



HAL
open science

A Definition of Open-Ended Learning Problems for Goal-Conditioned Agents

Olivier Sigaud, Gianluca Baldassarre, Cédric Colas, Stephane Doncieux, Richard Duro, Pierre-Yves Oudeyer, Nicolas Perrin-Gilbert, Vieri Giuliano Santucci

► **To cite this version:**

Olivier Sigaud, Gianluca Baldassarre, Cédric Colas, Stephane Doncieux, Richard Duro, et al.. A Definition of Open-Ended Learning Problems for Goal-Conditioned Agents. 2024. hal-04810195

HAL Id: hal-04810195

<https://inria.hal.science/hal-04810195v1>

Preprint submitted on 28 Nov 2024









HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

A DEFINITION OF OPEN-ENDED LEARNING FOR GOAL-CONDITIONED AGENTS

 **Olivier Sigaud**^{(1)*},  **Gianluca Baldassarre**^{(3)*},  **Cédric Colas**^(2,5),
 **Stéphane Doncieux**⁽¹⁾,  **Richard Duro**⁽⁴⁾,  **Pierre-Yves Oudeyer**⁽²⁾,
 **Nicolas Perrin-Gilbert**⁽¹⁾,  **Vieri Giuliano Santucci**⁽³⁾,

(1) Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique, F-75005 Paris, France

(2) INRIA Bordeaux Sud-Ouest, équipe FLOWERS, France

(3) Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy

(4) CITIC, Universidade da Coruña, Spain

(5) Massachusetts Institute of Technology, Cambridge, USA

(*) Equally contributing authors, the other authors are listed in alphabetical order

Correspondence to Olivier.Sigaud@isir.upmc.fr

ABSTRACT

A lot of recent machine learning research papers have “open-ended learning” in their title. But very few of them attempt to define what they mean when using the term. Even worse, when looking more closely there seems to be no consensus on what distinguishes open-ended learning from related concepts such as continual learning, lifelong learning or autotelic learning. In this paper, we contribute to fixing this situation. After illustrating the genealogy of the concept and more recent perspectives about what it truly means, we outline that open-ended learning is generally conceived as a composite notion encompassing a set of diverse properties. In contrast with previous approaches, we propose to isolate a key elementary property of open-ended processes, which is to produce elements from time to time (e.g., observations, options, reward functions, and goals), over an infinite horizon, that are considered novel from an observer’s perspective. From there, we build the notion of open-ended learning problems and focus in particular on the subset of open-ended goal-conditioned reinforcement learning problems in which agents can learn a growing repertoire of goal-driven skills. Finally, we highlight the work that remains to be performed to fill the gap between our elementary definition and the more involved notions of open-ended learning that developmental AI researchers may have in mind.

1 INTRODUCTION

Most existing software agents and robots suffer from insufficient versatility, limiting the potential introduction of these agents and robots in our everyday life Plappert et al. (2018). In most cases, some expertise is required from a human engineer to design their behavior in anticipation of the situations they may encounter. As a consequence, these agents cannot address novel or unforeseen situations.

An alternative to this specific expertise requirement would be to build an agent which would be ready to address *any* problem, fulfilling the long-standing dream of Artificial General Intelligence (AGI) Fjelland (2020). The corresponding design effort would be extraordinary, so there is a consensus that the way to reach AGI should be through learning. But the AGI perspective does not account for one of the core limitations of human intelligence: some humans can solve problems that others cannot, and vice versa, because humans have a limited lifespan for learning whereas being proficient in all potential tasks would require a potentially infinite amount of learning time.

Taking into account the above core limitation, open-ended learning (OEL) is a framework where autonomous agents face new tasks and learn how to solve them over their lifespan, driven by some intrinsic motivations or external guidance. Doing so, they build their own competence along some *developmental trajectory* which endows them with some capabilities, but not all capabilities that one may dream of Lungarella et al. (2003). Their developmental trajectory consists of a curriculum of tasks, and the curriculum learning challenge consists in finding when to learn which tasks so as to better extend the agent’s capabilities.

Though the OEL topic has been around for more than a decade (e.g., see the Intrinsically Motivated Open-ended Learning Workshops and Community¹ and Section 2), the interest for this framework is growing very fast in the machine learning community, as illustrated by more than 200 machine learning papers using the term in 2022, the emergence of new workshops (e.g. ALOE²) or even the very related COLLAS top-tier conference³. This growth is simultaneous with the emergence of related topics in Computational Neuroscience (e.g., see Niv (2019); Rmus et al. (2021)).

In this paper we start by showing that, despite a growing number of research works using the term “open-ended learning”, there have been very few attempts at providing a formal definition for the corresponding concept. This results in a lack of consensus on what OEL truly is and in a risk of confusion with several very related topics such as lifelong learning, continual learning or autotelic learning.

2 DEFINING OEL: ELEMENTS FROM THE LITERATURE

In this section, we draw from the literature some elements of a definition of OEL. We investigate the genealogy of the notion through its first occurrences in Artificial Intelligence (AI) and Artificial Life (AL) papers and we highlight the conceptual shift that arose in more recent papers. This investigation helps us set the stage for providing a better definition than the few existing ones in a subsequent section.

2.1 GENEALOGY OF THE NOTION

Searching for the very first instances of AI or AL works mentioning open-endedness is difficult because OEL is also a concept used in education sciences, with a far longer history. The first reference we could find where the term “open-ended” referred to intelligent machines within a cognitive science context is from Reader (1969):

“So far as is known, the range of tasks which the human intellect can master is open-ended (infinite), and therefore an intelligent machine can never be proved to be intelligent by comparing its task performance with that of the human intellect since the process would not terminate. [...] The only way a machine can continue to impersonate the open-ended ability of the human intellect is by also being open-ended. [...] The human intellect has an open-ended ability because it is capable of learning and the intelligent machine must also be capable of learning.”

This view for which an intelligent machine should learn forever as humans do is clearly reminiscent of Turing’s claim in his seminal paper that the process of making a machine intelligent should follow “the normal teaching of a child” Turing (1950).

Much later, an interesting early appearance of open-endedness in a pure AI paper is Meyer and Wilson (1991), a text from the founders of the “From animals to animats” community at the time where the corresponding conference emerged. They mention “an open-ended space of network architectures” in a paper dedicated to the artificial evolution of behaviours. The notion of “open-ended evolution” soon became the topic of several papers in the AL

¹IMOL - Intrinsically Motivated Open-ended Learning Workshops and Community: <https://www.imol-community.org/archive/>

²Agent Learning in Open-Endedness (ICLR 2022 Workshop): <https://sites.google.com/view/aloe2022>

³Conference on Lifelong Learning Agents: <https://lifelong-ml.cc/>

community, where the goal was to explain how evolution can generate increasingly complex and capable creatures in a context where most evolutionary algorithms were disappointingly converging to some unsatisfactory fixed-point or cyclic patterns. For instance, Standish (2003) mentions such discussions as central to the AL conferences in the early 2000’s, and open-ended evolution is still an active topic in the AL community, see for example Taylor et al. (2016); Packard et al. (2019). To extract elements of a definition from this line of research, we retain the definition from Standish (2003), where we find:

“The issue of open-ended evolution can be summed up by asking under what conditions will an evolutionary system continue to produce novel forms.”

Importantly, we see that the focus is on the conditions on the environment rather than on the evolution process itself (e.g., see also Soros and Stanley (2014), for similar conclusions) and that the expected outcome is the production of novelty.

One perspective on the conditions for open-ended evolution or OEL is thus to consider an rich enough environment which provides an open-ended sequence of problems. A similar idea of generating sequences of problems dates back to Schmidhuber (2013) where the author defines “Open-Ended PowerPlay”. The goal of the approach is the “Invention of new problems” Srivastava et al. (2012) to continually challenge a learning agent. This is much more closely related to what the machine learning community now means with OEL, but note that the focus is on generating problems in the environment rather than on the necessary mechanisms to solve them. This work is often cited in the machine learning works interested in the automated generation of curricula, and particularly in those which focus on generating a sequence of challenging problems in some environment rather than on solving them. For instance, this is the case of the POET algorithm Wang et al. (2019; 2020) where the open-ended invention of challenging environments is conceptualized as separated from the agent:

“How can progress in machine learning and reinforcement learning be automated to generate its own never-ending curriculum of challenges without human intervention? The recent emergence of quality diversity (QD) algorithms offers a glimpse of the potential for such continual open-ended invention. [...] The Paired Open-Ended Trailblazer (POET) algorithm introduced in this paper combines these principles to produce a practical approach to generating an endless progression of diverse and increasingly challenging environments while at the same time explicitly optimizing their solutions.”

This perspective about open-ended invention is now widely adopted in the machine learning community, and can be seen as one of the most prevalent understandings of what OEL truly means, for example see Dharna et al. (2020); Bontrager and Togelius (2021); Kepes et al. (2022); Dharna et al. (2022). This view has been recently strengthened and further broadened with the proposal of a theoretical perspective claiming that the current challenges of machine learning, including the production of general artificial intelligence, can be addressed shifting focus from learning algorithms to *generalized exploration*. This allows both reinforcement learning (RL) and supervised learning (SL) algorithms to automatically generated new data Jiang et al. (2023):

“Importantly, generalized exploration is a necessary objective for maintaining open-ended learning processes, which in continually learning to discover and solve new problems, provides a promising path to more general intelligence.”

Within this framework, the authors propose that promising approaches involve processes for the open-ended generation of novel data that maximize the agent’s learning potential while also leading it to explore areas of the problem space that are at the same time novel and grounded (i.e., relevant for the use of the system). This approach might for example involve the open-ended automatic generation of novel and relevant tasks in RL, for example based on parameterised environments, and novel and relevant new data in SL, for example based on the generation of synthetic training data.

A different perspective, pioneered by Weng et al. (2000; 2001), shifts the focus of the “openness” from the environment to the agent:

“What is autonomous mental development? With time, a brain-like natural or an artificial embodied system, under the control of its intrinsic developmental program (coded in the genes or artificially designed) develops mental capabilities through autonomous real-time interactions with its environments (including its own internal environment and components) by using its own sensors and effectors. Traditionally, a machine is not autonomous when it develops its skills, but a human is autonomous throughout its lifelong mental development. [...] A mental developmental process is also an open-ended cumulative process”

In Weng’s perspective, the autonomous progressive learning of the agent should be guided by an internal *developmental program*. This program is characterised by various elements that are relevant for OEL: *“Sensor-specific and effector-specific; task-nonspecific; tasks unknown at programming time; generate representation automatically; animal-like online learning; open-ended learning of more new tasks.”* This approach led to the birth of the new field of “Developmental Robotics” and the related “Development and Learning” interdisciplinary community and conference Lungarella et al. (2003).

In this context, other seminal works identified the key computational ingredients that allow the actual implementation of the developmental program, namely *intrinsic motivations* Barto et al. (2004); Singh et al. (2004). This is also the case of Prince et al. (2005) who introduces a notion of *ongoing emergence* which is explicitly related to open-ended learning and intrinsic motivations. These works managed to ignite a research effort directed to systematically draw concepts on different intrinsic motivations from psychology, and translate them into specific machine-learning algorithms Barto et al. (2004):

“Psychologists distinguish between extrinsic motivation, which means being moved to do something because of some specific rewarding outcome, and intrinsic motivation, which refers to being moved to do something because it is inherently enjoyable. Intrinsic motivation leads organisms to engage in exploration, play, and other behavior driven by curiosity in the absence of explicit reward. [...] Although these arguments are compelling, developmental approaches to artificial agent design have been slow to penetrate the mainstream of the machine learning community.”

Intrinsic motivations can thus support the possibly-open *autonomous* acquisition of knowledge and skills:

“According to this approach, an agent undergoes an extended developmental period during which collections of reusable skills are autonomously learned that will be useful for a wide range of later challenges.”

These works followed previous pioneering works on specific intrinsic motivation mechanisms which initially were overlooked Schmidhuber (1990; 1991). The renewed research effort led to connect educational science and developmental psychology concepts of OEL in children to the notion of intrinsic motivations in machine learning and robots Kaplan and Oudeyer (2007). In addition, it led to distinguish and formalise different classes of intrinsic motivations, in particular related to *prediction*, *novelty*, and *competence* Oudeyer and Kaplan (2007); Barto et al. (2013). Intrinsic motivations can thus form the ‘motivational engine’ guiding *autonomous* open-ended learning in organisms and robots Baldassarre and Mirolli (2013).

Another critical step for the definition of OEL agents was the investigation of the relation between intrinsic motivations and evolution, which we have seen to be two key areas where open-endedness can manifest. In general, evolution is proposed to lead to the emergence of both extrinsic motivations (those serving typical biological needs such as hunger and sex) and intrinsic motivations as both can guide the acquisition of behaviour adapted to the environmental conditions Schembri et al. (2007); Singh et al. (2010). In organisms, both types of motivations are supported by brain mechanisms responding to different general principles Baldassarre (2011). In particular, *extrinsic motivations* support the acquisition of material resources having a direct adaptive advantage, and to this purpose monitor visceral body states. Instead, intrinsic motivations, emerged later in evolution, support the acquisition of knowledge and skills which are only later useful to increase fitness, and thus monitor the

brain *information gain*. For this reason, only intrinsic motivations, if suitably translated into algorithms, have the “*potential to produce open-ended learning machines and robots*” Baldassarre (2011). This idea led to the start of the IMOL Workshop series mentioned above.

Aside from intrinsic motivations, research has highlighted that a second key element can foster OEL, namely *goals*. Goals are internal representations of desired states of the environment that can guide the agent’s action. Goals represent a key element in classic symbolic AI systems, for example to support planning Russell and Norvig (2016), but have been less employed in ML systems. A main exception is the RL option framework, where goals might be associated to the termination condition of options Sutton et al. (1999a), or *goal-conditioned policies*, where RL action policies are goal indexed Baldassarre (2001); Liu et al. (2022); Colas et al. (2022). Within this context, the potential relevance of goals for OEL was highlighted in Barto et al. (2004), where the learning of the agent was guided by the autonomous generation of options (hence goals) when a “salient event” was encountered. Later, self-generated goals have been indicated to be a fundamental means usable by OEL agents to autonomously generate curricula driven by competence-based intrinsic motivations Santucci et al. (2012; 2016). The EU-funded project GOAL-Robots⁴ pivoted on this idea to build *GOAL Agents*, where ‘GOAL’ stands for ‘Goal-based Open-ended Autonomous Learning’. The capacity of OEL agents to autonomously generate, discover or select goals has been recently called *autotelic* Colas et al. (2022), and plays an important role in the framework proposed here.

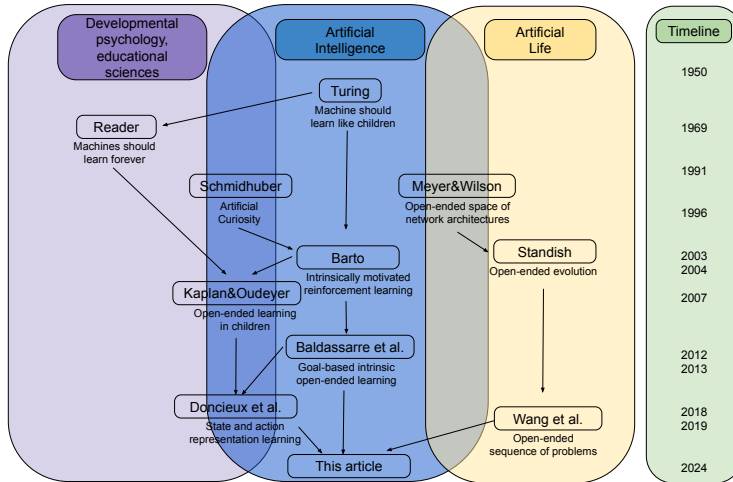


Figure 1: Schematic overview of the genealogy of the notion of open-endedness in the Developmental AI and Artificial Life literature.

We summarize the above overview of the genealogy of the notion of open-endedness in the AI/AL literature in Figure 1. From the historical investigation above, we can retain that the open-endedness property can be bound to the environment or to the agent. In the latter case, open-ended learning relies on intrinsic motivations able to drive the autonomous unbounded learning of increasingly complex and new knowledge and skills. In this case, goals have emerged as a key means to support open-ended learning.

2.2 CURRENT PERSPECTIVES

Though all the perspectives considered in the previous section still correspond to active research lines, we observe a renewal of the questions related to open-endedness in the more recent literature. For instance, though they do not define what they mean by OEL, Fan

⁴GOAL-Robots - Goal-based Open-ended learning Autonomous Robots: <https://www.goal-robots.eu/>

et al. (2022) have OEL environments, tasks, goals, and task suites, the main idea being to have a “wide variety” of such elements.

A relevant recent paper from the “Open-ended learning group” at DeepMind seems to adopt the above perspective. Though the term “open-ended” hardly appears apart from the title, introduction, and conclusion, we can find in Stooke et al. (2021) the following claim:

“We show that through constructing an open-ended learning process, which dynamically changes the training task distributions and training objectives such that the agent never stops learning, we achieve consistent learning of new behaviours.”

One can see that an OEL process should provide a goal distribution shift, or curriculum, from which the agent can continually learn to achieve new goals. As outlined in Section 4, this may be confused with continual or lifelong learning. The same is true of a more recent article from the same company Adaptive Agent Team (2023).

The work Doncieux et al. (2018) present another perspective that is relevant for our attempt to build a stronger definition of OEL. The authors define an *OEL process* as a process building appropriate MDPs depending on the current situation and the external reward function. This perspective has two major features. First, it is the only paper we found that attempts to provide an explicit definition of OEL. Second, it departs from all other existing frameworks by considering that each task should come with its own state and action spaces. This is in sharp contrast with all the multitask RL frameworks we know about, where a unique policy or set of policies with the same input-output format are used, implying that the state and action spaces are the same for all tasks. As outlined in a previous work tackling the issue of representations within the RL framework Konidaris (2019), having a necessary and sufficient representation specific to each task corresponds to the necessity of state and action *abstraction*: it is easier to solve a task if we abstract away all sensory information and potential actions that are not relevant for task achievement.

The same idea can be found in a recent trend in computational neuroscience research which focuses on the executive functions that help us determine an adequate representation of the task we want to solve Niv (2019); Rmus et al. (2021). In Doncieux et al. (2018), the capability to abstract away adequate state and action spaces for each task is based on a more general representational redescription capability inspired by Karmiloff-Smith (1994).

However, the definition of OEL provided in Doncieux et al. (2018) is not entirely satisfactory. A key point is that the framework considers that, when an agent addresses an open-ended sequence of tasks, each task comes with its own externally defined reward function. But if the tasks cannot be anticipated, the corresponding reward functions cannot be provided in advance by the agent designers. Moreover, and most importantly, although it is an interesting condition, it is in no way obvious why abstraction and representational redescription capabilities would be a necessary condition for exhibiting OEL capabilities. Finally, in Doncieux et al. (2018) goals and motivations are built from the reward. Instead, intrinsic motivations frameworks do the contrary by deriving goals from intrinsic motivations and rewards from intrinsic motivations and goal achievement Barto et al. (2004); Santucci et al. (2016); Colas et al. (2022).

This final remark can be generalized to many works cited above. Indeed, a key issue with previous attempts at defining OEL is that they all try to bind the concept to a large conjunction of properties, such as being autonomous, being endowed with intrinsic motivations or being capable of state abstraction. Relying on such conjunctions does not help pinpointing what is specific to OEL, which we believe is a key requirement for faster progress in designing agents capable of this property.

The work that best responds to this remark is Romero et al. (2023b), where the authors put forward the Lifelong Open-Ended Learning Autonomy (LOLA) framework. We think that by isolating OEL from lifelong learning and autonomous learning, the LOLA framework of Romero et al. (2023b) takes a step in the right direction. However, for the OEL component of their framework they refer to Doncieux et al. (2018) without providing a proper definition, though the referred paper does not provide a satisfactory OEL definition.

Thus the key contribution in Section 3 consists in providing such a core definition, on which we hope to build a solid OEL framework. Then in Section 4, we combine the just defined OEL property with other properties mentioned in the literature. In particular, we show how the property can be combined with lifelong learning and that it removes the need for the continual RL framework of Abel et al. (2023). In addition, we discuss the fact that an OEL agent might be autotelic or it could be teachable (that is, it receives goals from the environment). Finally, in Section 6 we discuss the limits of our proposal. We summarise all these elements and their relations in Figure 2, which expresses the relationships between the definitions we propose.

3 OPEN-ENDED LEARNING: A DEFINITION

Our investigation of the literature about the definition of OEL in AI and ALife has revealed that some authors were trying to define *OEL problems* whereas others were focusing on defining *OEL solutions*. The coexistence of both perspectives is not surprising as it is unclear whether the conditions for the emergence of OEL are more on the side of the environment – which defines the problem the agent should solve – or on the side of the agent – where a class of agents defines the solution to the OEL problem. In this paper, we want to define OEL problems in the most general way, so we consider both perspectives. To clearly distinguish definitions about the OEL problem from those about OEL solutions, we highlight the former in blue and the latter in red.

To introduce our framework, we first position it with respect to the definition of a continual reinforcement learning problem proposed in Abel et al. (2023). The authors formalize a class of problems where, to be optimal, an agent should never stop learning. In more details, and without strictly following their terminology, they define a learning agent as a trajectory in policy space generated by a history of interactions with the environment. In their framework, a set of agents corresponds to the set of policies that are generated by the corresponding interaction histories and learning algorithms. A triplet formed by an environment, a performance measure, and a set of such agents defines a continual RL problem if the best performing agents in that set never converge to a single policy. That is, in a continual RL problem the best agents must continually learn. Typical problems that require these types of solutions involve non-stationary scenarios presenting ever changing features to which the agents need to continuously adapt Kauvar et al. (2013); Romero et al. (2023a).

These definitions are helpful, but a few remarks need to be made. First, in the given definition a continual RL agent may switch between a finite set of policies forever without generating anything new. By contrast, in Section 3.1 we propose the idea that a key element of OEL is the never ending learning of new knowledge, building on a notion of open-ended process.

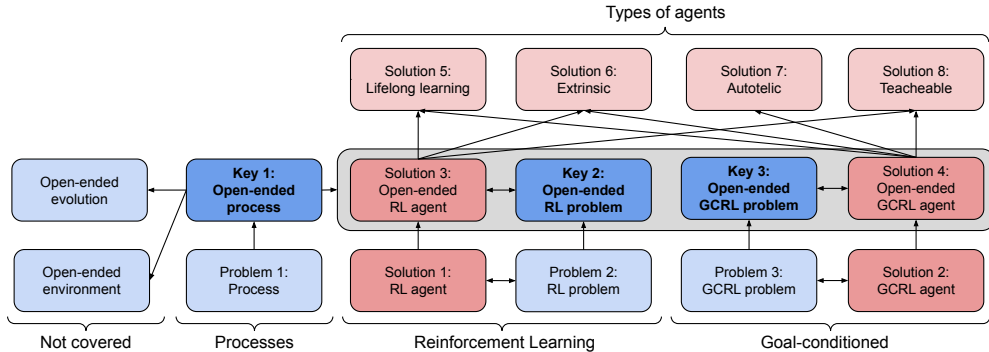


Figure 2: The set of definitions used in this paper. In blue, notions related to defining the open-ended goal-conditioned RL *problems* (key definitions are darker). In red, notions related to defining open-ended GCRL agents, that is *solutions*. An arrow expresses that a definition depends on another.

Second, the notion of continual learning given in Abel et al. (2023) does not imply the idea of a growing set of goals that an agent may face. To account for such an idea, in Section 3.3 we plug our notion of open-ended process into the goal-conditioned reinforcement learning (GCRL) framework, where the open-ended process plays the role of generating goals.

3.1 OPEN-ENDED PROCESS

The key building block of our proposal is a general notion of *open-ended process*. To arrive at this notion we first define a process.

Definition Pro1: A process is an entity that generates a set of tokens through time.

There is no constraint on what a token is and no time limit, and the same token may appear several times. Note that a process may generate several tokens at the same time.

To define an *open-ended process*, we introduce the idea that the process generates some novelty through time. But we would like to stress that novelty is a property relative to an observer: some observers will consider a token as new if it is different enough from all previous tokens according to their own perspective, but other observers may consider that the same token is too similar to previous tokens to be considered distinct. The notion of observer proposed here is broad as it might be a non-transparent subjective evaluator (e.g., a human user) or alternatively an objective mechanism that incorporates some measure of novelty (e.g., an algorithm). As a consequence, our key definition is the following:

Key definition 1: An observer considers a process as open-ended if, for any time t , there exists a time $t' > t$ at which the process generates a token that is new according to this observer’s perspective.

Differently from the definition of continual learning of Abel et al. (2023), this definition captures the idea that something new must happen at least from time to time. Moreover, the definition implies that an open-ended process works over an infinite time horizon.

A property of open-endedness defined this way is that any open-ended process necessarily generates an infinite number of distinct tokens. As a simple mathematical proof, if $|\text{Tok}_t|$ denotes the number of distinct tokens generated up to time t , by definition there exists a time $t' > t$ such that $|\text{Tok}_{t'}| \geq |\text{Tok}_t| + 1$. This property can be applied an arbitrary number of times, so that for any $N = 1, 2, \dots$ there exists a $t' > t$ such that $|\text{Tok}_{t'}| \geq |\text{Tok}_t| + N$, proving that the number of distinct tokens generated throughout the process is unbounded.

The above definition is very general and can capture many examples of open-endedness. For instance, in open-ended evolution, the tokens are agent phenotypes. In standard RL (i.e., non goal-conditioned), an OEL agent is an agent whose behavior generation process is open-ended. There might be several ways to implement this condition within our definition, for example the tokens might be reward functions, Markov decision processes, policies or trajectories. We now first give a definition of such an OEL framework based on standard RL, and then we focus on GCRL and open-ended goal generation.

3.2 OPEN-ENDED RL

To define RL problems we borrow the formalism of the problem definitions from Abel et al. (2023) that we introduced in Section 3. This formalism is used here because it has a broader scope than the more frequently used Markov Decision Processes. However, unlike in Abel et al. (2023) where the distinction between problems and solutions is not considered, our definitions insist on this distinction. As in Abel et al. (2023), we define an agent-environment “interface” as the $(\mathbf{O} \times \mathbf{A})$ pair where \mathbf{O} is the space of observations and \mathbf{A} is the space of actions resulting from interactions of this agent with this environment. We then define the set of possible histories of the agent-environment interactions as $\mathbf{H} = \bigcup_{t=0}^{\infty} (\mathbf{O} \times \mathbf{A})^t$. A single history of such a set, $h_t \in \mathbf{H}_t \subset \mathbf{H}$, involves a sequence of length t of observation-action pairs.

From these definitions, we define an RL problem as follows.

Definition Pro2: An RL problem is a tuple $\langle \mathbf{O}, \mathbf{A}, \mathbf{T}, \mathbf{R} \rangle$ where \mathbf{O} is the set of the agent’s observations, \mathbf{A} is the set of the agent’s primitive actions, $\mathbf{T} : \mathbf{H}_t \rightarrow \Delta(\mathbf{O})$ is a transition function mapping the history up to time t to a probability distribution of observations (where $\Delta(\mathbf{O})$ is the probability simplex of \mathbf{O}), and $\mathbf{R} : \mathbf{H}_t \rightarrow \mathbb{R}$ is a reward function.

To build an open-ended RL problem from the RL problem as defined above, we need to consider at least a sequence of such RL problems. Such problems can be organized purely sequentially, but also hierarchically, when an agent is building complex problem representations on top of simpler representations. In the case of a hierarchical organization, one may call upon the options framework Sutton et al. (1999b).

Under this hierarchical perspective, the tokens generated by an open-ended process could be RL problems themselves, that is an open-ended RL problem might generate an open-ended sequence of RL subproblems. Based on these elements, we can give the following general and potentially recursive definition of an open-ended RL problem.

Key definition 2: An open-ended RL problem is a sequence of RL problems in which any of the following defining elements is generated by an open-ended process: the observation space, the action space or some hierarchical options, the transition function, the reward function or RL problems.

3.3 OPEN-ENDED GCRL

From these definitions, we proceed to define GCRL problems. That is, we introduce into RL problems a notion of goal and goal-conditioned reward function (see Colas et al. (2022)). We characterize goals as elements of an arbitrary goal space \mathbf{G} but we do not specify where they come from. For example, they could be generated by the environment, another agent or the agent itself.

Definition Pro3: A GCRL problem is a tuple $\langle \mathbf{O}, \mathbf{A}, \mathbf{T}, \mathbf{G}, \mathbf{R} \rangle$, where \mathbf{O} is the set of the agent’s observations, \mathbf{A} is the set of the agent’s actions, $\mathbf{T} : \mathbf{H}_t \rightarrow \Delta(\mathbf{O})$ is a transition function mapping the history up to time t to a probability distribution of observations, \mathbf{G} is the set of goals, and $\mathbf{R} : \mathbf{H}_t \times \mathbf{G} \rightarrow \mathbb{R}$ is the goal-conditioned reward function.

Similarly, using goals we can apply our general definition of an open-ended process to a goal generation process, and get the following definition:

Key definition 3: A GCRL problem is an open-ended GCRL problem if the goals come from an open-ended generation process.

As noted above, for goal-generation processes whose set of distinct goals for the observer is finite, the problem cannot be open-ended. However, in practice agents generally have a finite lifespan. Thus a process generating a finite set of goals might be seen as giving rise to an open-ended goal generation process even if the infinite horizon condition is not met. We come back to this issue in Section 5.

The above definitions are very general, and capture a large set of OEL problems, among which many are not satisfactory models of developmental processes. To go further towards modelling open-ended learning from a developmental perspective, we must work on distinguishing trivial cases of open-ended GCRL problems from more interesting ones.

3.4 FIRST-ORDER AND SECOND-ORDER OPEN-ENDED GCRL

Given the above definitions, there are two simplistic cases of open-ended GCRL problems.

Case 1: The goal space is discrete, but infinite. A simple example of this case is when the goal is to count from 1 to N , where N is a natural number. Generating as goals an infinite sequence of growing values of N is enough for that problem to be open-ended.

Case 2: The goal space is continuous. A simple example of this case is asking a physical agent to travel at a given speed, where the speed is a real number. Targeting higher and higher traveling speeds that converge asymptotically to the highest possible traveling speed of the agent is a problem that already verifies the open-endedness property.

These two simplistic examples qualify as open-ended GCRL problems for naive observers who do not require a strong notion of novelty, but they do not account for the developmental perspective AI authors generally have in mind when they try to define OEL agents. We see two approaches to address this issue.

A first approach to avoid trivial OEL problems consists in putting constraints on what an observer may consider as a “new token” or not. Demanding observers may require new tokens to be “substantially new” or “interestingly new”.

A second approach, which is not orthogonal to the previous one, consists in considering that a developmental process implies that goal tokens are sampled from more and more interesting *spaces*. With this second approach, we may consider two classes of problems, following a strategy inspired from the work of Etcheverry et al. (2021). The first class involves problems as the usually considered ones:

Definition Pro4: A first-order open-ended GCRL problem is a GCRL problem where all goals \mathbf{g}_i are sampled from the same goal space \mathbf{G} .

By contrast, to define the second class, we first need to define a variety of goal spaces $\Omega_j \in \Omega$ where Ω is the set of all possible goal spaces. Then we can introduce the idea of goals generated from different spaces.

Definition Pro5: A second-order open-ended GCRL problem is a GCRL problem where a first open-ended process generates an open-ended sequence of goal representation spaces $\Omega_j \subset \Omega$, and a second open-ended process generates goals $\mathbf{g}_{i,j}$ within each generated space.

Below for simplicity we consider the case of one goal space, and hence \mathbf{G} instead of \mathbf{G}_j , but our definitions could be extended to the case of the generation of multiple goal spaces.

A different way to express the same thing is that in first-order GCRL problems tokens of the open-ended process are goals from a single goal space, whereas in the second-order case tokens of the open-ended processes are goal spaces and goals.

Note that with the above definition, second-order open-ended GCRL problems can still generate a trivial diversity of goal spaces, so again we have to add further conditions so that the generated goals and goal spaces ensure an interesting developmental trajectory.

4 COMBINING OPEN-ENDED LEARNING WITH OTHER PROPERTIES

We have proposed a definition for the elementary property of an OEL problem. Now it is time to ask whether and how we can combine this property with other, complementary ones. In particular, we would like to build an equivalent of the LOLA framework of Romero et al. (2023b), where LOLA stands for Lifelong Open-Ended Learning Autonomy. Below we do so by first building in Section 4.1 on the fully-formalized notion of continual learning proposed in Abel et al. (2023). Then we stress the need for accounting for the lifelong learning property. Finally, we also expand in Section 4.3 on the fact that the autonomous

learning property is too strict a requirement, as one may consider agents that are both autotelic and teachable, i.e. capable of adopting social partner’s goals.

4.1 GOAL-CONDITIONED CONTINUAL REINFORCEMENT LEARNING

Given our definitions above, we can now connect our definition of OEL to the definition of continual learning from Abel et al. (2023), investigating solutions to the continual and open-ended GCRL problem. An RL agent can then be defined as follows:

Definition Sol1: An RL agent is a function $\lambda : \mathbf{H}_t \rightarrow \Delta(\mathbf{A})$ mapping a history of observations and actions up to a given time t , to an action selection probability distribution.

An goal-conditioned agent can then be defined as follows:

Definition Sol2: A GCRL agent is a function $\kappa : \mathbf{H}_t \times \mathbf{G} \rightarrow \Delta(\mathbf{A})$ mapping a history of observations and actions up to a given time t , and the goal selected at that time, to the action selection probability distribution.

As classically done in the GCRL literature Campos et al. (2020), we distinguish ‘behavioral goals’, which are used to condition a policy, and ‘achieved goals’, which are the goals fulfilled by the behavior generated by the policy.

In Abel et al. (2023) the authors define ‘continual problems’ as problems where the optimal agents never stop learning (i.e., they continuously change their policies). A weakness of this definition is that the continual learning RL problem is defined in terms of the type of solutions (agents) to use to solve them. From that perspective, the examples given by the authors in their work all involve non-stationary environments, and thus the non-stationary property should actually be used to define their problems. Given this feature of such environments, it should follow that the optimal agents (solutions) solving them have to keep changing their policies. Furthermore, one should also consider that if the environment continues to change, but after some time it repeats the same configurations, then the best agents might actually learn all the solutions for all environment configurations, and then stop learning. This unless it is assumed that the agents forget previous solutions when the environment changes, an assumption that the authors seem to implicitly make. Only problems that continue to propose novel challenges to the agents, as the true OEL problems considered here, actually require agents that keep changing: these agents have to keep changing not because they forget previous solutions, but because they need to indefinitely acquire new knowledge. This is a signature feature of OEL agents and also of any solution to truly OEL problems.

We can define open-ended RL agents (on the side of solutions) as follows:

Definition Sol3: An open-ended RL agent is an RL agent that solves an open-ended RL problem.

This definition is admittedly very general. We can be more specific when defining open-ended GCRL agents:

Definition Sol4: An open-ended GCRL agent is an GCRL agent that uses goals from an open-ended goal generation process.

From this definition, it should be obvious that an open-ended GCRL agent defined as above is the solution to a continual RL problem as defined by Abel et al. (2023). Indeed, to be optimal, these agents should continuously learn in order to solve the continuously generated novel goals.

By contrast, if we remove the condition on OEL and if the environment is stationary, we may account for convergence to an universally efficient agent that has learned enough to become capable, after some time, to achieve any forthcoming goal in a zero-shot manner.

Given the considerations above, we would like the agents not to (completely) forget the tasks it previously achieved when it addresses new ones. This property corresponds to lifelong learning. Thus we should also include the lifelong learning property into our framework.

4.2 LIFELONG OPEN-ENDED LEARNING

Though the names suggest they are similar, the lifelong learning property and the continual learning property are not exactly the same. Indeed, the authors who use the first term generally have in mind the capability to overcome catastrophic forgetting (e.g., see Parisi et al. (2019)) whereas the authors of the continual learning definition mentioned above do not even mention this aspect Abel et al. (2023).

The lifelong learning property captures the idea of a growing repertoire of skills, although it mainly focuses on avoiding to forget the previous skills when learning new ones. Thus, characterizing such a property requires being more specific on the learning processes from the side of the agent, hence the definition must be from the side of solutions.

Definition Sol5: A lifelong open-ended RL or GCRL agent is an open-ended RL or GCRL agent which does not forget how to solve RL problems or achieve goals when it learns how to solve other RL problems or achieves other goals.

4.3 THE ORIGIN OF GOALS: EXTRINSIC, AUTOTELIC, AND TEACHABLE OPEN-ENDED LEARNING AGENTS

So far, we have not specified where the goals come from. In this respect, there are three main possibilities.

First, the goals may come from the environment itself. In this case we say they are extrinsic. This corresponds to the `GoalEnv` category of environments in the standard `gym` interface for RL agents. In this case, the environment provides the agent with a state, a goal, and a reward. Extrinsic OEL agents are thus as follows.

Definition Sol6: An open-ended RL or GCRL agent is extrinsic if the goal generation process is external to the agent.

An important case of extrinsic OEL problem is one where the environment generates an open-ended sequence of goals that maximise the speed and quality of the learning process of the agents Jiang et al. (2023). Extrinsic OEL agents, however, are not the main focus here.

The second case corresponds to agents who set and learn to pursue their own tasks or goals autonomously. In this case we say the goals generated by the agents are intrinsic. An interesting case of this is when the agents autonomously set their own goals, in which case we have autotelic agents.

Definition Sol7: An open-ended GCRL agent is autotelic if its goal generation process is fully internal to the agent (autonomous).

To discover new goals, autotelic OEL agents generally expand the set of behavioral goals they consider by sampling from the set of goals they have already achieved Campos et al. (2020), which defines a curriculum over goals. More precisely, the most efficient curricula consist in sampling at the frontier of the domain of currently achieved goals Seepanomwan et al. (2017); Pitis et al. (2020); Castanet et al. (2022). This interaction should ensure that the agent achieves enough new goals to continue expanding its behavioral goal domain.

From a slightly different perspective, for an autotelic agent facing a second-order open-ended GCRL problem where it generates goal spaces from its own experience, the environment has to be rich enough to foster this open-ended goal space generation process.

A third case involves agents that autonomously generate goals but whose goal generator can be influenced by interactions with social partners. By giving demonstrations, descriptions, instructions or feedback, social partners can suggest to these agents to adopt some goals rather than others. Following Sigaud et al. (2023), focusing on autotelic agents, we say that such agents are teachable.

Definition Sol8: An OEL GCRL agent is teachable if its internal goal generation process can be influenced by other agents.

It should be clear that ensuring the OEL property is more difficult for purely autonomous agents than for teachable agents, as the former must discover new goals or goal spaces without external support, whereas the latter can benefit from social interactions to do so. That is, being teachable may relax lot of the requirements on the goal discovery capabilities of agents.

Note that some agents can be teachable without autonomously generating goals, meaning that they fully depend on external agents for goal generation. Note also that autonomously generating goals is not a sufficient condition for being an OEL agent, as an autonomous agent may always sample from the same limited set of goals.

Also, it is not a sufficient condition to be a continual learner according to the definition of Abel et al. (2023), as the behavior may become stationary.

This line of thoughts suggests studying in the future the different properties of the types of agents developmental AI researchers have in mind, and the way they interact with each other.

5 EVALUATING OPEN-ENDED LEARNING AGENTS

The OEL framework poses two important challenges: (1) How can we make sure that a learning agent is an OEL agent? (2) How can we compare the performance of two OEL agents?

5.1 ASSESSING THE OPEN-ENDED LEARNING PROPERTY

A key difficulty with OEL agents is that, based on the definition given here, it is not easy to evaluate in practice if an agent is showing open-ended learning or not. Indeed, such an agent should address new goals forever, so we should wait forever to determine if it succeeds in doing so. Of course, this is not doable in practice. A related viewpoint is that the definition does not take into account the fact that concrete agents generally have a finite lifespan.

A first approach to address this problem was proposed by Cartoni and Baldassarre (2018), where the authors attempt to define open-ended learning problems in a concrete way. The agent’s “life” is divided into two phases. In a first, very long “intrinsic phase”, the agent is set in a given environment and receives no learning guidance (e.g., reward functions or goals) nor any hardwired knowledge. In a second “extrinsic phase”, the agent is tested with a set of tasks (e.g., goals to achieve) that are “randomly drawn” from the same environment, in the sense that these tasks are *representative of all possible tasks* that might be generated in that environment. The key idea is that, since in the intrinsic phase the agent does not know the tasks it will have to solve in the extrinsic phase, it will have to possibly rely on intrinsic motivations to maximise its knowledge and skill acquisition and thus have a high performance with tasks in the extrinsic phase. Importantly, this performance represents a measure of the knowledge-gain capacity of the OEL processes in the intrinsic phase. This approach is definitely practical, but we have to be aware that it does not truly evaluate the OEL property in itself.

Another possible approach is more closely related to the definition of OEL given here. It consists in measuring the rate of novel goals that the agent discovers and/or learns to master throughout its life, and then extrapolating from its dynamics if the agent would continue to discover new goals forever if given infinite time. In particular, if the number of seen goals asymptotically converges to a constant, the agent is not an OEL agent. Otherwise, if the curve shows a logarithmic behavior, a linear behavior with positive slope, or even better an increasing slope, then the agent is an OEL agent. The practical application of this approach requires the possibility of marking goals as novel in order to compute the goal-generation rate, an operation which would pose the problems previously discussed.

5.2 COMPARING OPEN-ENDED LEARNING AGENTS

Besides, comparing the goal-reaching capabilities of purely autotelic agents is particularly difficult since different agents may follow a completely different curriculum, thus there might be no common ground on which to compare them. However, the method proposed by Cartoni and Baldassarre (2018) was designed to also address this problem. Indeed, the tasks used in the extrinsic phase to measure the agent performance should be representative and cover the whole space of problems that might be generated in the environment. Thus, the average performance of different agents on these tasks should be representative of the effectiveness of their OEL processes independently of their specific curricula which generated specific task-trajectories across the task space.

In the case of teachable agents, the problem can be mitigated as one may determine a common social interaction policy for an interacting social partner and measure to which extent various teachable agents manage to learn the goals the social partner is trying to suggest.

6 DISCUSSION AND CONCLUSION

In this paper we have investigated the genealogy of the notion of open-ended learning and isolated a core property that may be general enough to account for a large number of open-ended learning phenomena. From there, we have focused on a goal-conditioned reinforcement learning perspective, and proposed a framework from which open-ended GCRL problems and agents could be defined.

In addition, we have outlined the need for combining our core property with other properties that would help developmental IA authors to capture the idea of a growing repertoire of capabilities or skills, which is so important in the OEL process of children.

As a consequence of our choice, the main limitation of our work is that our definition of OEL problems does not imply any form of performance progress from the side of the agent. An intuition is that the agent should maximize some performance measure on the behavioral goals it has already seen so far. Thus, to go further, we should characterize a goal discovery process, explain how an agent may discover new goal spaces Pong et al. (2019), maximize its competence in these goal spaces Santucci et al. (2016), introduce representational redescription processes Doncieux et al. (2018), abstraction capabilities Konidaris (2019); Shanahan and Mitchell (2022), and even creativity Boden (1998). Finding adequate performance measures for all these further capabilities is left for future work, and trying to integrate all these capabilities into a common framework may reveal some deeper limitations of the definitions we have proposed here.

ACKNOWLEDGEMENTS

This work has received funding from the European Commission’s Horizon Europe Framework Program under grant agreement *N*° 101070381 (PILLAR-robots project).

REFERENCES

- Abel, D., Barreto, A., Van Roy, B., Precup, D., van Hasselt, H., and Singh, S. (2023). A definition of continual reinforcement learning. *arXiv preprint arXiv:2307.11046*.
- Adaptive Agent Team, D. (2023). Human-timescale adaptation in an open-ended task space. *arXiv preprint arXiv:2301.07608*.
- Baldassarre, G. (2001). A planning modular neural-network robot for asynchronous multi-goal navigation tasks. In *Proceedings of the 2001 Fourth European Workshop on Advanced Mobile Robots (EUROBOT2001)*, pages 223–230. Lund, Sweden, 19-21 September 2001.
- Baldassarre, G. (2011). What are intrinsic motivations? a biological perspective. In *Proceedings of the International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob-2011)*, pages E1–8. Frankfurt am Main, Germany, 24–27/08/11.
- Baldassarre, G. and Mirolli, M., editors (2013). *Intrinsically motivated learning in natural and artificial systems*. Springer-Verlag, Berlin.
- Barto, A., Mirolli, M., and Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, 4(907):E1–15.
- Barto, A. G., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In Triesch, J. and Jebara, T., editors, *International Conference on Developmental Learning (ICDL2004)*, pages 112–119. UCSD Institute for Neural Computation, IEEE. LaJolla, CA.
- Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial intelligence*, 103(1-2):347–356.
- Bontrager, P. and Togelius, J. (2021). Learning to generate levels from nothing. In *2021 IEEE Conference on Games (CoG)*, pages 1–8. IEEE.
- Campos, V., Trott, A., Xiong, C., Socher, R., Giró-i Nieto, X., and Torres, J. (2020). Explore, discover and learn: Unsupervised discovery of state-covering skills. In *International Conference on Machine Learning*, pages 1317–1327. PMLR.
- Cartoni, E. and Baldassarre, G. (2018). Autonomous discovery of the goal space to learn a parameterized skill. *arXiv preprint arXiv:1805.07547*.
- Castanet, N., Lamprier, S., and Sigaud, O. (2022). Stein variational goal generation for reinforcement learning in hard exploration problems. *arXiv preprint arXiv:2206.06719*.
- Colas, C., Karch, T., Sigaud, O., and Oudeyer, P.-Y. (2022). Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research*, 74:1159–1199.
- Dharna, A., Summers, C., Dasari, R., Togelius, J., and Hoover, A. K. (2022). Watts: Infrastructure for open-ended learning. *arXiv preprint arXiv:2204.13250*.
- Dharna, A., Togelius, J., and Soros, L. B. (2020). Co-generation of game levels and game-playing agents. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 16, pages 203–209.
- Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., Roijers, D. M., Girard, B., Perrin, N., and Sigaud, O. (2018). Open-ended learning: a conceptual framework based on representational redescription. *Frontiers in Robotics and AI*, 12.
- Etcheverry, M., Chan, B. W.-C., Moulin-Frier, C., and Oudeyer, P.-Y. (2021). Meta-diversity search in complex systems, a recipe for artificial open-endedness?
- Fan, L., Wang, G., Jiang, Y., Mandlkar, A., Yang, Y., Zhu, H., Tang, A., Huang, D.-A., Zhu, Y., and Anandkumar, A. (2022). Minedojo: Building open-ended embodied agents with internet-scale knowledge. *arXiv preprint arXiv:2206.08853*.

-
- Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7(1):1–9.
- Jiang, M., Rocktäschel, T., and Grefenstette, E. (2023). General intelligence requires rethinking exploration. *Royal Society Open Science*, 10(6):230539.
- Kaplan, F. and Oudeyer, P.-Y. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in neuroscience*, page 17.
- Karmiloff-Smith, A. (1994). Beyond modularity: A developmental perspective on cognitive science. *European journal of disorders of communication*, 29(1):95–105.
- Kauvar, I., Doyle, C., Zhou, L., and Haber, N. (2013). Curious replay for model-based adaptation. *arXiv preprint arXiv:2306.15934*.
- Kepes, M., Guttenberg, N., and Soros, L. (2022). Chemgrid: An open-ended benchmark domain for an open-ended learner. In *2022 IEEE Conference on Games (CoG)*, pages 221–228. IEEE.
- Konidaris, G. (2019). On the necessity of abstraction. *Current opinion in behavioral sciences*, 29:1–7.
- Liu, M., Zhu, M., and Zhang, W. (2022). Goal-conditioned reinforcement learning: Problems and solutions. *arXiv preprint arXiv:2201.08299*.
- Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, 15(4):151–190.
- Meyer, J.-A. and Wilson, S. W. (1991). The artificial evolution of behaviour. In *First international conference on the Simulation of Adaptive Behaviour (SAB)*. MIT Press.
- Niv, Y. (2019). Learning task-state representations. *Nature neuroscience*, 22(10):1544–1553.
- Oudeyer, P.-Y. and Kaplan, F. (2007). What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurobotics*, 1(6):E1–13.
- Packard, N., Bedau, M. A., Channon, A., Ikegami, T., Rasmussen, S., Stanley, K. O., and Taylor, T. (2019). An overview of open-ended evolution: Editorial introduction to the open-ended evolution ii special issue. *Artificial life*, 25(2):93–103.
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71.
- Pitis, S., Chan, H., Zhao, S., Stadie, B., and Ba, J. (2020). Maximum entropy gain exploration for long horizon multi-goal reinforcement learning. In *International Conference on Machine Learning*, pages 7750–7761. PMLR.
- Plappert, M., Andrychowicz, M., Ray, A., McGrew, B., Baker, B., Powell, G., Schneider, J., Tobin, J., Chociej, M., Welinder, P., et al. (2018). Multi-goal reinforcement learning: Challenging robotics environments and request for research. *arXiv preprint arXiv:1802.09464*.
- Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., and Levine, S. (2019). Skew-fit: State-covering self-supervised reinforcement learning. *arXiv preprint arXiv:1903.03698*.
- Prince, C., Helder, N., and Hollich, G. (2005). Ongoing emergence: A core concept in epigenetic robotics. In *Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling cognitive development in robotic systems, Nara, Japan. ed. L. Berthouze, F. Kaplan, H. Kozima, H. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov, & C. Balkenius*, page 6370.
- Reader, A. V. (1969). Steps towards genuine artificial intelligence. *Acta Psychologica*, 29:279–289.

-
- Rmus, M., McDougle, S. D., and Collins, A. G. (2021). The role of executive function in shaping reinforcement learning. *Current Opinion in Behavioral Sciences*, 38:66–73.
- Romero, A., Baldassarre, G., Duro, R. J., and Santucci, V.-G. (2023a). Learning multiple tasks with non-stationary interdependencies in autonomous robots. In *In Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 2547–2549.
- Romero, A., Bellas, F., and Duro, R. J. (2023b). A perspective on lifelong open-ended learning autonomy for robotics through cognitive architectures. *Sensors*, 23(3):1611.
- Russell, S. J. and Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson Education, Harlow, UK, third edition.
- Santucci, V.-G., Baldassarre, G., and Mirolli, M. (2012). Intrinsic motivation mechanisms for competence acquisition. In *Proceeding of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob 2012)*, pages 1–6. IEEE. San Diego, CA, USA, 7-9 November 2012.
- Santucci, V. G., Baldassarre, G., and Mirolli, M. (2016). Grail: A goal-discovering robotic architecture for intrinsically-motivated learning. *IEEE Transactions on Cognitive and Developmental Systems*, 8(3):214–231.
- Schembri, M., Mirolli, M., and Baldassarre, G. (2007). Evolution and learning in an intrinsically motivated reinforcement learning robot. In *Proceedings of the 9th European Conference on Artificial Life (ECAL2007)*, pages 294–303. Springer Verlag, Berlin. Lisbon, Portugal, September 2007.
- Schmidhuber, J. (1990). Making the world differentiable: on using fully recurrent self-supervised neural networks for dynamic reinforcement learning and planning in non-stationary environments. Technical Report FKI-126-90, Technische Universität München, München.
- Schmidhuber, J. (1991). Curious model-building control systems. In *Proceedings of the International Joint Conference on Artificial Neural Networks*, volume 2, pages 1458–1463. IEEE. Piscataway, NJ.
- Schmidhuber, J. (2013). Powerplay: Training an increasingly general problem solver by continually searching for the simplest still unsolvable problem. *Frontiers in psychology*, 4:313.
- Seepanomwan, K., Santucci, V. G., and Baldassarre, G. (2017). Intrinsically motivated discovered outcomes boost user’s goals achievement in a humanoid robot. In *2017 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 178–183. IEEE.
- Shanahan, M. and Mitchell, M. (2022). Abstraction for deep reinforcement learning. *arXiv preprint arXiv:2202.05839*.
- Sigaud, O., Caselles-Dupré, H., Colas, C., Akakzia, A., Oudeyer, P.-Y., and Chetouani, M. (2023). Towards teachable autonomous agents. *IEEE Transactions in Cognitive and Developmental Systems*.
- Singh, S., Lewis, R. L., Barto, A. G., and Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, 2(2):70–82.
- Singh, S. P., Barto, A. G., and Chentanez, N. (2004). Intrinsically motivated reinforcement learning. In *Advances in neural information processing systems (NIPS-2004)*, pages 1281–1288.

-
- Soros, L. B. and Stanley, K. O. (2014). Identifying necessary conditions for open-ended evolution through the artificial life world of chromaria. In *ALIFE 14: The Fourteenth International Conference on the Synthesis and Simulation of Living Systems*, pages 793–800. MIT Press.
- Srivastava, R. K., Steunebrink, B. R., Stollenga, M., and Schmidhuber, J. (2012). Continually adding self-invented problems to the repertoire: first experiments with powerplay. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, pages 1–6. IEEE.
- Standish, R. K. (2003). Open-ended artificial evolution. *International Journal of Computational Intelligence and Applications*, 3(02):167–175.
- Stooke, A., Mahajan, A., Barros, C., Deck, C., Bauer, J., Sygnowski, J., Trebacz, M., Jaderberg, M., Mathieu, M., et al. (2021). Open-ended learning leads to generally capable agents. *arXiv preprint arXiv:2107.12808*.
- Sutton, R. S., Precup, D., and Singh, S. (1999a). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211.
- Sutton, R. S., Precup, D., and Singh, S. (1999b). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211.
- Taylor, T., Bedau, M., Channon, A., Ackley, D., Banzhaf, W., Beslon, G., Dolson, E., Froese, T., Hickinbotham, S., Ikegami, T., et al. (2016). Open-ended evolution: Perspectives from the oee workshop in york. *Artificial life*, 22(3):408–423.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- Wang, R., Lehman, J., Clune, J., and Stanley, K. O. (2019). POET: open-ended coevolution of environments and their optimized solutions. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 142–151.
- Wang, R., Lehman, J., Rawal, A., Zhi, J., Li, Y., Clune, J., and Stanley, K. (2020). Enhanced poet: Open-ended reinforcement learning through unbounded invention of learning challenges and their solutions. In *International Conference on Machine Learning*, pages 9940–9951. PMLR.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2000). Computational autonomous mental development: A white paper for suggesting a new initiative. Technical report, Michigan State University, Department of Computer Science/Engineering.
- Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291(5504):599–600.