Grounding an Ecological Theory of Artificial Intelligence in Human Evolution

Eleni Nisioti, Katia Jodogne-del Litto, Clément Moulin-Frier

Grounding an Ecological Theory of Artificial Intelligence in Human Evolution

Eleni Nisioti
Flowers Team, Inria and Ensta ParisTech
Bordeaux, France
eleni.nisioti@inria.fr

Katia Jodogne-del Litto
Flowers Team, Inria and Ensta ParisTech
Bordeaux, France

Clément Moulin-Frier
Flowers Team, Inria and Ensta ParisTech
Bordeaux, France
clement.moulin-frier@inria.fr

Abstract

Recent advances in Artificial Intelligence (AI) have revived the quest for agents able to acquire an open-ended repertoire of skills. Although this ability is fundamentally related to the characteristics of human intelligence, research in this field rarely considers the processes and ecological conditions that may have guided the emergence of complex cognitive capacities during the evolution of the species.

Research in Human Behavioral Ecology (HBE) seeks to understand how the behaviors characterizing human nature can be conceived as adaptive responses to major changes in our ecological niche. In this paper, we propose a framework highlighting the role of environmental complexity in open-ended skill acquisition, grounded in major hypotheses from HBE and recent contributions in Reinforcement learning (RL). We use this framework to highlight fundamental links between the two disciplines, as well as to identify feedback loops that bootstrap ecological complexity and create promising research directions for AI researchers. We also present our first steps towards designing a simulation environment that implements the climate dynamics necessary for studying key HBE hypotheses relating environmental complexity to skill acquisition.

1 Introduction

Be it morphological, behavioral or cultural, the open-endedness of biological life has been a puzzle for researchers in natural sciences trying to analyze it [1, 2] and an inspiration for researchers in Artificial Intelligence (AI) trying to implement it [3, 4]. While a definition of AI has long eluded scientists, it has been proposed that a key property of an intelligent agent may be its ability to adapt to an open-ended set of environments [5, 6]. From a reinforcement learning (RL) perspective, an agent with open-ended learning abilities should be able to adapt to an unbounded set of diverse tasks [4, 7], a new paradigm that comes in contrast to the classical approach of explicitly engineering learning algorithms upon encountering a new task. A significant part of past and present RL literature is concerned with the design of new: (i) algorithms and architectures for learning [8, 9] (ii) cost functions [10, 11, 9] (iii) benchmarks and environments [12, 13, 8, 14].

A seminal observation in the study of the evolution of cognitive processes such as perception and problem-solving, is that they did not emerge out of an unconstrained optimization process but were rather largely shaped by the ecologies they inhabited [15]. In the AI community, a similar intellectual dialogue is unfolding: intelligence is only as general as its environment requires, an observation that

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

suggests that we should abandon the quest for artificial general intelligence [6, 16, 5] and progress our understanding of how the environment biases the learning abilities of agents [17, 18]. In deep RL in particular, past experiences of agents bias their learning abilities by shaping internal reward mechanisms, studied under the term of intrinsic motivation [19] and representation functions [20] and acting as a curriculum [21, 18]. Under this new paradigm, it has been proposed to optimize RL benchmarks for their curriculum building ability [18] and study how environmental properties in reset-free, sparse reward settings impact the problem-solving abilities of agents [22].

Despite this progress, research in RL does not often acknowledge that open-ended skill acquisition is fundamentally related to the characteristics of human intelligence [3]. In this work, we take a step back from our computer-scientific lens and turn our attention towards Human Behavioral Ecology (HBE), a field concerned with the effect that ecology has had on the evolution of the human species [23, 2, 24]. Works in this field have studied among others how speciation, extinction and dispersal arose in the human history [23], cooperative groups [25], resource management [26], tool use [27], the development of human language [28] and the emergence of cultural norms and institutions [29].

Admittedly, there are many paths to the acquisition of open-ended skills in AI; grounding our study in human ecology seems to be but one of the options. But there's a number of reasons that may persuade us to explore it: (i) examining *all* possible ecologies is infeasible considering our modern and foreseeable computational power [18] (ii) ecologies that are more familiar to ours make it easier to define evaluation criteria. For example, human-ecology inspired metrics such as equality, sustainability and social welfare have been employed to evaluate agents on their ability to forage [30], find optimal taxation strategies [31] and play games [32] (iii) Darwinian evolution offers an existence proof for human-like open-ended skill acquisition [18], as well as empirical data and testable hypotheses (iv) 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 [33, 34] and embodied language acquisition [35] have had a significant impact on modern AI research (v) the potential of knowledge transfer between HBE and RL has already been recognized [36], with the transfer of ideas having the opposite direction from the one proposed here. A proposal to study major evolutionary transitions in ecology in order to understand the general laws that underlie innovation and transfer insights to artificial evolution is presented in [37]. Our proposal follows a similar direction but focuses on highlighting the overlap between concepts in RL and HBE. In addition, a number of works in RL have recently resorted to theories from ecology, psychology and economics for inspiration [38–40, 30, 41].

Our proposal can be seen as an attempt to ground an "ecological theory of RL" in hypotheses from HBE, motivated by the observation that paleoclimatology data offer us insights into how environments offered affordances for humans to evolve cognitive mechanisms that could potentially drive artificial agents towards open-ended skill acquisition. In this work, we: (i) examine a range of hypotheses from HBE and map them to key research question in RL (ii) present a conceptual framework that formalizes links between HBE and RL (iii) identify desiderata for a simulation environment that enables experimentation with HBE hypotheses related to environmental complexity and present a preliminary examination of resource availability patterns emerging in such an environment.

Section 2 provides related background, in particular offering a review of recent advances in AI from the perspective of open-ended skill acquisition and a bird's eye view of the field of HBE, focusing on hypotheses associating climate variability to major events in human evolution that took place at the Rift Valley approximately 5 million years ago. Finally, Section 3 presents our contributions towards a dialogue between the AI and HBE communities, in the form of a shared conceptual framework and an ecologically-valid playground for the study of skill acquisition in AI.

2 Two monologues on open-ended skill acquisition

Although both fields of AI and HBE study open-ended skill acquisition today, they differ significantly in the trajectories they have followed: open-endedness has been central for HBE since its birth, as the curiosity for it in essence defined the field. Research on AI on the other hand, has only recently reached the maturity levels required to experiment with open-endeness in simulated worlds.

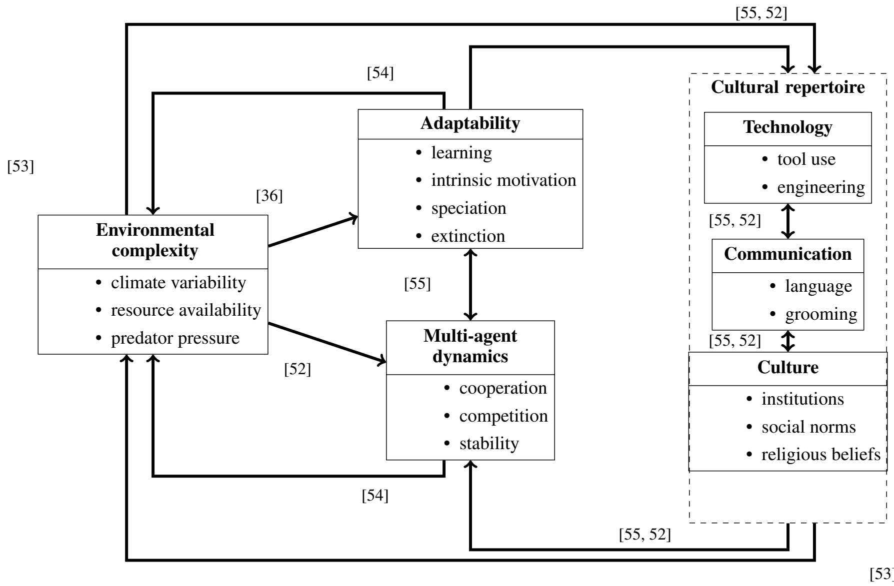

Figure 1: Environmental complexity as a main driver in human behavioral ecology. Feed-forward and feedback arrows indicate relationships between the different ecological components, analyzed in the corresponding references from HBE literature provided as labels.

2.1 Perspectives from Artificial Intelligence

The interplay between environmental complexity and open-ended skill acquisition in intelligent agents has been investigated from various perspectives. Single-agent settings have focused on different elements of an agent's architecture: (i) the neural networks employed for function approximation, whose generalization abilities are an ongoing debate [42, 43]. It has been proposed that the ability of a neural network to generalize emerges in a complex environment, characterized by multi-modal signals situated in temporally and physically rich spaces that allow for diversity in the agent's perspective [44] (ii) the cost function or type of intrinsic motivation considered. Useful skills may emerge as an agent minimizes future surprise, attempting to counteract the uncertainty inherent in its environment [10]. In curiosity-driven exploration, learning progress generates intrinsic rewards that push an agent to explore and create its own learning curricula [19, 45].

In the multi-agent reinforcement learning (MARL) literature, the automatic discovery of new environments is achieved by multi-agent autocurricula, where environmental complexity arises due to the co-existence of multiple agents [17, 21, 32, 46]. In addition to self-play originally used in two-player problem settings [47], the presence of multiple agents can give rise to an arms race [32] or create population dynamics that lead to the emergence of cooperation [30] and a drive for exploration [46].

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 RL algorithm operating on a given environment, analog to a developmental learning process. The outer level is tuning the parameters of the inner loop such that it performs well on a distribution of environments, analog to an evolutionary process. Mechanisms are either gradient-based [48] or memory-based [49].

The recent introduction of quality-diversity algorithms [50] has signified a departure from a purely performance-based view of artificial evolution and has renewed interest in mechanisms related to the preservation of diversity arising in natural evolution. An important link between artificial and natural evolution is made by *behavioral niches*, which, resembling ecological niches, introduce local competition in the evolutionary dynamics and, thus, contribute to higher diversity [51].

2.2 Perspectives from Human Behavioral Ecology

Which factors contributed to the manifested ability of humans to generalize? What differentiated the human species from others that went extinct due to their inability to adapt to novel environments? These are some of the important questions that have preoccupied HBE, a field that emerged from anthropology and is today closely related to evolutionary psychology and cultural evolution. The spotlight is on the Rift Valley at East Africa during a period that lasted approximately from 7 to 2 million years ago. This period constitutes a turning point in our evolutionary trajectory characterized by the first appearance of modern humans and their expansion to other geographical areas [23]. This evolutionary leap was originally studied under theories that layed 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 [56, 23, 57].

In Figure 1, we introduce a conceptual framework that recognizes important ecological components and the feedforward and feedback links relating them. In the remainder of this section, we discuss hypotheses studying these relationships in the human ecosystem and, in Section 3.1, associate them with research questions in the study of artificial ecosystems. Under this framework, environmental complexity is essentially driven by climate variability, which implies instability in the ecological conditions, in particular through changes in resource availability and exposition to predators [54, 36, 2]. This complexity has a strong influence on two major phenomena. First, it drives adaptability both at the evolutionary time scale, through speciation and extinction, [24, 56] and at the developmental time scale through cognitive mechanisms for exploration, learning and abstraction [58, 36]. Second, varying the levels of resource availability and exposition to predators has a strong influence on multi-agent dynamics through the modulation of cooperation and competition pressures [55, 59].

The influence of environmental complexity on adaptation and multi-agent dynamics can then have feedback and feedforward effects on the ecological system. First, increased morphological and cognitive complexity due to adaptation, as well as increased complexity in the multi-agent dynamics, feed back to environmental complexity through the modification of resource availability and predation pressure [53, 52]. For example, the Red Queen hypothesis [60] proposes that competition among different species is a major drive of evolution, possibly driving an arms race between co-adapting species. Second, adaptation and multi-agent dynamics can bootstrap in a feedforward manner the emergence of more advanced behaviors related to technology (e.g. tool use [27]), communication (e.g. language [59, 28]) and culture (e.g. social norms [59], institutions [59] and religions [61]). Here again, the emergence of these new behaviors feeds back into environmental complexity through the process of social niche construction [53], thus creating a positive feedback loop potentially driving the ever-expanding social complexity of human ecology [54, 62].

3 A proposed dialogue between Artificial Intelligence and Ecology

3.1 Towards a shared conceptual framework

There seems to be a significant overlap between the questions that RL research poses in its study of the acquisition of open-ended behavior and the hypotheses examined by HBE. In this section, we focus on three emergent phenomena attracting the interest of the RL community: adaptability of individual agents, multi-agent dynamics of groups and their cultural repertoire. By referring to concepts depicted in Figure 1 and highlighted in this section, we initiate a dialogue between the two fields and identify key links that we believe deem further investigation by the RL community.

3.1.1 Adaptability

Insights from ecology Under the PCV framework discussed in Section 2.2, environmental factors such as **climate variability**, **resource availability** and **predation pressure** have served as a drive for the ability of humans to adapt to complex environments. Adaptability is achieved through mechanisms whose form 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: based on environmental observations an agent may be able to switch between different behaviors following innate, and not learned instructions [58, 63, 36].

In cases where the environment changes noticeably across generations but slowly enough within a generation , behavioral plasticity is guided by a process of developmental selection, an example of which is the *learning* process, where an agent's past behavior guide its future behavior. Another example is intrinsic motivation, i.e. the evolution of an internal mechanism rewarding e.g. play and exploration, independently of any external rewards [64]. Thus, adaptation to environmental conditions operates on two scales: the evolutionary one drives **speciation** and **extinction**, shaping the developmental one which drives **learning** and **exploration**. Adaptation feeds back into environmental complexity by affecting how environmental changes affect species, equipped with different skill repertoires. For example, during dry periods, the extinction rates of generalist species would reduce as they would be better able to find resources, while specialist species would struggle having lost their environmental niche and their competitive advantage [2].

State of the art in RL A recent work studied how environmental dynamics and, in particularly, the initial state and transition dynamics, affect the behavior of deep RL agents in non-episodic settings [22]. An important observation is that the artificial reset mechanism classically employed in RL to address the problem of sparse rewards can be replaced by environmental shaping in a non-episodic setting, which shifts the focus of the design from the algorithmic setup to the environment. The outer and inner loop optimization procedure that meta-RL algorithms follow matches well with the aforementioned biological mechanisms of adaptation. There is a lot of flexibility in the choice of algorithms used to optimize the two loops, with both evolutionary and gradient-based optimization being applicable on the outer loop [65, 66, 48, 67]. This comes in agreement with recent proposals to view evolution as equivalent to learning [68] and development [69]. Based on ecological insights, we can indicate the following research directions for adaptability: (i) as the optimal choice of the adaptation mechanism depends on dynamic properties of environmental variability, the community can explore how different mechanisms emerge for different ecological condition. This could for example be done by studying the environmental conditions that favor the emergence of innate, learned and intrinsically motivated behaviors, following the spirit of recent works [70, 71] (ii) tools from RL can be used to test the predictions proposed by hypotheses under the PCV framework in simulation environments, potentially offering insights to both communities.

3.1.2 Multi-agent dynamics

Insights from ecology 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* [72], altruism emerged in small-scale groups due to kin selection or reciprocity. In contrast, the interdependence hypothesis [52] 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** between co-existing groups and species also gives rise to arms races, where reciprocal selection and adaptation lead to co-evolution [73]. Even at this small scale, the multi-agent dynamics feed back into environmental complexity through the source of social niche construction: predation, nutrient excretion and habitat modification populations alter their environment and further influence future populations [54].

State of the art in RL The emergence of cooperation has attracted significant interest in the MARL community. In particular, sequential social dilemmas [] have attracted a significant amount of attention. Here, groups of adaptive agents need to solve tasks such as foraging and hunting of big animals [30]. An important departure from works on the classical social dilemmas studied by game theory, is that the payoff matrices here are not given by the human designer but emerge from the resource availability patterns and multi-agent dynamics, having the form of empirical payoff matrices. In addition, Recent works have studied the role of intrinsic motivation based on the theories of assortative matching and group selection [38], inequity aversion based on fairness norms [39] and social influence [40]. Ecology-inspired hierarchical organizations have been used to facilitate decentralized learning [74]. The feedback effect that multi-agent dynamics have in environmental complexity has been studied from the perspective of multi-agent autocurricula [17] and arms races between competing groups [32]. The effect that population dynamics have on the environment was investigated in [46], where increase in population size indirectly lead to exploration. As our brief discussion of related HBE literature however reveals, there exist a number of hypotheses and observations that researchers can leverage to further advance research in MARL: (i) according to the inter-dependence hypothesis, the human drive to cooperate was born neither in scenarios that required altruism nor in social dilemmas, which have served as an application ground for the majority of works in MARL. Rather, cooperation arose in Stag hunt type situations, which favored mutualistic collaboration [52]. Thus, human ecology can offer us insights on the types of social dilemmas we should focus on and, in particular, in the order in which we need to attack them.; (ii) group properties such as size and social structure are directly related to multi-agent dynamics such as the speed of information spread [75]. Thus, their influence on the emerging multi-agent autocurricula requires investigation.

3.1.3 Cultural repertoire

Insights from ecology Non-human species often exhibit impressive behavioral repertoires [76]. However, human ecology is characterized by a uniquely large behavioral repertoire: **engineering**, **language**, **social norms**, **institutions** and **religious beliefs** constitute a complex cultural ecosystem that has lead scientists on the search of factors that differentiated us from other species [52, 77, 61]. 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 [76] 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 [55]. The feedforward and feedback links associated with **tool use** have also been investigated under a number of, often contesting, hypotheses. Based on the data analysis in [53], environmental variability such as 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 [78].

Another important link at this level is the relationship between tool use, language and adaptability. Studies of biological motor systems and language acquisition in infants have revealed that action and language representation share a similar compositional structure [35]. To understand how this similarity between two apparently distinct systems arose, one needs to turn to the origins of this relationship in human ecology. According to the Corballis hypothesis [79], the ability of primates' to manipulate tools may have played a pivotal role in the evolution of language by creating the cognitive representations that compositionality requires. At the same time, the compositional structure of language is hypothesized to be an enabler of flexible and adaptable behavior, thus feeding back to adaptability [35]. Finally, it has also been proposed that intrinsic motivation, a mechanism that might have evolved for quickly adapting to rapidly changing environments [80], can guide and constrain evolution by constituting a reservoir of behavioral and cognitive innovations which can be later on recruited for functions not yet anticipated [19].

State of the art in RL The AI community has been studying language from two distinct perspectives: Natural Language Processing and emergent communication [81]. While the former has achieved impressive results in tasks like translation and text generation, it ignores functional properties of communication, focusing on structural properties of language. In contrast, the emergence of communication in MARL systems is closer to real-world settings, but has mostly been applied in environments that are relatively simple [82] or not ecologically-valid [83]. In a similar spirit, MARL has also studied social norms and conventions [41, 84, 85]. The effect that the structure of organization has on communication learning in groups of deep reinforcement learning agents is investigated in [86]. This work constitutes an important first step in the realm of the social complexity hypothesis [55], but there remain a number of research directions lying at the intersection of MARL and meta RL: (i) the feedback effect that the cultural repertoire has on environmental variability through cultural niche construction [53] can potentially create more powerful autocurricula than the already studied ones based on niche construction in small-scale groups [32, 30]; (ii) studying the stabilization effect of cultural niche construction can provide important insights to the problem of scaling up artificial multi-agent systems ; (iii) the relationship between action/language compositionality and the ability of agents to generalize and adapt needs to be further investigated in order to transfer insights from human language acquisition to intelligent agents [45].

3.2 Towards an ecologically-valid environment

Climate variability is a key element in the framework of Figure 1. A variety of long-standing hypotheses in HBE highlight the importance of climate dynamics in providing a wide diversity of environmental constraints and opportunities for evolution. We, therefore, believe that the first step in a dialogue between the HBE and AI communities is to model such climate dynamics and propose the following desiderata for an ecologically-valid environment, in particular the implementation of: (i) unbounded and realistic climate dynamics. Rather than requiring explicit design, patterns of resource availability and exposition to predators thus emerge naturally, potentially exhibiting complexdynamics (ii) spatial open-endedness, a requirement for the appearance of multiple niches. This will enable the study of dispersal, which has been linked to evolvability [87] (iii) a variety of tasks relevant to human evolution, such as navigation, harvesting, hunting and crafting through tool use. This is crucial for enabling behavioral diversity, an important property of an open-ended system [50, 88] (iv) environmental variability at multiple spatiotemporal scales, e.g., seasonal fluctuations and day/night variation. This feature will enable the study of the interaction between evolution, development and learning [89].

The recent surge of the RL community for open-ended skill acquisition has led to the creation of many exciting environments [90, 91]. 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 [92]. 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. Specifically, the creation and deletion of items is 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 patterns, such as clusters and custom, spatially non-stationary distributions.

To enable the empirical experimentation of hypotheses proposed under the PCV framework, we have equipped JBW with climate dynamics. Our objective is to observe the appearance of lakes with interesting dynamics, such as quick expansion and chaotic contraction [93], which in its turn will modulate the presence of resources available to the agents. To achieve this, we have extended the existing item creation mechanism 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") that grow near lakes and trees, which cannot be consumed by the agent and act as obstacles. The creation of these items follows interaction and intensity functions influenced by the two newly introduced control mechanisms of precipitation and humidity. Specifically, the climate dynamics follow a four-step process: (i) the user inputs the pattern of precipitation, a function of time and position that can be specially and temporally variable. The user can tune this function in order to model a desired relationship between the timescales at which climate variablity and learning dynamics take place. (ii) each cell of the grid-world exhibits a humidity level computed based on precipitation and the proximity of lakes. Humidity has a similar effect to precipitation and acts as a buffer mechanism that slows down the death rate of water cells during periods of low precipitation. (iii) the birth and death of water cells that form lakes depends on the levels of precipitation and humidity, with clustered water cells being less likely to disappear compared to isolated ones; (iv) the presence of resources is also influenced by humidity levels and the presence of lakes. Resources are more likely to appear near bigger lakes and interact with other items in their neighborhood. They may disappear due to their predetermined lifetime or due to low levels of humidity. Some resources are directly consumable (e.g. jelly-beans and bananas), while others act as barriers and enable the growth of resources around them (e.g. trees)..

A simplified model of our proposed climate dynamics is depicted in Figure 2(a), while Figure 2(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 2(c)) and high precipitation (2(d)).

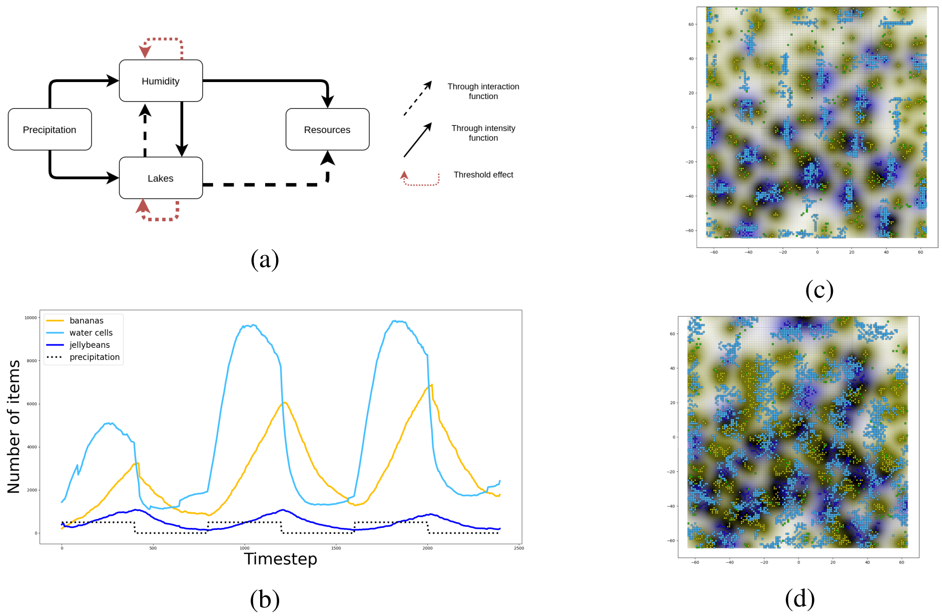

Figure 2: Climate dynamics in our proposed environment: (a) simplified model of the climate dynamics (b) temporal patterns of lake and item presence during simulations with a precipitation function having a pulse form (c) a top-view of a gridworld where an agent navigates in a grid-world populated by 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) amd a high-precipitation period (d)

This first step towards an ecologically-valid environment has obvious limitations: it only addresses the first level in our conceptual framework and the implementation of multi-scale variability and tool use has been left for future work. Nevertheless, by customizing the precipitation function the current form our playground can be used to test various hypotheses related to resource consumption, speciation and dispersal patterns, which can lead to novel insights in open-ended skill acquisition. Our prediction is that our proposed environment will pose a challenge for standard methods in RL: deep RL agents will struggle with non-stationarity while meta RL agents are ill-suited for the sequential nature of the climate dynamics, as they are traditionally applied on independently sampled tasks. We believe that bi-level optimization [94, 95] is an interesting direction as it can model the interaction between evolutionary and developmental processes.

4 Discussion

Our proposal is but a preliminary step towards realizing the potential of a cross-disciplinary dialogue between the HBE and RL communities, a glimpse of which has already been offered by recent works [36]. On one hand, our discussion reveals that the potential of RL as a computational tool for enriching the analytical toolbox of HBE has not been fully realized. On the other hand, our proposal illustrates that AI research can gain more inspiration once a better overall picture and a thorough examination of feedforward and feedback links taking place at different ecological levels has taken place. We believe that conceptual frameworks, such as the one that guided our current analysis in Figure 1, can serve as an important basis for the proposed inter-disciplinary dialogue, with different questions zooming in on different sub-parts and potentially revealing lower-level relationships, and that designing simulation environments with realistic climate dynamics, as the one proposed in this work, is key to moving forward. We believe that this conceptual framework can lead to even broader perspectives, in particular by modeling the interaction between ecological, environmental, developmental and cultural dynamics, modulating the game-theoretic structure of the environment and allowing for item compositionality. This will require the design of appropriate environments, following the spirit of our current proposal.

Acknowledgements

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,

References

[1] Charles Darwin. *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.

[2] Gillian R. Brown, Thomas E. Dickins, Rebecca Sear, and Kevin N. Laland. Evolutionary accounts of human behavioural diversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1563):313–324, February 2011. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2010.0267. URL `https://royalsocietypublishing.org/doi/10.1098/rstb.2010.0267`.

[3] Wolfgang Banzhaf, Bert Baumgaertner, Guillaume Beslon, René Doursat, James A. Foster, Barry McMullin, Vinicius Veloso de Melo, Thomas Miconi, Lee Spector, Susan Stepney, and Roger White. Defining and simulating open-ended novelty: requirements, guidelines, and challenges. *Theory in Biosciences*, 135(3):131–161, September 2016. ISSN 1431-7613, 1611-7530. doi: 10.1007/s12064-016-0229-7. URL `http://link.springer.com/10.1007/s12064-016-0229-7`.

[4] Rui Wang, Joel Lehman, Jeff Clune, and Kenneth O. Stanley. Paired Open-Ended Trailblazer (POET): Endlessly Generating Increasingly Complex and Diverse Learning Environments and Their Solutions. *arXiv:1901.01753 [cs]*, February 2019. URL `http://arxiv.org/abs/1901.01753`. arXiv: 1901.01753.

[5] François Chollet. On the measure of intelligence. *CoRR*, abs/1911.01547, 2019. URL `http://arxiv.org/abs/1911.01547`.

[6] Joel Lehman, Jeff Clune, and Sebastian Risi. An Anarchy of Methods: Current Trends in How Intelligence Is Abstracted in AI. *IEEE INTELLIGENT SYSTEMS*, page 7, 2014.

[7] Open Ended Learning Team, Adam Stooke, Anuj Mahajan, Catarina Barros, Charlie Deck, Jakob Bauer, Jakub Sygnowski, Maja Trebacz, Max Jaderberg, Michael Mathieu, Nat McAleese, Nathalie Bradley-Schmieg, Nathaniel Wong, Nicolas Porcel, Roberta Raileanu, Steph Hughes-Fitt, Valentin Dalibard, and Wojciech Marian Czarnecki. Open-ended learning leads to generally capable agents, 2021.

[8] Maruan Al-Shedivat, Trapit Bansal, Yuri Burda, Ilya Sutskever, Igor Mordatch, and Pieter Abbeel. Continuous Adaptation via Meta-Learning in Nonstationary and Competitive Environments. *arXiv:1710.03641 [cs]*, February 2018. URL `http://arxiv.org/abs/1710.03641`. arXiv: 1710.03641.

[9] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. volume 80 of *Proceedings of Machine Learning Research*, pages 1861–1870, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL `http://proceedings.mlr.press/v80/haarnoja18b.html`.

[10] Glen Berseth, Daniel Geng, Coline Devin, Nicholas Rhinehart, Chelsea Finn, Dinesh Jayaraman, and Sergey Levine. SMiRL: Surprise Minimizing Reinforcement Learning in Dynamic Environments. *arXiv:1912.05510 [cs, stat]*, July 2020. URL `http://arxiv.org/abs/1912.05510`. arXiv: 1912.05510.

[11] Abbas Abdolmaleki, Jost Tobias Springenberg, Yuval Tassa, Remi Munos, Nicolas Heess, and Martin Riedmiller. Maximum a Posteriori Policy Optimisation. *arXiv:1806.06920 [cs, math, stat]*, June 2018. URL `http://arxiv.org/abs/1806.06920`. arXiv: 1806.06920.

[12] Joseph Suarez, Yilun Du, Phillip Isola, and Igor Mordatch. Neural MMO: A Massively Multiagent Game Environment for Training and Evaluating Intelligent Agents. *arXiv:1903.00784 [cs, stat]*, March 2019. URL `http://arxiv.org/abs/1903.00784`. arXiv: 1903.00784.

[13] Jonathan Schwarz, Daniel Altman, Andrew Dudzik, Oriol Vinyals, Yee Whye Teh, and Razvan Pascanu. Towards a natural benchmark for continual learning. page 13, 2019.

[14] Karl Cobbe, Oleg Klimov, Chris Hesse, Taehoon Kim, and John Schulman. Quantifying generalization in reinforcement learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1282–1289, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL `http://proceedings.mlr.press/v97/cobbe19a.html`.

[15] James J. Gibson. *The Ecological Approach to Visual Perception: Classic Edition*. Houghton Mifflin, 1979.

[16] Roman V. Yampolskiy. Human $\neq$ agi, 2020.

[17] Joel Z Leibo, Edward Hughes, Marc Lanctot, and Thore Graepel. Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. *arXiv preprint arXiv:1903.00742*, 2019.

[18] Jeff Clune. AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence. *arXiv:1905.10985 [cs]*, January 2020. URL `http://arxiv.org/abs/1905.10985`. arXiv: 1905.10985.

[19] Pierre-Yves Oudeyer and Linda B. Smith. How Evolution May Work Through Curiosity-Driven Developmental Process. *Topics in Cognitive Science*, 8(2):492–502, 2016. ISSN 1756-8765. doi: 10.1111/tops.12196.

[20] Max Schwarzer, Nitarshan Rajkumar, Michael Noukhovitch, Ankesh Anand, Laurent Charlin, Devon Hjelm, Philip Bachman, and Aaron Courville. Pretraining Representations for Data-Efficient Reinforcement Learning. *arXiv:2106.04799 [cs]*, June 2021. URL `http://arxiv.org/abs/2106.04799`. arXiv: 2106.04799.

[21] Remy Portelas, Cédric Colas, Lilian Weng, Katja Hofmann, and Pierre-Yves Oudeyer. Automatic Curriculum Learning For Deep RL: A Short Survey. *arXiv:2003.04664 [cs, stat]*, May 2020. URL `http://arxiv.org/abs/2003.04664`. arXiv: 2003.04664.

[22] John D. Co-Reyes, Suvansh Sanjeev, Glen Berseth, Abhishek Gupta, and Sergey Levine. Ecological reinforcement learning, 2020.

[23] Mark A. Maslin, Susanne Shultz, and Martin H. Trauth. A synthesis of the theories and concepts of early human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1663):20140064, March 2015. doi: 10.1098/rstb.2014.0064. URL `https://royalsocietypublishing.org/doi/10.1098/rstb.2014.0064`. Publisher: Royal Society.

[24] Rebecca Sear, David W. Lawson, and Thomas E. Dickins. 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. URL `https://akjournals.com/doi/10.1556/jep.2007.1019`.

[25] Colin A Chapman and Lauren J Chapman. Constraints on Group Size in Red Colobus and Red-tailed Guenons: Examining the Generality of the Ecological Constraints Model. page 21, 2000.

[26] J. A. J. Gowlett. The discovery of fire by humans: a long and convoluted process. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1696):20150164, June 2016. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2015.0164. URL `https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0164`.

[27] Sophie A. de Beaune. The Invention of Technology: Prehistory and Cognition. *Current Anthropology*, 45(2):139–162, April 2004. ISSN 0011-3204, 1537-5382. doi: 10.1086/381045. URL `https://www.journals.uchicago.edu/doi/10.1086/381045`.

[28] Todd M. Freeberg, Robin I. M. Dunbar, and Terry J. Ord. 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, 1471-2970. doi: 10.1098/rstb.2011.0213. URL `https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0213`.

[29] Michael Tomasello. *The cultural origins of human cognition*. Harvard university press, 2009.

[30] Julien Pérolat, Joel Z. Leibo, Vinicius Zambaldi, Charles Beattie, Karl Tuyls, and Thore Graepel. A multi-agent reinforcement learning model of common-pool resource appropriation. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 3643–3652, 2017.

[31] Stephan Zheng, Alexander Trott, Sunil Srinivasa, Nikhil Naik, Melvin Gruesbeck, David C. Parkes, and Richard Socher. The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies. *arXiv:2004.13332 [cs, econ, q-fin, stat]*, April 2020. URL `http://arxiv.org/abs/2004.13332`. arXiv: 2004.13332.

[32] Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, and Igor Mordatch. Emergent Tool Use From Multi-Agent Autocurricula. 2020. URL `https://openreview.net/forum?id=SkxpxJBKwS`. tex.ids: Baker2019 arXiv: 1909.07528.

[33] P. Oudeyer, F. Kaplan, and V. V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, 2007. doi: 10.1109/TEVC.2006.890271.

[34] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[35] Angelo Cangelosi, Giorgio Metta, Gerhard Sagerer, Stefano Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Tony Belpaeme, Giulio Sandini, Francesco Nori, Luciano Fadiga, Britta Wrede, Katharina Rohlfing, Elio Tuci, Kerstin Dautenhahn, Joe Saunders, and Arne Zeschel. Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics. *IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT*, 2(3):29, 2010.

[36] Willem E. Frankenhuis, Karthik Panchanathan, and Andrew G. Barto. Enriching behavioral ecology with reinforcement learning methods. *Behavioural Processes*, 161:94–100, 2019. ISSN 0376-6357. doi: 10.1016/j.beproc.2018.01.008. URL `http://www.sciencedirect.com/science/article/pii/S0376635717303637`.

[37] Ricard Solé. 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. URL `https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0175`.

[38] Jane X. Wang, Edward Hughes, Chrisantha Fernando, Wojciech M. Czarnecki, Edgar A. Duenez-Guzman, and Joel Z. Leibo. Evolving intrinsic motivations for altruistic behavior. *arXiv:1811.05931 [cs]*, March 2019. URL `http://arxiv.org/abs/1811.05931`. arXiv: 1811.05931.

[39] Edward Hughes, Joel Z. Leibo, Matthew G. Phillips, Karl Tuyls, Edgar A. Duéñez-Guzmán, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin R. McKee, Raphael Koster, Heather Roff, and Thore Graepel. Inequity aversion improves cooperation in intertemporal social dilemmas. *arXiv:1803.08884 [cs, q-bio]*, September 2018. URL `http://arxiv.org/abs/1803.08884`. arXiv: 1803.08884.

[40] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro A. Ortega, DJ Strouse, Joel Z. Leibo, and Nando de Freitas. 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. URL `https://openreview.net/forum?id=B1lG42C9Km`. arXiv: 1810.08647.

[41] Raphael Köster, Kevin R. McKee, Richard Everett, Laura Weidinger, William S. Isaac, Edward Hughes, Edgar A. Duéñez-Guzmán, Thore Graepel, Matthew Botvinick, and Joel Z. Leibo. Model-free conventions in multi-agent reinforcement learning with heterogeneous preferences. *arXiv:2010.09054 [cs]*, October 2020. URL `http://arxiv.org/abs/2010.09054`. arXiv: 2010.09054.

[42] Gary F. Marcus. Rethinking Eliminative Connectionism. *Cognitive Psychology*, 37(3):243–282, December 1998. ISSN 00100285. doi: 10.1006/cogp.1998.0694. URL `https://linkinghub.elsevier.com/retrieve/pii/S0010028598906946`.

[43] Behnam Neyshabur, Srinadh Bhojanapalli, David Mcallester, and Nati Srebro. Exploring generalization in deep learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 5947–5956. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/10ce03a1ed01077e3e289f3e53c72813-Paper.pdf`.

[44] Felix Hill, Andrew Lampinen, Rosalia Schneider, Stephen Clark, Matthew Botvinick, James L McClelland, and Adam Santoro. ENVIRONMENTAL DRIVERS OF SYSTEMATICITY AND GENERALISATION IN A SITUATED AGENT. page 15, 2020.

[45] Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Clément Moulin-Frier, Peter Ford Dominey, and Pierre-Yves Oudeyer. Language as a cognitive tool to imagine goals in curiosity-driven exploration, 2020.

[46] Joel Z. Leibo, Julien Perolat, Edward Hughes, Steven Wheelwright, Adam H. Marblestone, Edgar Duéñez-Guzmán, Peter Sunehag, Iain Dunning, and Thore Graepel. Malthusian Reinforcement Learning. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, 2:1099–1107, December 2018. URL `http://arxiv.org/abs/1812.07019`. arXiv: 1812.07019 Publisher: International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

[47] David Silver, Aja Huang, Christopher J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529:484–503, 2016. URL `http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html`.

[48] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *arXiv:1703.03400 [cs]*, July 2017. URL `http://arxiv.org/abs/1703.03400`. arXiv: 1703.03400.

[49] Jane X. Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn. *arXiv:1611.05763 [cs, stat]*, January 2017. URL `http://arxiv.org/abs/1611.05763`. arXiv: 1611.05763.

[50] Justin K. Pugh, Lisa B. Soros, and Kenneth O. Stanley. Quality diversity: A new frontier for evolutionary computation. *Frontiers Robotics AI*, 3:40, 2016.

[51] Kenneth O. Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary Computation*, 10(2):99–127, 2002. doi: 10.1162/106365602320169811.

[52] Michael Tomasello, Alicia P Melis, Claudio Tennie, Emily Wyman, and Esther Herrmann. Two Key Steps in the Evolution of Human Cooperation. 53(6):21, 2020.

[53] Laurel Fogarty and Nicole Creanza. 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, December 2017. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2016.0428. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0428.

[54] David M. Post and Eric P. Palkovacs. 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, June 2009. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2009.0012. URL https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0012.

[55] R. I. M. Dunbar. 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. URL https://www.cambridge.org/core/product/identifier/S0140525X00032325/type/journal_article.

[56] Richard Potts. Hominin evolution in settings of strong environmental variability. *Quaternary Science Reviews*, 73:1–13, August 2013. ISSN 02773791. doi: 10.1016/j.quascirev.2013.04.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0277379113001340.

[57] Mark Collard, Michael Kemery, and Samantha Banks. Causes of Toolkit Variation Among Hunter-Gatherers: A Test of Four Competing Hypotheses. page 20, 2005.

[58] Dean Frederick Hougen and Syed Naveed Hussain Shah. The Evolution of Reinforcement Learning. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1457–1464, December 2019. doi: 10.1109/SSCI44817.2019.9003146.

[59] Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5):675–691, October 2005. ISSN 0140-525X, 1469-1825. doi: 10.1017/S0140525X05000129. URL https://www.cambridge.org/core/product/identifier/S0140525X05000129/type/journal_article.

[60] Paul N Pearson. *Red Queen Hypothesis*. American Cancer Society, 2001. ISBN 9780470015902. doi: https://doi.org/10.1038/npg.els.0001667. URL https://novel-coronavirus.onlinelibrary.wiley.com/doi/abs/10.1038/npg.els.0001667.

[61] Carlos A. Botero, Beth Gardner, Kathryn R. Kirby, Joseph Bulbulia, Michael C. Gavin, and Russell D. Gray. 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. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1408701111.

[62] K. N. Laland, J. Odling-Smee, and M. W. Feldman. Cultural niche construction and human evolution. *Journal of Evolutionary Biology*, 14(1):22–33, 2001. doi: https://doi.org/10.1046/j.1420-9101.2001.00262.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1420-9101.2001.00262.x.

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

[64] E L Deci and Richard M Ryan. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press.

[65] Scott Niekum, Lee Spector, and Andrew Barto. Evolution of reward functions for reinforcement learning. In *Proceedings of the 13th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '11, page 177–178, New York, NY, USA, 2011. Association for Computing Machinery. ISBN 9781450306904. doi: 10.1145/2001858.2001957. URL https://doi.org/10.1145/2001858.2001957.

[66] Djordje Grbic and Sebastian Risi. Towards continual reinforcement learning through evolutionary meta-learning. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, GECCO '19, page 119–120, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367486. doi: 10.1145/3319619.3322044. URL `https://doi.org/10.1145/3319619.3322044`.

[67] Louis Kirsch, Sjoerd van Steenkiste, and Jürgen Schmidhuber. Improving Generalization in Meta Reinforcement Learning using Learned Objectives. *arXiv:1910.04098 [cs, stat]*, February 2020. URL `http://arxiv.org/abs/1910.04098`. arXiv: 1910.04098.

[68] Richard A. Watson and Eörs Szathmáry. How Can Evolution Learn? *Trends in Ecology & Evolution*, 31(2):147–157, February 2016. ISSN 0169-5347. doi: 10.1016/j.tree.2015.11.009. URL `http://www.sciencedirect.com/science/article/pii/S0169534715002931`.

[69] C. Fields and M. Levin. Does Evolution Have a Target Morphology? *Organisms. Journal of Biological Sciences*, Vol 4:57–76 Pages, September 2020. doi: 10.13133/2532-5876/16961. URL `https://www.rspi.uniroma1.it/index.php/Organisms/article/view/16961`. Artwork Size: 57-76 Pages Publisher: Organisms. Journal of Biological Sciences.

[70] Robert Tjarko Lange and Henning Sprekeler. Learning not to learn: Nature versus nurture in silico, 2021.

[71] Ferran Alet, Martin F. Schneider, Tomas Lozano-Perez, and Leslie Pack Kaelbling. Meta-learning curiosity algorithms, 2020.

[72] Terence Burnham and Dominic Johnson. The biological and evolutionary logic of human cooperation. *Analyse & Kritik*, 27:113–35, 12 2005. doi: 10.1515/auk-2005-0107.

[73] C. Benkman, T. Parchman, Amanda Favis, and A. Siepielski. Reciprocal selection causes a coevolutionary arms race between crossbills and lodgepole pine. *The American Naturalist*, 162: 182 – 194, 2003.

[74] Sanjeevan Ahilan and Peter Dayan. Feudal multi-agent hierarchies for cooperative reinforcement learning. *CoRR*, abs/1901.08492, 2019. URL `http://arxiv.org/abs/1901.08492`.

[75] Winter A. Mason, Andy Jones, and Robert L. Goldstone. 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. URL `http://doi.apa.org/getdoi.cfm?doi=10.1037/a0012798`.

[76] Indrikis Krams, Tatjana Krama, Todd M Freeberg, Cecilia Kullberg, and Jeffrey R Lucas. Linking social complexity and vocal complexity: a parid perspective. page 13, 2012.

[77] M. Tomasello. *Becoming Human: A Theory of Ontogeny*. Harvard University Press, 2019. ISBN 978-0-674-98085-3. URL `https://books.google.fr/books?id=ZnhyDwAAQBAJ`.

[78] Mark Collard, Briggs Buchanan, April Ruttle, and Michael J O'Brien. Niche Construction and the Toolkits of Hunter–Gatherers and Food Producers. page 9.

[79] Michael Corballis. From hand to mouth: The origins of language. 01 2002.

[80] Satinder Singh, Richard L Lewis, Andrew G Barto, and Jonathan Sorg. Intrinsically motivated reinforcement learning: An evolutionary perspective. 2(2):70–82.

[81] Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. Multi-agent communication meets natural language: Synergies between functional and structural language learning, 2020.

[82] Angeliki Lazaridou and Marco Baroni. Emergent Multi-Agent Communication in the Deep Learning Era. *arXiv:2006.02419 [cs]*, July 2020. URL `http://arxiv.org/abs/2006.02419`. arXiv: 2006.02419.

[83] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. *arXiv:1605.06676 [cs]*, May 2016. URL `http://arxiv.org/abs/1605.06676`. arXiv: 1605.06676.

[84] Samhar Mahmoud, Simon Miles, and Michael Luck. Cooperation emergence under resource-constrained peer punishment. In *Proceedings of the 2016 International Conference on Autonomous Agents; Multiagent Systems*, AAMAS '16, page 900–908, Richland, SC, 2016. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450342391.

[85] Ismael T. Freire, Clement Moulin-Frier, Marti Sanchez-Fibla, Xerxes D. Arsiwalla, and Paul F. M. J. Verschure. Modeling the formation of social conventions from embodied real-time interactions. 15(6):e0234434, 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0234434. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0234434`.

[86] Marina Dubova, Arseny Moskvichev, and Robert Goldstone. Reinforcement communication learning in different social network structures, 07 2020.

[87] Joel Lehman and Kenneth Stanley. Evolvability is inevitable: Increasing evolvability without the pressure to adapt. *PloS one*, 8:e62186, 05 2013. doi: 10.1371/journal.pone.0062186.

[88] Arend Hintze. Open-endedness for the sake of open-endedness. *Artificial Life*, 25:198–206, 2019.

[89] David Stephens. Change, regularity, and value in the evolution of animal learning. *Behavioral Ecology*, 2, 03 1991. doi: 10.1093/beheco/2.1.77.

[90] Danijar Hafner. Benchmarking the spectrum of agent capabilities. *arXiv preprint arXiv:2109.06780*, 2021.

[91] Matthew Johnson, Katja Hofmann, Tim Hutton, and David Bignell. The malmo platform for artificial intelligence experimentation. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 4246–4247. AAAI Press, 2016. ISBN 9781577357704.

[92] Emmanouil Antonios Platanios, Abulhair Saparov, and Tom Mitchell. Jelly bean world: A testbed for never-ending learning, 2020.

[93] Martin H. Trauth, Mark A. Maslin, Alan L. Deino, Annett Junginger, Moses Lesoloyia, Eric O. Odada, Daniel O. Olago, Lydia A. Olaka, Manfred R. Strecker, and Ralph Tiedemann. Human evolution in a variable environment: the amplifier lakes of Eastern Africa. *Quaternary Science Reviews*, 29(23-24):2981–2988, November 2010. ISSN 02773791. doi: 10.1016/j.quascirev.2010.07.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0277379110002581`.

[94] David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Sebastian Risi, and Kenneth O. Stanley. Indirectly Encoding Neural Plasticity as a Pattern of Local Rules. In Stéphane Doncieux, Benoît Girard, Agnès Guillot, John Hallam, Jean-Arcady Meyer, and Jean-Baptiste Mouret, editors, *From Animals to Animats 11*, volume 6226, pages 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. URL `http://link.springer.com/10.1007/978-3-642-15193-4_50`. Series Title: Lecture Notes in Computer Science.

[95] Elias Najarro and Sebastian Risi. Meta-Learning through Hebbian Plasticity in Random Networks. *arXiv:2007.02686 [cs]*, March 2021. URL `http://arxiv.org/abs/2007.02686`. arXiv: 2007.02686.