Behavior* through the modeling of complex interactions between environmental, adaptive, multi-agent and cultural dynamics. In particular:

- The framework analyzes relationships between different fields studying open-ended skill acquisition (OESA) in humans (in particular human behavioral ecology) and in AI (Sections 3.2 and 3.3.1),
- It structures an interdisciplinary dialog among these fields as well as a roadmap for progressing towards OESA in AI (Sections 3.3, 3.4, 3.5 and 3.6),
- It considers AI methods as computational tools to better understand the dynamics of the human species at multiple spatiotemporal scales (Section 3.7).

The ORIGINS framework is structured around three levels of dynamics (Figure 3.3), each associated with a main research question.

Level-1 dynamics: What are the ecological conditions favoring the evolution of autotelic agents with intrinsically motivated goal exploration strategies? The theoretical arguments related to this question are developed in Section 3.4. My contributions and future work on the topic are presented in Chapter 5. The main hypothesis at this level is that environmental complexity is the main driver of behavioral diversity, as proposed by recent theories in human behavioral ecology as well as AI experiments in e.g. meta reinforcement learning and curriculum learning. Multi-agent dynamics through feedback effects of co-adaptation also play an important role in promoting behavioral diversity at this level. These mechanisms contribute to the evolution of autotelic agents equipped with an intrinsic motivation to discover and explore new niches and skills as a solution to adapt to abrupt environmental changes.

Level-2 dynamics: What is the joint role of intrinsically motivated goal exploration and multi-agent dynamics in the formation of a cultural repertoire? The theoretical arguments related to this question are developed in Section 3.5. My contributions and future work on the topic are presented in Chapter 6. The main hypothesis at this level is that intrinsically motivated goal exploration coupled with varying cooperation and competition pressures (emerged from level-1 dynamics above) result in the acquisition of complex skills that are not necessarily linked to evolutionary fitness. These contribute to the formation of a cultural repertoire, i.e. an expanding collection of behaviors that are transmitted repeatedly through social or observational learning to become a population-level characteristic. These skills include technology, communication and social organization, originally of relatively low complexity.

Level-3 dynamics: What is the role of feedback effects from a cultural repertoire in the open-ended dynamics of human skill acquisition? The theoretical arguments related to this question are developed in Section 3.6. My contributions and future work on the topic are presented in Chapter 7. The main hypothesis at this level is that a cultural repertoire, when reaching a certain level of complexity, can bootstrap positive feedback loops continuously increasing the complexity of the environment, of the agents' cognitive abilities and of their multi-agent dynamics. These feedback loops, which are especially prominent in the human species due to the complexity of our cultural repertoire, are what make human skill acquisition truly open-ended.

In the discussion at the end of Chapter 3 (Section 3.7) I detail the interdisciplinary aspects and the limitations of the proposed ORIGINS framework.

In the longer term, the objective is to study the interactions between these levels in an integrated simulation environment. This will require the integration of (1) complex environmental dynamics inspired by paleo-climatology and human behavioral ecology, providing diverse constraints and opportunities through compositional dynamics (Section 5.6.2) ; (2) artificial evolution able to generate morphogenetic and neural structures with diverse functionalities, including the potential evolution of autotelic agents (Section 5.6.2) (3) reciprocal influences between agent populations modifying their environmental niches (Section 6.4.2) ; (4) cross-generational influence mediated by environmental changes and agent's interactions as a bootstrap of cultural evolution (Section 7.4.2).

I believe that simulating these mechanisms is the most promising path toward achieving human-like open-ended skill acquisition in artificial systems. Beyond this perspective for computer science, I also believe that the ORIGINS framework can help us to better understand the dynamics of our own species within its eco-evolutionary and cultural context (see *a humanist perpective on AI* in Section 3.7).

Bibliography

Alet, F., Schneider, M. F., Lozano-Perez, T., and Kaelbling, L. P.
Meta-learning curiosity algorithms.
In *International Conference on Learning Representations (ICLR 2020)*, 2020.

Anastassacos, N., García, J., Hailes, S., and Musolesi, M.
Cooperation and reputation dynamics with reinforcement learning.
pp. 9, 2021.

Arbib, M. A.
From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics.
Behavioral and Brain Sciences, 28:105–167, 2005.

Arbib, M. A. and Moulin-Frier, C.
Recognizing speech in a novel accent: The motor theory of speech perception reframed.
In *Neurobiology of Language Conference, San Diego, USA*, 2010.

Arsiwalla, X. D., Sole, R., Moulin-Frier, C., Herreros, I., Sanchez-Fibla, M., and Verschure, P.
The Morphospace of Consciousness.
arXiv:1705.11190 [cond-mat, physics:physics, q-bio], November 2018.

Baker, B., Gupta, O., Naik, N., and Raskar, R.
Designing neural network architectures using reinforcement learning.
arXiv:1611.02167 [cs], 2017.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., and Mordatch, I.
Emergent Tool Use From Multi-Agent Autocurricula.
In *International Conference on Learning Representations*, 2020.

Baker, B., Akkaya, I., Zhokhov, P., Huizinga, J., Tang, J., Ecoffet, A., Houghton, B., Sampedro, R., and Clune, J.
Video PreTraining (VPT): Learning to act by watching unlabeled online videos.
arXiv preprint arXiv: Arxiv-2206.11795, 2022.

Baldassarre, G. and Mirolli, M. (eds.).
Intrinsically Motivated Learning in Natural and Artificial Systems.
Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
ISBN 978-3-642-32374-4.
doi: 10.1007/978-3-642-32375-1.

Bansal, T., Pachocki, J., Sidor, S., Sutskever, I., and Mordatch, I.
Emergent Complexity via Multi-Agent Competition.
In *International Conference on Learning Representations*, 2018.

Banzhaf, W., Baumgaertner, B., Beslon, G., Doursat, R., Foster, J. A., McMullin, B., de Melo, V. V., Miconi, T., Spector, L., Stepney, S., and White, R.
Defining and simulating open-ended novelty: Requirements, guidelines, and challenges.
Theory in Biosciences, 135(3):131–161, September 2016.
ISSN 1611-7530.
doi: 10.1007/s12064-016-0229-7.

Baranes, A. and Oudeyer, P.-Y.
Active Learning of Inverse Models with Intrinsically Motivated Goal Exploration in Robots.
Robotics and Autonomous Systems, 61(1):49–73, 2013.
doi: 10.1016/j.robot.2012.05.008.

Barde, P., Karch, T., Nowrouzezahrai, D., Moulin-Frier, C., Pal, C., and Oudeyer, P.-Y.
Learning to Guide and to Be Guided in the Architect-Builder Problem.
In *Tenth International Conference on Learning Representations (ICLR 2022)*, 2022.

Baron-Cohen, S.
Mindblindness: An Essay on Autism and Theory of Mind.
MIT press, 1997.
ISBN 0-262-52225-X.

Barto, A., Singh, S., and Chenatez, N.
Intrinsically Motivated Learning of Hierarchical Collections of Skills.
In *Proc. 3rd Int. Conf. Dvp. Learn.*, pp. 112–119, San Diego, CA, 2004.

Barto, A. G.
Intrinsic Motivation and Reinforcement Learning.
In Baldassarre, G. and Mirolli, M. (eds.), *Intrinsically Motivated Learning in Natural and Artificial Systems*, pp. 17–47. Springer, Berlin, Heidelberg, 2013.
ISBN 978-3-642-32375-1.
doi: 10.1007/978-3-642-32375-1_2.

Bateson, G.
Culture Contact and Schismogenesis.
Man, 35:178–183, 1935.
ISSN 0025-1496.
doi: 10.2307/2789408.

Beer, R. D.
Autopoiesis and Cognition in the Game of Life.
Artificial Life, 10(3):309–326, July 2004.
ISSN 1064-5462.
doi: 10.1162/1064546041255539.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., and Munos, R.
Unifying count-based exploration and intrinsic motivation.
In *Advances in Neural Information Processing Systems*, pp. 1471–1479, 2016.

Bellemare, M. G., Naddaf, Y., Veness, J., and Bowling, M.
The Arcade Learning Environment: An Evaluation Platform for General Agents.
Journal of Artificial Intelligence Research, 47:253–279, June 2013.
ISSN 1076-9757.
doi: 10.1613/jair.3912.

Bergmüller, R. and Taborsky, M.
Animal personality due to social niche specialisation.
Trends in Ecology & Evolution, 25(9):504–511, September 2010.
ISSN 0169-5347.
doi: 10.1016/j.tree.2010.06.012.

Berk, L. E.
Why Children Talk to Themselves.
Scientific American, 1994.

Berlyne, D. E.
A theory of human curiosity.
British Journal of Psychology, 45:180–191, 1954.

Berrah, A.-R.
Evolution d'une Société Artificielle d'agents de Parole : Un Modéle Pour l'émergence Des Structures Phonétiques.
PhD thesis, Institut National Polytechnique de Grenoble, 1998.

Berrah, A.-R., Glotin, H., Laboissière, R., Bessière, P., and Boë, L.-J.
From Form to Formation of Phonetic Structures: An evolutionary computing perspective.
In Fogarty, T. and Venturini, G. (eds.), *ICML '96 Workshop on Evolutionary Computing and Machine Learning*, pp. 23–29, Bari, 1996.

Borgerhoff Mulder, M. and Schacht, R.
Human behavioural ecology.
e LS, 2001.

Botero, C. A., Gardner, B., Kirby, K. R., Bulbulia, J., Gavin, M. C., and Gray, R. D.
The ecology of religious beliefs.
Proceedings of the National Academy of Sciences, 111(47):16784–16789, November 2014.
ISSN 0027-8424, 1091-6490.
doi: 10.1073/pnas.1408701111.

Brooks, R. A.
Intelligence without representation.
Artificial Intelligence, 47(1-3):139–159, 1991.
ISSN 00043702.
doi: 10.1016/0004-3702(91)90053-M.

Brown, G. R., Dickins, T. E., Sear, R., and Laland, K. N.
Evolutionary accounts of human behavioural diversity.
Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1563): 313–324, February 2011.
ISSN 0962-8436.
doi: 10.1098/rstb.2010.0267.

Brown, J. L.
Optimal group size in territorial animals.
Journal of Theoretical Biology, 95(4):793–810, April 1982.
ISSN 0022-5193.
doi: 10.1016/0022-5193(82)90354-X.

Bruner, J.
The narrative construction of reality.
Critical inquiry, pp. 1–21, 1991.

Burnham, T. and Johnson, D.
The biological and evolutionary logic of human cooperation.
Analyse & Kritik, 27:113–35, December 2005.
doi: 10.1515/auk-2005-0107.

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., Tani, J., Belpaeme, T., Sandini, G., Nori, F., Fadiga, L., Wrede, B., Rohlfing, K., Tuci, E., Dautenhahn, K., Saunders, J., and Zeschel, A.
Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics.
IEEE Transactions on Autonomous Mental Development, 2(3):167–195, September 2010.
ISSN 1943-0612.
doi: 10.1109/TAMD.2010.2053034.

Carroll, S. M.
The Big Picture: On the Origins of Life, Meaning, and the Universe Itself.
Dutton est. 1852, an imprint of Penguin Random House LLC, New York, New York, 2016.
ISBN 978-0-525-95482-8.

Carruthers, P.
Modularity, Language, and the Flexibility of Thought.
Behavioral and Brain Sciences, (6), 2002.
ISSN 0140-525X, 1469-1825.

Chapman, C. A. and Chapman, L. J.
Constraints on group size in red colobus and red-tailed guenons: Examining the generality of the ecological constraints model.
International Journal of Primatology, 21(4):565–585, 2000.

Chollet, F.
On the Measure of Intelligence, 2019.

Christianos, F., Schäfer, L., and Albrecht, S.
Shared experience actor-critic for multi-agent reinforcement learning.
In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H. (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 10707–10717. Curran Associates, Inc., 2020.

Chu, J. and Schulz, L.
Exploratory play, rational action, and efficient search.
Preprint, PsyArXiv, June 2020.

Clark, A.
Word, Niche and Super-Niche: How Language Makes Minds Matter More.
Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia, 20(3):255–268, 2005.

Clune, J.
AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence.
arXiv:1905.10985 [cs], January 2020.

Co-Reyes, J. D., Sanjeev, S., Berseth, G., Gupta, A., and Levine, S.
Ecological Reinforcement Learning.
arXiv:2006.12478 [cs, stat], June 2020.

Cobbe, K., Klimov, O., Hesse, C., Kim, T., and Schulman, J.
Quantifying generalization in reinforcement learning.
In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1282–1289, Long Beach, California, USA, June 2019. PMLR.

Colas, C., Fournier, P., Chetouani, M., Sigaud, O., and Oudeyer, P.-Y.
CURIOUS: Intrinsically motivated modular multi-goal reinforcement learning.
In Chaudhuri, K. and Salakhutdinov, R. (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 1331–1340. PMLR, June 2019.

Colas, C., Karch, T., Lair, N., Moulin-Frier, C., Dussoux, J.-M., Dominey, P. F., and Oudeyer, P.-Y.
Language as a Cognitive Tool to Imagine Goals in Curiosity Driven Exploration.
In *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 2020.

Colas, C., Hejblum, B., Rouillon, S., Thiébaut, R., Oudeyer, P.-Y., Moulin-Frier, C., and Prague, M.

EpidemiOptim: A Toolbox for the Optimization of Control Policies in Epidemiological Models.
Journal of Artificial Intelligence Research, 71:479–519, July 2021.
ISSN 1076-9757.
doi: 10.1613/jair.1.12588.

Colas, C., Karch, T., Moulin-Frier, C., and Oudeyer, P.-Y.
Vygotskian Autotelic Artificial Intelligence: Language and Culture Internalization for Human-Like AI.
Nature Machine Intelligence (to appear), 2022a.
doi: 10.48550/arXiv.2206.01134.

Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y.
Autotelic Agents with Intrinsically Motivated Goal-Conditioned Reinforcement Learning: A Short Survey.
Journal of Artificial Intelligence Research, 74:1159–1199, 2022b.
ISSN 1076-9757.
doi: 10.1613/jair.1.13554.

Collard, M., Kemery, M., and Banks, S.
Causes of Toolkit Variation Among Hunter-Gatherers: A Test of Four Competing Hypotheses.
pp. 20, 2005.

Collard, M., Buchanan, B., Ruttle, A., and O'Brien, M. J.
Niche Construction and the Toolkits of Hunter–Gatherers and Food Producers.
Biological Theory, 6(3):251–259, September 2011.
ISSN 1555-5550.
doi: 10.1007/s13752-012-0034-6.

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., and Friston, K. J.
A variational approach to niche construction.
Journal of The Royal Society Interface, 15(141):20170685, 2018.
ISSN 1742-5689, 1742-5662.
doi: 10.1098/rsif.2017.0685.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R., and Rice, K.
Consensus on consensus: A synthesis of consensus estimates on human-caused global warming.
Environmental Research Letters, 11(4):048002, April 2016.
ISSN 1748-9326.
doi: 10.1088/1748-9326/11/4/048002.

Csikszentmihalyi, M.
Creativity: Flow and the Psychology of Discovery and Invention.
HarperCollins, 1997.
ISBN 978-0-06-092820-9.

Cully, A., Clune, J., Tarapore, D., and Mouret, J.-B.
Robots that can adapt like animals.
Nature, 521(7553):503–507, May 2015.
ISSN 1476-4687.
doi: 10.1038/nature14422.

Cuypers, T. D., Rutten, J. P., and Hogeweg, P.
Evolution of evolvability and phenotypic plasticity in virtual cells.
BMC Evolutionary Biology, 17(1):60, February 2017.
ISSN 1471-2148.
doi: 10.1186/s12862-017-0918-y.

Darwin, C.
On the Origin of Species by Means of Natural Selection or the Natural Selection of Favoured Races in the Struggle for Life.
New York: D. Appleton and Company, 1859.

de Boer, B.
Self-organization in vowel systems.
Journal of Phonetics, 28(4):441–465, 2000.
ISSN 0095-4470.
doi: 10.1006/jpho.2000.0125.

De Jaegher, H. and Di Paolo, E.
Participatory sense-making.
Phenomenology and the Cognitive Sciences, 6(4):485–507, December 2007.
ISSN 1572-8676.
doi: 10.1007/s11097-007-9076-9.

deBeaune, S., Davidson, I., Hardy, B., McGrew, W., Marchant, L., Reader, S., Stout, D., Vauclair, J., and DeBeaune, S.
The invention of technology: Prehistory and cognition.
Current Anthropology, 45(2):139–162, 2004.

Deci, E. L. and Ryan, R. M.
Intrinsic Motivation and Self-Determination in Human Behavior.
Plenum Press, New York, 1985.

Demirel, B., Moulin-Frier, C., Arsiwalla, X. D., Verschure, P. F. M. J., and Sánchez-Fibla, M.
Distinguishing Self, Other, and Autonomy From Visual Feedback: A Combined Correlation and Acceleration Transfer Analysis.
Frontiers in Human Neuroscience, 15, 2021.
ISSN 1662-5161.

Dennett, D. C.
Consciousness Explained.
Penguin uk, 1993.

Derex, M. and Boyd, R.
Partial connectivity increases cultural accumulation within groups.
Proceedings of the National Academy of Sciences of the United States of America, 113 (11):2982–2987, March 2016.
ISSN 1091-6490.
doi: 10.1073/pnas.1518798113.

Dolson, E. L., Vostinar, A. E., Wiser, M. J., and Ofria, C. A.
The MODES toolbox: Measurements of open-ended dynamics in evolving systems.
Artificial Life, 25:50–73, 2019.

Dunbar, R. I. M.
Coevolution of neocortical size, group size and language in humans.
Behavioral and Brain Sciences, 16(4):681–694, December 1993.
ISSN 0140-525X, 1469-1825.
doi: 10.1017/S0140525X00032325.

Dunbar, R. I. M.
How conversations around campfires came to be.
Proceedings of the National Academy of Sciences, 111(39):14013, September 2014.
doi: 10.1073/pnas.1416382111.

Ecoffet, A., Huizinga, J., Lehman, J., Stanley, K. O., and Clune, J.
First return, then explore.
Nature, 590(7847):580–586, 2021.

Eppe, M. and Oudeyer, P.-Y.
Intelligent behavior depends on the ecological niche: Scaling up AI to human-like intelligence in socio-cultural environments.
KI - Künstliche Intelligenz, 35(1):103–108, 2021.
ISSN 0933-1875, 1610-1987.
doi: 10.1007/s13218-020-00696-1.

Eskridge, B. E. and Hougen, D. F.
Nurturing promotes the evolution of learning in uncertain environments.
In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pp. 1–6, 2012.
doi: 10.1109/DevLrn.2012.6400847.

Etcheverry, M., Moulin-Frier, C., and Oudeyer, P.-Y.
Hierarchically-Organized Latent Modules for Exploratory Search in Morphogenetic Systems.
In *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 2020.

Etcheverry, M., Wang-Chak Chan, B., Moulin-Frier, C., and Oudeyer, P.-Y.
Meta-Diversity Search in Complex Systems,A Recipe for Artificial Open-Endedness ?, June 2021.

Fan, L., Wang, G., Jiang, Y., Mandlekar, A., Yang, Y., Zhu, H., Tang, A., Huang, D.-A., Zhu, Y., and Anandkumar, A.

MineDojo: Building Open-Ended Embodied Agents with Internet-Scale Knowledge, 2022.

Finn, C., Abbeel, P., and Levine, S.
Model-agnostic meta-learning for fast adaptation of deep networks.
In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, pp. 1126–1135, Sydney, NSW, Australia, August 2017. JMLR.org.

Fischer, T., Puigbò, J.-Y., Camilleri, D., Nguyen, P. D. H., Moulin-Frier, C., Lallée, S., Metta, G., Prescott, T. J., Demiris, Y., and Verschure, P. F. M. J.
iCub-HRI: A Software Framework for Complex Human–Robot Interaction Scenarios on the iCub Humanoid Robot.
Frontiers in Robotics and AI, 5:22, 2018.
ISSN 2296-9144.
doi: 10.3389/frobt.2018.00022.

Flack, J. C., Girvan, M., de Waal, F. B. M., and Krakauer, D. C.
Policing stabilizes construction of social niches in primates.
Nature, 439(7075):426–429, January 2006.
ISSN 1476-4687.
doi: 10.1038/nature04326.

Fogarty, L. and Creanza, N.
The niche construction of cultural complexity: Interactions between innovations, population size and the environment.
Philosophical Transactions of the Royal Society B: Biological Sciences, 372(1735): 20160428, 2017.
ISSN 0962-8436, 1471-2970.
doi: 10.1098/rstb.2016.0428.

Forestier, S. and Oudeyer, P.-Y.
Modular active curiosity-driven discovery of tool use.
In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference On*, pp. 3965–3972. IEEE, 2016.

Forestier, S., Mollard, Y., and Oudeyer, P.-Y.
Intrinsically motivated goal exploration processes with automatic curriculum learning.
CoRR, abs/1708.02190, 2017.

Frankenhuis, W. E., Panchanathan, K., and Barto, A. G.
Enriching behavioral ecology with reinforcement learning methods.
Behavioural Processes, 161:94–100, 2019.
ISSN 0376-6357.
doi: 10.1016/j.beproc.2018.01.008.

Freeberg, T. M., Dunbar, R. I. M., and Ord, T. J.
Social complexity as a proximate and ultimate factor in communicative complexity.
Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1597): 1785–1801, July 2012.
ISSN 0962-8436.
doi: 10.1098/rstb.2011.0213.

Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D., and Verschure, P. F. M. J.
Modeling the formation of social conventions from embodied real-time interactions.
PLOS ONE, 15(6):e0234434, 2020.
ISSN 1932-6203.
doi: 10.1371/journal.pone.0234434.

Froese, T. and Ziemke, T.
Enactive artificial intelligence: Investigating the systemic organization of life and mind.
Artificial Intelligence, 173(3):466–500, March 2009.
ISSN 0004-3702.
doi: 10.1016/j.artint.2008.12.001.

Garcia Ortiz, M., Jankovics, V., Caselles-Dupre, H., and Annabi, L.
Simple-playgrounds, 2021.

Gärdenfors, P.
Cooperation and the Evolution of Symbolic Communication.
Lund University, 2002.

Gentner, D. and Hoyos, C.
Analogy and Abstraction.
Topics in Cognitive Science, (3), 2017.
ISSN 17568757.

Ghazanfar, A. A. and Takahashi, D. Y.
The evolution of speech: Vision, rhythm, cooperation.
Trends in Cognitive Sciences, 18(10):543–553, 2014.
ISSN 1879307X.
doi: 10.1016/j.tics.2014.06.004.

Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., and Taddei, F.
Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut.
Science, 291(5513):2606–2608, March 2001.
doi: 10.1126/science.1056421.

Goertzel, B. and Pennachin, C.
Artificial General Intelligence, volume 2.
Springer, 2007.

Gopnik, A.
Childhood as a solution to explore–exploit tensions.
Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1803): 20190502, July 2020.
doi: 10.1098/rstb.2019.0502.

Gopnik, A., Meltzoff, A. N., and Kuhl, P. K.
The Scientist in the Crib: Minds, Brains, and How Children Learn.
William Morrow & Co, 1999.

Gottlieb, J. and Oudeyer, P.-Y.
Towards a neuroscience of active sampling and curiosity.
Nature Reviews Neuroscience, 19(12):758–770, December 2018.
ISSN 1471-0048.
doi: 10.1038/s41583-018-0078-0.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., and Baranes, A.
Information-seeking, curiosity, and attention: Computational and neural mechanisms.
Trends in Cognitive Sciences, 17(11):585–593, 2013.
ISSN 13646613.
doi: 10.1016/j.tics.2013.09.001.

Gowlett, J. A. J.
The discovery of fire by humans: A long and convoluted process.
Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1696): 20150164, 2016.

Graeber, D. and Wengrow, D.
The Dawn of Everything: A New History of Humanity.
Penguin UK, 2021.

Grove, M.
Evolution and dispersal under climatic instability: A simple evolutionary algorithm.
Adaptive Behavior, 22(4):235–254, August 2014.
ISSN 1059-7123.
doi: 10.1177/1059712314533573.

Gruber, T., Reynolds, V., and Zuberbühler, K.
The knowns and unknowns of chimpanzee culture.
Communicative & Integrative Biology, 3(3):221–223, 2010.
ISSN 1942-0889.

Gupta, A., Kumar, V., Lynch, C., Levine, S., and Hausman, K.
Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning.
In Kaelbling, L. P., Kragic, D., and Sugiura, K. (eds.), *3rd Annual Conference on Robot Learning, CoRL 2019, Osaka, Japan, October 30 - November 1, 2019, Proceedings*, volume 100 of *Proceedings of Machine Learning Research*, pp. 1025–1037. PMLR, 2019.

Gupta, A., Savarese, S., Ganguli, S., and Fei-Fei, L.
Embodied intelligence via learning and evolution.
Nature Communications, 12(1):5721, December 2021.
ISSN 2041-1723.
doi: 10.1038/s41467-021-25874-z.

Ha, D. and Tang, Y.
Collective Intelligence for Deep Learning: A Survey of Recent Developments.
arXiv:2111.14377 [cs], December 2021.

Hafner, D.
Benchmarking the spectrum of agent capabilities.
arXiv preprint arXiv:2109.06780, 2021.

Hamon, G., Etcheverry, M., Chan, B. W.-C., Moulin-Frier, C., and Oudeyer, P.-Y.
Learning Sensorimotor Agency in Cellular Automata (blog post: https://developmentalsystems.org/sensorimotor-lenia/), 2022.

Harari, Y. N.
Sapiens.
2014.
ISBN 978-1-84655-824-5.

Held, D., Geng, X., Florensa, C., and Abbeel, P.
Automatic Goal Generation for Reinforcement Learning Agents.
May 2017.

Hesse, M.
The cognitive claims of metaphor.
The journal of speculative philosophy, 1988.

Hintze, A.
Open-endedness for the sake of open-endedness.
Artificial Life, 25:198–206, 2019.

Horgan, D., Quan, J., Budden, D., Barth-Maron, G., Hessel, M., van Hasselt, H., and Silver, D.
Distributed Prioritized Experience Replay.
arXiv:1803.00933 [cs], March 2018.

Hougen, D. F. and Shah, S. N. H.
The Evolution of Reinforcement Learning.
In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1457–1464, December 2019.
doi: 10.1109/SSCI44817.2019.9003146.

Hughes, E., Leibo, J. Z., Phillips, M. G., Tuyls, K., Duéñez-Guzmán, E. A., Castañeda, A. G., Dunning, I., Zhu, T., McKee, K. R., Koster, R., Roff, H., and Graepel, T.
Inequity aversion improves cooperation in intertemporal social dilemmas.
arXiv:1803.08884 [cs, q-bio], September 2018.

Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Risi, S., and Stanley, K. O.
Indirectly Encoding Neural Plasticity as a Pattern of Local Rules.
In Doncieux, S., Girard, B., Guillot, A., Hallam, J., Meyer, J.-A., and Mouret, J.-B. (eds.), *From Animals to Animats 11*, volume 6226, pp. 533–543. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
ISBN 978-3-642-15192-7 978-3-642-15193-4.
doi: 10.1007/978-3-642-15193-4_50.

Iriki, A. and Taoka, M.
Triadic (ecological, neural, cognitive) niche construction: A scenario of human brain evolution extrapolating tool use and language from the control of reaching actions.
Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1585):10–23, 2012.
ISSN 0962-8436.
doi: 10.1098/rstb.2011.0190.

Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P. A., Strouse, D., Leibo, J. Z., and de Freitas, N.
Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning.
In *Proceedings of the 35 Th International Conference on Machine Learning, Stockholm, Sweden*, 2019.

Jiang, M., Luketina, J., Nardelli, N., Minervini, P., Torr, P. H. S., Whiteson, S., and Rocktäschel, T.
WordCraft: An environment for benchmarking commonsense agents.
arXiv:2007.09185 [cs], 2020.

Johnson, M., Hofmann, K., Hutton, T., and Bignell, D.
The Malmo Platform for Artificial Intelligence Experimentation.
International joint conference on artificial intelligence (IJCAI), pp. 4246–4247, 2016.
ISSN 10450823.

Johnston, T. D.
Selective Costs and Benefits in the Evolution of Learning.
In Rosenblatt, J. S., Hinde, R. A., Beer, C., and Busnel, M.-C. (eds.), *Advances in the Study of Behavior*, volume 12, pp. 65–106. Academic Press, January 1982.
doi: 10.1016/S0065-3454(08)60046-7.

Kaplan, F. and Oudeyer, P.-Y.
In search of the neural circuits of intrinsic motivation.
Frontiers in neuroscience, 1(1):225, 2007.

Karch, T., Colas, C., Teodorescu, L., Moulin-Frier, C., and Oudeyer, P.-Y.
Deep Sets for Generalization in RL.
In *Beyond "Tabula Rasa" in Reinforcement Learning (BeTR-RL) Workshop*, 2020.

Karch, T., Teodorescu, L., Hofmann, K., Moulin-Frier, C., and Oudeyer, P.-Y.
Grounding spatio-temporal language with transformers.
In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 5236–5249. Curran Associates, Inc., 2021.

Kendal, J., Tehrani, J. J., and Odling-Smee, J.
Human niche construction in interdisciplinary focus.
Philosophical Transactions of the Royal Society B: Biological Sciences, 366(1566): 785–792, March 2011.
doi: 10.1098/rstb.2010.0306.

Kidd, C. and Hayden, B. Y.
The psychology and neuroscience of curiosity.
Neuron, 88(3):449–460, 2015.

Kirby, S., Griffiths, T., and Smith, K.
Iterated learning and the evolution of language.
Current Opinion in Neurobiology, 28:108–114, October 2014.
ISSN 0959-4388.
doi: 10.1016/j.conb.2014.07.014.

Kirsch, L. and Schmidhuber, J.
Meta learning backpropagation and improving it.
arXiv:2012.14905 [cs, stat], October 2021.

Köster, R., Hadfield-Menell, D., Everett, R., Weidinger, L., Hadfield, G. K., and Leibo, J. Z.
Spurious normativity enhances learning of compliance and enforcement behavior in artificial agents.
Proceedings of the National Academy of Sciences, 119(3):e2106028118, January 2022.
doi: 10.1073/pnas.2106028118.

Kovac, G., Portelas, R., Hofmann, K., and Oudeyer, P.-Y.
SocialAI: Benchmarking socio-cognitive abilities in deep reinforcement learning agents.
CoRR, abs/2107.00956, 2021.

Krakauer, D. C., Page, K. M., Erwin, D. H., Rice, A. E. S. H., and Whitlock, E. M. C.
Diversity, Dilemmas, and Monopolies of Niche Construction.
The American Naturalist, 173(1):26–40, 2009.
ISSN 0003-0147.
doi: 10.1086/593707.

Krams, I., Krama, T., Freeberg, T. M., Kullberg, C., and Lucas, J. R.
Linking social complexity and vocal complexity: A parid perspective.
Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1597): 1879–1891, July 2012.
ISSN 0962-8436.
doi: 10.1098/rstb.2011.0222.

Kuhl, P. K.
Early language acquisition: Cracking the speech code.
Nature Reviews Neuroscience, 5(11):831–843, 2004.
ISSN 1471-003X.
doi: 10.1038/nrn1533.

Lakoff, G. and Johnson, M.
Metaphors We Live By.
University of Chicago press, 2008.

Laland, K. N. and O'Brien, M. J.
Cultural Niche Construction: An Introduction.

Biological Theory, 6(3):191–202, September 2011.
ISSN 1555-5550.
doi: 10.1007/s13752-012-0026-6.

Laland, K. N., Odling-Smee, J., and Feldman, M. W.
Cultural niche construction and human evolution.
Journal of Evolutionary Biology, 14(1):22–33, 2001.
doi: 10.1046/j.1420-9101.2001.00262.x.

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., and Odling-Smee, J.
The extended evolutionary synthesis: Its structure, assumptions and predictions.
Proceedings of the Royal Society B: Biological Sciences, 282(1813):20151019, August 2015.
doi: 10.1098/rspb.2015.1019.

Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Perolat, J., Silver, D., and Graepel, T.
A Unified Game-Theoretic Approach to Multiagent Reinforcement Learning.
In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 4190–4203, 2017.

Lange, R. T. and Sprekeler, H.
Learning not to learn: Nature versus nurture in silico.
arXiv:2010.04466 [cs, q-bio], 2021.

Larrasoaña, J. C.
A Northeast Saharan Perspective on Environmental Variability in North Africa and its Implications for Modern Human Origins.
In Hublin, J.-J. and McPherron, S. P. (eds.), *Modern Origins: A North African Perspective*, Vertebrate Paleobiology and Paleoanthropology, pp. 19–34. Springer Netherlands, Dordrecht, 2012.
ISBN 978-94-007-2929-2.
doi: 10.1007/978-94-007-2929-2_2.

Lazaridou, A. and Baroni, M.
Emergent Multi-Agent Communication in the Deep Learning Era.
arXiv:2006.02419 [cs], June 2020.

LeCun, Y.
A Path Towards Autonomous Machine Intelligence, 2022.

Legg, S. and Hutter, M.
Universal Intelligence: A Definition of Machine Intelligence.
Minds and Machines, 17(4):391–444, December 2007.
ISSN 1572-8641.
doi: 10.1007/s11023-007-9079-x.

Legg, S., Hutter, M., and Others.
A collection of definitions of intelligence.
Frontiers in Artificial Intelligence and applications, 157:17, 2007.

Lehman, J. and Stanley, K.
Evolvability is inevitable: Increasing evolvability without the pressure to adapt.
PloS one, 8:e62186, May 2013.
doi: 10.1371/journal.pone.0062186.

Lehman, J. and Stanley, K. O.
Evolving a diversity of virtual creatures through novelty search and local competition.
In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation*, GECCO '11, pp. 211–218, New York, NY, USA, July 2011. Association for Computing Machinery.
ISBN 978-1-4503-0557-0.
doi: 10.1145/2001576.2001606.

Leibo, J. Z., Zambaldi, V., Lanctot, M., Marecki, J., and Graepel, T.
Multi-agent Reinforcement Learning in Sequential Social Dilemmas.
In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 464–473. International Foundation for Autonomous Agents and Multiagent Systems, February 2017.

Leibo, J. Z., Hughes, E., Lanctot, M., and Graepel, T.
Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research.
arXiv preprint arXiv:1903.00742, 2019.

Lemesle, Y., Karch, T., Laroche, R., Moulin-Frier, C., and Oudeyer, P.-Y.
Emergence of Shared Sensory-motor Graphical Language from Visual Input, October 2022.

Li, F. and Bowling, M.
Ease-of-Teaching and Language Structure from Emergent Communication.
Technical report, 2019.

Lupyan, G.
What Do Words Do? Toward a Theory of Language-Augmented Thought.
In *Psychology of Learning and Motivation*. Elsevier, 2012.

Lynch, M., Ackerman, M. S., Gout, J.-F., Long, H., Sung, W., Thomas, W. K., and Foster, P. L.
Genetic drift, selection and the evolution of the mutation rate.
Nature Reviews Genetics, 17(11):704–714, November 2016.
ISSN 1471-0064.
doi: 10.1038/nrg.2016.104.

Marcus, G. and Davis, E.
Rebooting AI: Building Artificial Intelligence We Can Trust.
Pantheon Books, New York, first edition edition, 2019.
ISBN 978-1-5247-4825-8 978-0-525-56604-5.

Marean, C. W., Anderson, R. J., Bar-Matthews, M., Braun, K., Cawthra, H. C., Cowling, R. M., Engelbrecht, F., Esler, K. J., Fisher, E., Franklin, J., Hill, K., Janssen, M., Potts, A. J., and Zahn, R.

A new research strategy for integrating studies of paleoclimate, paleoenvironment, and paleoanthropology.
Evolutionary Anthropology: Issues, News, and Reviews, 24(2):62–72, 2015.
ISSN 10601538.
doi: 10.1002/evan.21443.

Maslin, M. A., Shultz, S., and Trauth, M. H.
A synthesis of the theories and concepts of early human evolution.
Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1663): 20140064, 2015.
doi: 10.1098/rstb.2014.0064.

Mason, W. A., Jones, A., and Goldstone, R. L.
Propagation of innovations in networked groups.
Journal of Experimental Psychology: General, 137(3):422–433, 2008.
ISSN 1939-2222, 0096-3445.
doi: 10.1037/a0012798.

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

Mazyavkina, N., Sviridov, S., Ivanov, S., and Burnaev, E.
Reinforcement learning for combinatorial optimization: A survey.
Computers & Operations Research, 134:105400, 2021.

McCarthy, J.
From here to human-level AI.
Artificial Intelligence, 171(18):1174–1182, December 2007.
ISSN 0004-3702.
doi: 10.1016/j.artint.2007.10.009.

Mitchell, T. M.
Machine Learning.
McGraw-Hill Series in Computer Science. McGraw-Hill, New York, 1997.
ISBN 978-0-07-042807-2.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., Others, Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D.
Human-level control through deep reinforcement learning.
Nature, 518(7540):529–533, February 2015.
ISSN 0028-0836.
doi: 10.1038/nature14236.

Moore, R. K.
Spoken language processing: Piecing together the puzzle.
Speech Communication, 49(5):418–435, May 2007.

ISSN 0167-6393.
doi: 10.1016/j.specom.2007.01.011.

Mordatch, I. and Abbeel, P.
Emergence of Grounded Compositional Language in Multi-Agent Populations.
In *Thirty-Second AAAI Conference on Artificial Intelligence*, March 2017.

Moulin-Frier, C. and Arbib, M. A.
Recognizing speech in a novel accent: The motor theory of speech perception reframed.
Biological Cybernetics, 107(4):421–447, 2013.

Moulin-Frier, C. and Oudeyer, P.-Y.
Curiosity-driven phonetic learning.
In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob 2012)*, pp. 1–8, 2012.
ISBN 978-1-4673-4963-5.
doi: 10.1109/DevLrn.2012.6400583.

Moulin-Frier, C. and Oudeyer, P.-Y.
Exploration strategies in developmental robotics: A unified probabilistic framework.
In *International Conference on Development and Learning, ICDL/Epirob, Osaka, Japan*, pp. 1–6, 2013a.
ISBN 978-1-4799-1036-6.
doi: 10.1109/DevLrn.2013.6652535.

Moulin-Frier, C. and Oudeyer, P.-Y.
The role of intrinsic motivations in learning sensorimotor vocal mappings: A developmental robotics study.
In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 1268–1272, Lyon, France, 2013b.

Moulin-Frier, C. and Oudeyer, P.-Y.
Multi-Agent Reinforcement Learning as a Computational Tool for Language Evolution Research: Historical Context and Future Challenges.
In *Challenges and Opportunities for Multi-Agent Reinforcement Learning (COMARL), AAAI Spring Symposium Series, Stanford University, Palo Alto, California, USA*, 2021.

Moulin-Frier, C. and Verschure, P. F.
Two possible driving forces supporting the evolution of animal communication.
Physics of Life Reviews, 16:88–90, March 2016.
ISSN 15710645.
doi: 10.1016/j.plrev.2016.01.019.

Moulin-Frier, C., Schwartz, J.-L. J.-L., Diard, J., and Bessière, P.
Emergence of a language through deictic games within a society of sensory-motor agents in interaction.
In *8th International Seminar on Speech Production, ISSP 2008*, pp. 261–264, Strasbourg France, 2008.

Moulin-Frier, C., Schwartz, J.-L., Diard, J., and Bessière, P.
A unified theoretical Bayesian model of speech communication.
In *1st Conference on Applied Digital Human Modeling, Miami, USA*, 2010.

Moulin-Frier, C., Schwartz, J.-L., Diard, J., and Bessière, P.
Emergence of articulatory-acoustic systems from deictic interaction games in a "Vocalize to Localize" framework.
In Vilain, A., Schwartz, J.-L., Abry, C., and Vauclair, J. (eds.), *Primate Communication and Human Language: Vocalisations, Gestures, Imitation and Deixis in Humans and Non-Humans*, pp. 193–220. Advances in Interaction Studies' series by John Benjamins Pub. Co., 2011.

Moulin-Frier, C., Laurent, R., Bessière, P., Schwartz, J.-L., and Diard, J.
Adverse conditions improve distinguishability of auditory, motor and perceptuo-motor theories of speech perception: An exploratory Bayesian modeling study.
Language and Cognitive Processes, 27(7–8):1240–1263, 2012.

Moulin-Frier, C., Nguyen, S. M., and Oudeyer, P.-Y.
Self-Organization of Early Vocal Development in Infants and Machines: The Role of Intrinsic Motivation.
Frontiers in Psychology (Cognitive Science), 4(1006), 2014a.
ISSN 1664-1078.
doi: 10.3389/fpsyg.2013.01006.

Moulin-Frier, C., Rouanet, P., and Oudeyer, P.-Y.
Explauto: An open-source Python library to study autonomous exploration in developmental robotics.
In *IEEE ICDL-EPIROB 2014 - 4th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics*, pp. 171–172, 2014b.
ISBN 978-1-4799-7540-2.
doi: 10.1109/DEVLRN.2014.6982976.

Moulin-Frier, C., Diard, J., Schwartz, J.-L., and Bessière, P.
COSMO ('Communicating about Objects using Sensory-Motor Operations'): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems.
Journal of Phonetics, 53:5–41, 2015a.
ISSN 00954470.
doi: 10.1016/j.wocn.2015.06.001.

Moulin-Frier, C., Sanchez-Fibla, M., and Verschure, P.
Autonomous development of turn-taking behaviors in agent populations: A computational study.
In *International Conference on Development and Learning and Epigenetic Robotics, ICDL-EpiRob 2015*, 2015b.
ISBN 978-1-4673-9320-1.
doi: 10.1109/DEVLRN.2015.7346139.

Moulin-Frier, C., Arsiwalla, X. D., Puigbò, J.-Y. J.-Y., Sánchez-Fibla, M., Duff, A., Verschure, P. F. M. J. P., Sanchez-Fibla, M., Duff, A., and Verschure, P. F. M. J. P.

Top-down and bottom-up interactions between low-level reactive control and symbolic rule learning in embodied agents.
In *Proceedings of the Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches. 30th Annual Conference on Neural Information Processing Systems (NIPS 2016)*, volume 1773, 2016.

Moulin-Frier, C., Brochard, J., Stulp, F., and Oudeyer, P.
Emergent Jaw Predominance in Vocal Development through Stochastic Optimization.
IEEE Transactions on Cognitive and Developmental Systems, 2017a.
ISSN 23798939.
doi: 10.1109/TCDS.2017.2704912.

Moulin-Frier, C., Puigbò, J.-Y., Arsiwalla, X. D., Sanchez-Fibla, M., and Verschure, P. F. M. J.
Embodied Artificial Intelligence through Distributed Adaptive Control: An Integrated Framework.
In *International Conference on Development and Learning, ICDL/Epirob, Lisbon, Portugal*, 2017b.

Moulin-Frier, C., Fischer, T., Petit, M., Pointeau, G., Puigbo, J.-Y., Pattacini, U., Low, S. C., Camilleri, D., Nguyen, P., Hoffmann, M., Chang, H. J., Zambelli, M., Mealier, A.-L., Damianou, A., Metta, G., Prescott, T. J., Demiris, Y., Dominey, P. F., and Verschure, P. F. M. J.
DAC-h3: A Proactive Robot Cognitive Architecture to Acquire and Express Knowledge About the World and the Self.
IEEE Transactions on Cognitive and Developmental Systems, 10(4):1005–1022, 2018.
ISSN 2379-8920.
doi: 10.1109/TCDS.2017.2754143.

Muthukrishna, M. and Henrich, J.
Innovation in the collective brain.
Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1690): 20150192, 2016.
ISSN 0962-8436, 1471-2970.
doi: 10.1098/rstb.2015.0192.

Nair, A., Pong, V., Dalal, M., Bahl, S., Lin, S., and Levine, S.
Visual Reinforcement Learning with Imagined Goals.
arXiv:1807.04742 [cs, stat], December 2018.

Najarro, E. and Risi, S.
Meta-Learning through Hebbian Plasticity in Random Networks.
arXiv:2007.02686 [cs], March 2021.

Nettle, D., Gibson, M. A., Lawson, D. W., and Sear, R.
Human behavioral ecology: Current research and future prospects.
Behavioral Ecology, 24(5):1031–1040, 2013.
ISSN 1465-7279, 1045-2249.
doi: 10.1093/beheco/ars222.

Newell, A.
Unified Theories of Cognition.
Harvard University Press, 1994.

N'Guyen, S., Moulin-Frier, C., and Droulez, J.
Decision making under uncertainty: A quasimetric approach.
PLoS ONE, 8(12), 2013.
ISSN 19326203.

Nisioti, E. and Moulin-Frier, C.
Grounding Artificial Intelligence in the Origins of Human Behavior.
arXiv:2012.08564 [cs], December 2020.

Nisioti, E. and Moulin-Frier, C.
Plasticity and evolvability under environmental variability: The joint role of fitness-based selection and niche-limited competition.
In *Proceedings of the 2022 Genetic and Evolutionary Computation Conference (GECCO 2022)*, 2022.

Nisioti, E., Jodogne-del Litto, K., and Moulin-Frier, C.
Grounding an Ecological Theory of Artificial Intelligence in Human Evolution.
In *NeurIPS 2021 - Conference on Neural Information Processing Systems / Workshop: Ecological Theory of Reinforcement Learning*, virtual event, France, 2021.

Nisioti, E., Mahaut, M., Oudeyer, P.-Y., Momennejad, I., and Moulin-Frier, C.
Social Network Structure Shapes Innovation: Experience-sharing in RL with SAPIENS, 2022.

Niv, Y., Joel, D., Meilijson, I., and Ruppin, E.
Evolution of reinforcement learning in uncertain environments: A simple explanation for complex foraging behaviors.
pp. 20.

Odling-Smee, F. J., Laland, K. N., and Feldman, M. W.
Niche Construction: The Neglected Process in Evolution.
Monographs in Population Biology. Princeton University Press, Princeton, 2003.
ISBN 978-0-691-04438-5.

Oh, J., Hessel, M., Czarnecki, W. M., Xu, Z., van Hasselt, H., Singh, S., and Silver, D.
Discovering reinforcement learning algorithms.
arXiv:2007.08794 [cs], 2021.

Oller, D. K.
The Emergence of the Speech Capacity.
Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
ISBN 978-0-8058-2629-6.

OpenAI.
OpenAI Five, 2018.

Oudeyer, P.-Y.
The self-organization of speech sounds.
Journal of Theoretical Biology, 233(3):435–449, 2005.
ISSN 0022-5193.
doi: 10.1016/j.jtbi.2004.10.025.

Oudeyer, P.-Y.
Self-Organization in the Evolution of Speech, volume 6.
Oxford University Press, January 2006.
doi: 10.1093/acprof:oso/9780199289158.001.0001.

Oudeyer, P.-Y.
Computational theories of curiosity-driven learning.
In *The New Science of Curiosity*, Psychology of Emotions, Motivations and Actions, pp. 43–72. Nova Science Publishers, Hauppauge, NY, US, 2018.
ISBN 978-1-5361-3800-9.

Oudeyer, P.-Y. and Smith, L. B.
How Evolution May Work Through Curiosity-Driven Developmental Process.
Topics in Cognitive Science, 8(2):492–502, April 2016.
ISSN 1756-8765.
doi: 10.1111/tops.12196.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V.
Intrinsic Motivation Systems for Autonomous Mental Development.
IEEE Transactions on Evolutionary Computation, 11(2):265–286, 2007.

Pal, A., Philion, J., Liao, Y.-H., and Fidler, S.
Emergent road rules in multi-agent driving environments.
arXiv:2011.10753 [cs], November 2020.

Papavasileiou, E., Cornelis, J., and Jansen, B.
A systematic literature review of the successors of "NeuroEvolution of augmenting topologies".
Evolutionary Computation, 29(1):1–73, 2021.
ISSN 1063-6560, 1530-9304.
doi: 10.1162/evco_a_00282.

Pardo, F., Tavakoli, A., Levdik, V., and Kormushev, P.
Time Limits in Reinforcement Learning.
In *Proceedings of the 35th International Conference on Machine Learning*, pp. 4045–4054. PMLR, July 2018.

Pathak, D., Agrawal, P., Efros, A. A., and Darrell, T.
Curiosity-driven exploration by self-supervised prediction.
In *International Conference on Machine Learning (ICML)*, volume 2017, 2017.

Pearson, P. N.
Red queen hypothesis.
In *eLS*. American Cancer Society, 2001.

ISBN 978-0-470-01590-2.
doi: 10.1038/npg.els.0001667.

Pérolat, J., Leibo, J. Z., Zambaldi, V., Beattie, C., Tuyls, K., and Graepel, T.
A multi-agent reinforcement learning model of common-pool resource appropriation.
In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pp. 3643–3652, 2017.

Piaget, J.
The Language and Thought of the Child.
Routledge, 1926.
ISBN 0-415-26750-1.

Pigliucci, M. and Muller, G. B.
Evolution–the extended synthesis.
2010.

Platanios, E. A., Saparov, A., and Mitchell, T.
Jelly bean world: A testbed for never-ending learning, 2020.

Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., and Levine, S.
Skew-Fit: State-Covering Self-Supervised Reinforcement Learning.
arXiv:1903.03698 [cs, stat], 2020.

Portelas, R., Colas, C., Weng, L., Hofmann, K., and Oudeyer, P.-Y.
Automatic curriculum learning for deep RL: A short survey.
arXiv:2003.04664 [cs, stat], 2020.

Post, D. M. and Palkovacs, E. P.
Eco-evolutionary feedbacks in community and ecosystem ecology: Interactions between the ecological theatre and the evolutionary play.
Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1523): 1629–1640, 2009.
ISSN 0962-8436, 1471-2970.
doi: 10.1098/rstb.2009.0012.

Potts, R.
Hominin evolution in settings of strong environmental variability.
Quaternary Science Reviews, 73:1–13, August 2013.
ISSN 0277-3791.
doi: 10.1016/j.quascirev.2013.04.003.

Pugh, J. K., Soros, L. B., and Stanley, K. O.
Quality diversity: A new frontier for evolutionary computation.
Frontiers Robotics AI, 3:40, 2016.

Puigbò, J.-Y., Herreros, I., Moulin-Frier, C., and Verschure, P. .
Towards a Two-Phase Model of Sensor and Motor Learning.
In *4th International Conference on Biomimetic and Biohybrid Systems*, 2015a.

Puigbò, J.-Y., Moulin-Frier, C., Vouloutsi, V., Sanchez-Fibla, M., Herreros, I., and Verschure, P.
Skill refinement through cerebellar learning and human haptic feedback: An iCub learning to paint experiment.
In *IEEE-RAS Conference on Humanoids Robots (Humanoids 2015), Seoul, Korea*, 2015b.
ISBN 978-1-4799-6885-5.
doi: 10.1109/HUMANOIDS.2015.7363580.

Puigbò, J.-Y., Vouloutsi, V., Moulin-Frier, C., and Verschure, P. F. M. J.
Reactive and adaptive control loops for social learning in human-robot interaction, 2015c.

Puigbò, J.-Y., Moulin-Frier, C., and Verschure, P.
Towards self-controlled robots through distributed adaptive control.
In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9793, 2016.
ISBN 978-3-319-42416-3.
doi: 10.1007/978-3-319-42417-0_52.

Racaniere, S., Lampinen, A. K., Santoro, A., Reichert, D. P., Firoiu, V., and Lillicrap, T. P.
Automated curricula through setter-solver interactions.
arXiv preprint arXiv:1909.12892, 2019.

Reinke, C., Etcheverry, M., and Oudeyer, P.-Y.
Intrinsically motivated discovery of diverse patterns in self-organizing systems.
arXiv:1908.06663 [cs, stat], 2020-02-17, 2020-2-17.

Ren, Y., Guo, S., Labeau, M., Cohen, S. B., and Kirby, S.
Compositional languages emerge in a neural iterated learning model.
In *International Conference on Learning Representations*, March 2020.

Richerson, P. J. and Boyd, R.
Not by Genes Alone: How Culture Transformed Human Evolution.
University of Chicago press, 2008.

Risi, S. and Togelius, J.
Increasing Generality in Machine Learning through Procedural Content Generation.
arXiv:1911.13071 [cs], March 2020.

Roy, A. C. and Arbib, M. A.
The syntactic motor system.
Gesture, 5(1):7–37, 2005.
ISSN 15681475.
doi: 10.1075/gest.5.1.03roy.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G.
Sequential Thought Processes in Pdp Models.
Parallel distributed processing: explorations in the microstructures of cognition, 1986.

Saltz, J. B., Geiger, A. P., Anderson, R., Johnson, B., and Marren, R.
What, if anything, is a social niche?
Evolutionary Ecology, 30(2):349–364, 2016.
ISSN 1573-8477.
doi: 10.1007/s10682-015-9792-5.

Sánchez-Fibla, M., Moulin-Frier, C., and Verschure, P. F. M. J.
A sensorimotor account of visual and tactile integration for depth perception: An iCub robot experiment.
In *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp. 86–91, Lisbon, Portugal, 2017.
doi: 10.1109/DEVLRN.2017.8329792.

Schmidhuber, J.
A possibility for implementing curiosity and boredom in model-building neural controllers.
In Meyer, J. A. and Wilson, S. W. (eds.), *Proc. SAB'91*, pp. 222–227, 1991.

Schmidhuber, J.
Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-2010).
IEEE Transactions on Autonomous Mental Development, 2(3):230–247, 2010.

Schmidhuber, J., Zhao, J., and Wiering, M.
Shifting Inductive Bias with Success-Story Algorithm, Adaptive Levin Search, and Incremental Self-Improvement.
Machine Learning, 28:105–130, 1997.
ISSN 08856125.
doi: 10.1023/A:1007383707642.

Schwartz, J.-L., Moulin-Frier, C., and Oudeyer, P.-Y.
On the cognitive nature of speech sound systems.
Journal of Phonetics, 53:1–4, 2015.
ISSN 00954470.
doi: 10.1016/j.wocn.2015.09.008.

Sear, R., Lawson, D. W., and Dickins, T. E.
Synthesis in the human evolutionary behavioural sciences.
Journal of Evolutionary Psychology, 5(1):3–28, March 2007.
ISSN 1789-2082, 2060-5587.
doi: 10.1556/JEP.2007.1019.

Searle, J. R.
Minds, brains, and programs.
Behavioral and Brain Sciences, 3(3):417–424, September 1980.
ISSN 1469-1825, 0140-525X.
doi: 10.1017/S0140525X00005756.

Shultz, S. and Maslin, M.
Early Human Speciation, Brain Expansion and Dispersal Influenced by African Climate Pulses.

PLOS ONE, 8(10):e76750, 2013.
ISSN 1932-6203.
doi: 10.1371/journal.pone.0076750.

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., and Hassabis, D.
Mastering the game of Go with deep neural networks and tree search.
Nature, 529(7587):484–489, 2016.
ISSN 0028-0836.
doi: 10.1038/nature16961.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., and Hassabis, D.
Mastering the game of Go without human knowledge.
Nature, 550(7676):354–359, October 2017.
ISSN 0028-0836.
doi: 10.1038/nature24270.

Silver, D., Singh, S., Precup, D., and Sutton, R. S.
Reward is enough.
Artificial Intelligence, 299:103535, October 2021.
ISSN 0004-3702.
doi: 10.1016/j.artint.2021.103535.

Singh, S., Barto, A. G., and Chentanez, N.
Intrinsically motivated reinforcement learning.
18th Annual Conference on Neural Information Processing Systems (NIPS), 17:1281–1288, 2004.
ISSN 1943-0604.
doi: 10.1109/TAMD.2010.2051031.

Singh, S., Lewis, R. L., Barto, A. G., and Sorg, J.
Intrinsically motivated reinforcement learning: An evolutionary perspective.
IEEE Transactions on Autonomous Mental Development, 2(2):70–82, 2010.

Smith, E. A.
Communication and collective action: Language and the evolution of human cooperation.
Evolution and Human Behavior, 31(4):231–245, July 2010.
ISSN 1090-5138.
doi: 10.1016/J.EVOLHUMBEHAV.2010.03.001.

Solé, R.
The major synthetic evolutionary transitions.
Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1701): 20160175, August 2016.

ISSN 0962-8436, 1471-2970.
doi: 10.1098/rstb.2016.0175.

Solé, R.
Revisiting Leigh Van Valen's "A New Evolutionary Law" (1973).
Biological Theory, pp. s13752–021–00391–w, January 2022.
ISSN 1555-5542, 1555-5550.
doi: 10.1007/s13752-021-00391-w.

Solé, R. V., Valverde, S., Casals, M. R., Kauffman, S. A., Farmer, D., and Eldredge, N.
The evolutionary ecology of technological innovations.
Complexity, 18(4):15–27, 2013.
ISSN 1099-0526.
doi: 10.1002/cplx.21436.

Stanley, K. O.
Why Open-Endedness Matters.
Artificial Life, 25(3):232–235, August 2019.
ISSN 1064-5462.
doi: 10.1162/artl_a_00294.

Stanley, K. O. and Lehman, J.
Why Greatness Cannot Be Planned: The Myth of the Objective.
Springer Publishing Company, Incorporated, 2015.
ISBN 978-3-319-15523-4.

Steels, L.
The Artificial Life Roots of Artificial Intelligence.
Artificial Life Journal, 1(1):89–125, 1994.

Steels, L.
The synthetic modeling of language origins.
Evolution of Communication, 1(1):1–34, 1997.

Steels, L.
The Autotelic Principle.
In Iida, F., Pfeifer, R., Steels, L., and Kuniyoshi, Y. (eds.), *Embodied Artificial Intelligence: Dagstuhl Castle, Germany, July 7-11, 2003*, volume 3139, pp. 231–242. Springer Verlag, Berlin, 2004.

Steels, L.
Modeling the cultural evolution of language.
Physics of Life Reviews, 8(4):339–356, 2011.
ISSN 15710645.
doi: 10.1016/j.plrev.2011.10.014.

Stephens, D. W.
Change, regularity, and value in the evolution of animal learning.
Behavioral Ecology, 2(1):77–89, 1991.
ISSN 1465-7279, 1045-2249.
doi: 10.1093/beheco/2.1.77.

Sterling, P.
Allostasis: A model of predictive regulation.
Physiology and Behavior, 106(1):5–15, 2012.
ISSN 00319384.
doi: 10.1016/j.physbeh.2011.06.004.

Suarez, J., Du, Y., Isola, P., and Mordatch, I.
Neural MMO: A Massively Multiagent Game Environment for Training and Evaluating Intelligent Agents.
March 2019.

Taylor, J., Nisioti, E., and Moulin-Frier, C.
Socially Supervised Representation Learning: The Role of Subjectivity in Learning Efficient Representations.
In *International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2022)*, 2022.

Team, C. G. I., Bhoopchand, A., Brownfield, B., Collister, A., Lago, A. D., Edwards, A., Everett, R., Frechette, A., Oliveira, Y. G., Hughes, E., Mathewson, K. W., Mendolicchio, P., Pawar, J., Pislar, M., Platonov, A., Senter, E., Singh, S., Zacherl, A., and Zhang, L. M.
Learning Robust Real-Time Cultural Transmission without Human Data, March 2022.

Ten, A., Oudeyer, P.-Y., and Moulin-Frier, C.
Curiosity-driven exploration: Diversity of mechanisms and functions.
In *The Drive for Knowledge: The Science of Human Information Seeking*. Cambridge University Press, 2022.

Tessler, M. H., Madeano, J., Tsividis, P. A., Harper, B., Goodman, N. D., and Tenenbaum, J. B.
Learning to solve complex tasks by growing knowledge culturally across generations.
In *NeurIPS 2021 Cooperative AI Workshop*. arXiv, December 2021.
doi: 10.48550/arXiv.2107.13377.

Tomasello, M.
The Cultural Origins of Human Cognition.
Harvard university press, 2009.

Tomasello, M.
Becoming Human: A Theory of Ontogeny.
Harvard University Press, 2019.
ISBN 978-0-674-98085-3.

Tomasello, M. and Carpenter, M.
Shared intentionality.
Developmental Science, 10(1):121–125, 2007.
ISSN 1467-7687.
doi: 10.1111/j.1467-7687.2007.00573.x.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H.
Understanding and sharing intentions: The origins of cultural cognition.
The Behavioral and Brain Sciences, 28(5):675–735, 2005.
ISSN 0140-525X.
doi: 10.1017/S0140525X05000129.

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E.
Two Key Steps in the Evolution of Human Cooperation.
Current Anthropology, 53(6):673–692, December 2012.
ISSN 0011-3204.
doi: 10.1086/668207.

Trauth, M. H., Maslin, M. A., Deino, A. L., Junginger, A., Lesoloyia, M., Odada, E. O., Olago, D. O., Olaka, L. A., Strecker, M. R., and Tiedemann, R.
Human evolution in a variable environment: The amplifier lakes of Eastern Africa.
Quaternary Science Reviews, 29(23-24):2981–2988, 2010.
ISSN 02773791.
doi: 10.1016/j.quascirev.2010.07.007.

Van Valen, L.
A new evolutionary law.
1973.

Vanschoren, J.
Meta-learning: A survey.
arXiv:1810.03548 [cs, stat], 2018.

Varela, F. J.
Patterns of life: Intertwining identity and cognition.
Brain and cognition, 34(1):72–87, 1997.

Varela, F. J., Thompson, E., and Rosch, E.
The Embodied Mind: Cognitive Science and Human Experience.
MIT Press, 1991.
ISBN 0-262-26123-5.

Venkattaramanujam, S., Crawford, E., Doan, T., and Precup, D.
Self-supervised learning of distance functions for goal-conditioned reinforcement learning.
arXiv preprint arXiv:1907.02998, 2019.

Verschure, P. F. M. J., Pennartz, C. M. A., and Pezzulo, G.
The why, what, where, when and how of goal-directed choice: Neuronal and computational principles.
Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1655):1–14, 2014.

Vinyals, O., Babuschkin, I., Chung, J., Mathieu, M., Jaderberg, M., Czarnecki, W. M., Dudzik, A., Huang, A., Georgiev, P., Powell, R., Ewalds, T., Horgan, D., Kroiss, M., Danihelka, I., Agapiou, J., Oh, J., Dalibard, V., Choi, D., Sifre, L., Sulsky, Y.,

Vezhnevets, S., Molloy, J., Cai, T., Budden, D., Paine, T., Gulcehre, C., Wang, Z., Pfaff, T., Pohlen, T., Wu, Y., Yogatama, D., Cohen, J., McKinney, K., Smith, O., Schaul, T., Lillicrap, T., Apps, C., Kavukcuoglu, K., Hassabis, D., and Silver, D.
AlphaStar: Mastering the Real-Time Strategy Game StarCraft II, 2019.

Vrba, E. S.
Environment and evolution: Alternative causes of the temporal distribution of evolutionary events.
South African Journal of Science, 81:229–236, 1985.

Vygotsky, L. S.
Tool and Symbol in Child Development.
In *Mind in Society*. Harvard University Press, 1930.

Vygotsky, L. S.
Play and Its Role in the Mental Development of the Child.
Soviet Psychology, 1933.

Vygotsky, L. S.
Thought and Language.
MIT press, 1934.

Vygotsky, L. S.
Tool and Symbol in Child Development.
In *Mind in Society*, pp. 19–30. Harvard University Press, 1978.
ISBN 0-674-57629-2.

Walker, S. I., Packard, N., and Cody, G. D.
Re-conceptualizing the origins of life.
Philosophical transactions. Series A, Mathematical, physical, and engineering sciences, 375(2109):20160337, December 2017.
ISSN 1364-503X.
doi: 10.1098/rsta.2016.0337.

Wang, J. X., Hughes, E., Fernando, C., Czarnecki, W. M., Duenez-Guzman, E. A., and Leibo, J. Z.
Evolving intrinsic motivations for altruistic behavior.
arXiv:1811.05931 [cs], 2019a.

Wang, J. X., King, M., Porcel, N., Kurth-Nelson, Z., Zhu, T., Deck, C., Choy, P., Cassin, M., Reynolds, M., Song, F., Buttimore, G., Reichert, D. P., Rabinowitz, N., Matthey, L., Hassabis, D., Lerchner, A., and Botvinick, M.
Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents, 2021.

Wang, R., Lehman, J., Clune, J., and Stanley, K. O.
Paired Open-Ended Trailblazer (POET): Endlessly Generating Increasingly Complex and Diverse Learning Environments and Their Solutions.
arXiv:1901.01753 [cs], 2019b.

Wang, R. E., Kew, J. C., Lee, D., Lee, T.-W. E., Zhang, T., Ichter, B., Tan, J., and Faust, A.
Model-based reinforcement learning for decentralized multiagent rendezvous.
arXiv:2003.06906 [cs], 2020.

Weinbaum, D. and Veitas, V.
Open ended intelligence: The individuation of intelligent agents.
Journal of Experimental & Theoretical Artificial Intelligence, 29(2):371–396, 2017.
ISSN 0952-813X.
doi: 10.1080/0952813X.2016.1185748.

Whorf, B. L.
Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.
MIT press, 1956.

Wiessner, P. W.
Embers of society: Firelight talk among the Ju/'hoansi Bushmen.
Proceedings of the National Academy of Sciences, 111(39):14027, September 2014.
doi: 10.1073/pnas.1404212111.

Wikipedia.
Minecraft — Wikipedia, the free encyclopedia, 2022.

Yang, J., Nakhaei, A., Isele, D., Fujimura, K., and Zha, H.
CM3: Cooperative Multi-goal Multi-stage Multi-agent Reinforcement Learning.
In *International Conference on Learning Representations*, 2019.

Zheng, S., Trott, A., Srinivasa, S., Naik, N., Gruesbeck, M., Parkes, D. C., and Socher, R.
The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies.
arXiv:2004.13332 [cs, econ, q-fin, stat], April 2020.

List of Figures