

Curiosity-driven exploration: Diversity of mechanisms and functions

Alexandr Ten, Pierre-Yves Oudeyer, Clément Moulin-Frier

▶ **To cite this version:**

Alexandr Ten, Pierre-Yves Oudeyer, Clément Moulin-Frier. Curiosity-driven exploration: Diversity of mechanisms and functions. The Drive for Knowledge: The Science of Human Information Seeking, , 2022, 10.1017/9781009026949 . hal-03447896v2

HAL Id: hal-03447896
https://inria.hal.science/hal-03447896v2

Submitted on 25 Nov 2021

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.

Curiosity-driven exploration: Diversity of mechanisms and functions

Alexandr TEN, Pierre-Yves OUDEYER, and Clément MOULIN-FRIER

Inria / University of Bordeaux / Ensta ParisTech, France

Abstract

Intrinsically motivated information-seeking, also called curiosity-driven exploration, is widely believed to be a key ingredient for autonomous learning in the real world. Such forms of spontaneous exploration have been studied in multiple independent lines of computational research, producing a diverse range of algorithmic models that capture different aspects of these processes. These algorithms resolve some of the limitations of neurocognitive theories by formally describing computational functions and algorithmic implementations of intrinsically motivated learning. Moreover, they reveal a high diversity of effective forms of intrinsically motivated information-seeking that can be characterized along different mechanistic and functional dimensions. This chapter aims at reviewing different classes of algorithms and highlighting several important dimensions of variation among them. Identifying these dimensions provides means for structuring a comprehensive taxonomy of approaches. We believe this exercise to be useful in working towards a general computational account of information-seeking. Such an account should facilitate the proposition of new hypotheses about information-seeking in humans and complement the existing psychological theory of curiosity.

1 INTRODUCTION

If you are reading this chapter, chances are, you are seeking information; you might even be curious about what you are about to read. Whatever the case, you are certainly familiar – as much as the next person – with what being curious feels like. It is this paradoxically palpable gap you perceive in your knowledge that makes you long for information.

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

Precisely how the curiosity-inducing mechanism operates in biological organisms is poorly understood (Gottlieb and Oudeyer 2018), but the undeniable significance of curiosity for the eventual acquisition of useful knowledge continues to spur researchers in artificial intelligence (AI) to develop curious learning algorithms. These efforts have generated a great diversity of mechanisms enabling curiosity-driven learning in artificial agents. This chapter identifies common

structural dimensions of these mechanisms and relates variation along these dimensions to possible evolutionary functions. To be able to discuss these ideas, we need to recount how machines learn.

Although defining learning can be controversial (Barron et al. 2015), we will adopt a straightforward formulation from machine learning (ML; Jordan and Mitchell 2015), where learning is defined as improvement on a task (or a set of tasks) with experience. An artificial agent is said to have improved on a task if – according to some well-defined criterion – it is able to perform the task better than it did prior to receiving learning experience. Thus, what drives improvement is a combination of the experience that comes in the form of data that the agent can represent and the learning algorithm (also called learning, or update rule) that specifies how the agent's innards change by processing the incoming data.

For a long time, major ML paradigms have been elaborating increasingly efficient and powerful learning algorithms that optimize pre-specified task criteria. For example, all traditional supervised and unsupervised learning (respectively, SL and UL) algorithms depend on a formally defined objective function which evaluates the agent's responses to stimuli and thus drives structural changes that result in better responses in the future. While these algorithms can learn many different tasks (e.g., image classification, natural language processing, visual scene parsing etc.), they are typically trained on datasets and objective functions assigned by the engineer. On the other hand, active learning agents (Thrun 1994; Cohn, Ghahramani, and M. Jordan 1995) feature algorithms that actively sample the data they learn from. This is particularly the case in reinforcement learning (RL; Sutton and Barto 2018) where agents have control over the sampled data by virtue of causal interactions with their environments. Here, learning is driven by an evaluative objective criterion which comes in the form of a reward function. Like in SL and UL, what the agent ends up learning is determined by a predefined criterion, but additional com-plications arise because of the need to sample experiences that help the agent improve. In realistic settings, only a tiny fraction of all possible experiences are relevant, which can be further complicated by the sparsity or deceptiveness of rewards. Sampling relevant experiences while avoiding noise is an important problem which will receive much attention in this chapter.

In contrast to ML agents that learn externally assigned tasks, biological agents, particularly humans, often have autonomy not only in how they choose experiences to learn tasks, but also in choosing what tasks to learn. We seek information when we become curious without being told what to be curious about. Seeking information out of curiosity, rather than to achieve a separable outcome, like food or money, is characterized in psychology as intrinsically motivated (Harlow, Harlow, and Meyer 1950; Ryan and Deci 2000). In addition to information-seeking phenomena like curiosity (Bazhydai, Twomey, and Westermann 2020) and interest (Hidi and Renninger 2006), intrinsic motivation is also linked to other hallmarks of human behavior such as creativity (Gross, Zedelius, and Schooler 2020) and play (Chu and Schulz 2020).

Similarly to humans but unlike the traditional ML systems mentioned above, intrinsically motivated artificial agents control what they learn through autonomous and task-independent sampling of learning experiences. In order to collect learning experiences, agents explore learning situations. To better appreciate what we mean by learning situations are and how they relate to various mechanisms and functions of intrinsic motivation, consider the following scenarios:

1. A virtual robot trying random actions in its environment
2. A toddler trying to build the tallest possible tower with toy blocks

3. A curious student raising his hand to ask a question
4. A human infant looking at where her mother is pointing
5. A rat exploring a maze to get familiar with its environment
6. A scientist designing the next experiment

All of these scenarios describe an agent interacting with its environment and thereby engaging in a learning situation[1]. Importantly, these situations are not merely different due to the different kinds of agents involved, but because they involve different mechanisms that have distinct functions. Mechanisms of curiosity-driven exploration are composed of (a) the interface through which agents sample learning situations, i.e., the space in which they can make choices; (b) the principle by which agents rank learning situations within this space. We describe these dimensions more thoroughly in section **Mechanisms**, but the examples above already illustrate that agents can sample learning situations by choosing what actions to take (1), what goals to pursue (2), or whom to ask (3). Moreover, the decisions between the alternatives may be based on prior knowledge (3), driven by competence (2), or influenced socially (4). Note different mechanisms illustrated by the examples imply distinct consequences, suggesting that they may serve distinct functions. In section **Functions**, we locate these functions on the axis of causal proximity to evolutionary fitness: from the rat exploring a maze (5) to the scientist exploring an obscure domain (6). Finally, in sections **Usages** we discuss the relationships between AI and psychology and briefly survey some of the use cases for various intrinsically motivated algorithms outside the realm of AI research.

2 MECHANISMS

As mentioned in the introduction, researchers in AI have proposed a wealth of algorithms for curiosity-driven exploration. These algorithms differ by how they address two related subproblems:

- How to parameterize learning situations?
- How to choose learning situations?

The first subproblem is addressed by defining the choice space to drive exploration (Moulin-Frier and Oudeyer 2013). Intuitively, a choice space serves as a basis for assessing different learning situations. Actualizing choices in a choice space results in agent-environment interactions from which the agent can learn. This chapter reviews three kinds of choice spaces that consist of either actions, goal states, or social partners. These choice spaces correspond to different ways in which learning situations can be parameterized. Parameterizing learning situations on the basis of actions lets the agent consider what can be learned by performing these actions; parameterizing learning situations on the basis of goals lets the agent consider what can be learned by pursuing goals; parameterizing learning situations on the basis of social partners lets the agent consider what can be learned by interacting with others. We discuss the main ways in which choice spaces are specified and how they differ in the next subsection (**Exploration bases**).

Given a fully specified choice space, the agent can make decisions within that space. Specifying this decision-making process addresses the second subproblem. Agents choose what to learn by

[1] Of course, the agent may not learn upon every single interaction with the environment, but any interaction creates a situation where learning could happen.

following a certain strategy by which they assign “interestingness” to the available choices, thereby determining which learning situations are more likely to be approached. For example, choices can be driven by features of learning situations, such as the amount of knowledge a situation might bring or how novel it is; they can also be made completely at random (e.g., Colas, Sigaud, and Oudeyer 2018; Colas, Karch, Lair, et al. 2020). We provide a brief survey of these methods in the subsection on **Exploration strategies**.

2.1 EXPLORATION BASES

We stated before that a learning situation arises whenever an agent interacts with its environment. Active interactions entail that the agent has to decide how to act in a given context, and upon deciding and acting, gets to observe the effects of its actions. How the agent explores, therefore, depends primarily on how the agent contextualizes its interactions. Specifically, the agent can explore by considering what actions to take, what goals to pursue, or what social partners to engage. Sets of actions, goals, or social partners provide a basis for comparing potential interactions. The rest of this section reviews different ways in which such bases are defined, used, and represented.

2.1.1 Exploring by choosing actions

One family of approaches considers agents which observe the current state of the environment as a context, choose actions to execute, and observe the resulting following state. Here, the objective of exploration is to select the actions which generate informative data for learning an internal model of causal dynamics through SL, UL or RL. For example, several SL-based robotic agents (Oudeyer, Frdric Kaplan, and Hafner 2007; Caligiore et al. 2008; Baranes and Oudeyer 2009; Saegusa et al. 2009; Lefort and Gepperth 2015) maintain world models representing their knowledge about either forward dynamics (inferring future states from specific actions), or inverse dynamics (inferring the right actions to bring about specific states), or both. These systems learn from observations borne out of the actions they choose. Therefore, exploration of learning situations in these approaches corresponds to making choices in the action space. Examples of action-space exploration can also be found in RL settings (Singh et al. 2010; Bellemare et al. 2016; Jaderberg et al. 2016; Pathak et al. 2017; Tang et al. 2017; Burda et al. 2018; Haber et al. 2018; Bougie and Ryutaro Ichise 2020). In intrinsically motivated RL, agents learn behavioral policies by maximizing intrinsic rewards (e.g., rewards based on state novelty as in Tang et al. 2017; on model prediction error as in Pathak et al. 2017; or on surprise as in Berseth et al. 2021). In these systems, actions that bring about rewarding states get reinforced. Thus, the agent collects learning data (state transitions) by ranking actions according to their capacity to yield intrinsic rewards.

2.1.2 Exploring by choosing goals

Another family of approaches considers agents making choices in a goal space instead of an action space. In the general case, such agents learn to represent and sample their own goals, i.e., they are autotelic (Colas, Karch, Sigaud, et al. 2021). This family of approaches is referred to as Intrinsically Motivated Goal-Exploration Processes (or IMGEP for short; Forestier et al. 2017; Colas, Karch, Sigaud, et al. 2021). Here, the notion of a goal is generalized: it refers to any set of constraints on any set of future sensorimotor representations (Colas, Karch, Sigaud, et al. 2021). This abstract conceptualization enables researchers to express all kinds of goals, ranging from particular world states (e.g., coordinates of the agent’s hand must be equal to a specific x, y, and z), to constraints on entire behavioral trajectories and their linguistic descriptions (e.g., ”water the plant and then feed the dog”; Colas, Karch, Lair, et al. 2020). In all cases, goals are specified by two essential

components: (1) goal representation that specifies the criteria and (2) goal-achievement function that signals whether the criteria are met. Goals are usually represented as numerical vectors (sometimes called goal embeddings) that comprise abstract goal spaces from which specific goals can be sampled, while goal achievement is evaluated using logical operations.

Examples of IMGEPs can be found in SL and UL contexts (M. Jordan and Rumelhart 1992; Rolf, Steil, and Gienger 2010; Baranes and Oudeyer 2013; Moulin-Frier and Oudeyer 2013; Forestier et al. 2017; Takahashi et al. 2017; Reinke, Etcheverry, and Oudeyer 2020; Laversanne-Finot, Pere, and Oudeyer 2018). In these frameworks, agents autonomously engage in learning situations by attempting to reach self-selected goals. Note that this process is markedly different from the one described above, where the agent accesses learning situations by choosing among the available actions. In IMGEPs, the agent can consider any goal it may imagine, not just the states that its actions may bring about.

The domain of goal-conditioned RL works with agents that learn action policies conditioned on goals. These agents base their actions not only on the current state (as in traditional RL) but also the goal encoding, which means they can act differently in the same situation depending on what they are after (Schaul et al. 2016). Intrinsically motivated goal-conditioned RL builds upon that framework and allows agents to generate their own goals (see Colas, Karch, Sigaud, et al. 2021, for a recent review). Some notable examples include Colas, Sigaud, and Oudeyer 2018; Nair et al. 2018; Colas, Fournier, et al. 2019; Pong et al. 2020; Colas, Karch, Lair, et al. 2020. While there is a great deal of variability among strategies for sampling interesting goals (see **Exploration strategies**), all of these goal-oriented agents make decisions in a goal space rather than in an action space. Goal-oriented intrinsically motivated learning has a number of advantages compared to action-oriented intrinsically-motivated learning: it improves the performance and convergence time when learning inverse models in high-dimensional spaces with highly non-linear mappings (Baranes and Oudeyer 2013), it automatically generates learning curricula from easy to more complex skills (Moulin-Frier, Nguyen, and Oudeyer 2014); finally, it enables hindsight learning (Forestier et al. 2017; Andrychowicz et al. 2018; Colas, Fournier, et al. 2019). We discuss these positive practical implications in more detail in **Functions**.

2.1.3 Exploring by choosing social partners

A few contributions have explored how IM can be coupled with social interaction. The SGIM-ACTS architecture (Nguyen and Oudeyer 2012) considers an IM agent that is able to choose whether and when to learn from a social peer or by autonomous goal generation, from which social peer to learn, and what to ask the chosen social peer. Interacting with the social peer becomes part of the choice space and the agent makes hierarchical decisions: it first decides to interact or to self-explore its own goals, then which social peer or self-generated goal to focus on. The SGIM-ACTS framework was also applied to agents equipped with a realistic computer model of the human vocal tract and was able to reproduce the main developmental stages of infant vocal development (Moulin-Frier, Nguyen, and Oudeyer 2014). More recent contributions consider social influence as intrinsic motivation for achieving coordination and communication in multi-agent RL (Jaques et al. 2019). Other work in multi-agent RL have theorized and demonstrated how competition and cooperation display intrinsic dynamics resulting in a naturally emergent curriculum (Leibo et al. 2019; Baker et al. 2020).

Representing choice spaces Precisely what do choices in choice spaces represent? In the case of action-space exploration, actions can represent micro-actions responsible for transitions between temporally adjacent momentary states, like pixel images (e.g., Bellemare et al. 2016; Pathak et al.

2017). In this case, the agent can only learn from short-term transitions which limits their ability to efficiently learn regularities spanning larger time scales[2]. However, the agent can sample learning situations by making decisions in the space of macro-actions (e.g., action-policies), rather than micro-actions, which enables exploration of more temporally extended effects of its actions (see Baranes and Oudeyer 2013). In the case of goal sampling, the choice space can correspond entirely to the state space, so that a given goal represents a particular state. This can be problematic for very high-dimensional spaces. For example, if the state space is a space of pixel images, because sampling directly from this space yields white noise most of the time, just like blindly picking alphabet letters would produce mostly gibberish (Nair et al. 2018). One solution to this problem is to define the goal space as some high-level feature space (also called latent or embedding space) of the raw-image space (e.g., Laversanne-Finot, Péré, and Oudeyer 2021). Following the letter-picking analogy, this would correspond to composing and sampling from a higher-level space of syllables or words which is more likely to produce meaningful strings. A related issue in cases when the goal space is defined over some abstraction over the raw sensory experience (e.g., 2D positions of objects vs raw pixel images), is whether to assume that this abstract space is given to the agent like in Forestier et al. (2017) or whether the agent needs to learn a latent goal space from scratch (e.g., Nair et al. 2018; Laversanne-Finot, Péré, and Oudeyer 2021). Whether learned or given, latent goal representations help agents target valid achievable goals Pathak et al. 2017.

2.2 EXPLORATION STRATEGIES

Given a well-defined choice space for exploration, what strategies can an artificial agent follow to decide which learning situations are more or less interesting? Comprehensive reviews of different approaches can be found elsewhere (Oudeyer and Kaplan 2007; Mirolli and Baldassarre 2013; Aubret, Matignon, and Hassas 2019; Linke et al. 2020), so we only provide a short survey of different approaches with a focus on their diversity rather than precise implementations.

We group existing approaches into three main categories: undirected, knowledge-based, and competence-based exploration. While all learning situations are equally interesting in undirected exploration, directed exploration strategies scale interestingness with the agent's abilities. Directed exploration can be divided into two broad classes: knowledge-based and competence-based strategies Oudeyer and Kaplan 2007. Sometimes the distinction between the two can be subtle because knowledgeable systems can also be competent and competent systems can be knowledgeable Mirolli and Baldassarre 2013. The point of divergence for these families of mechanisms is that to a knowledge-based system, the interestingness of a learning situation is determined by its relation to the system's knowledge. On the other hand, to a competence-based

[2] Model-free RL agents (e.g., Bellemare et al. 2016; Pathak et al. 2017) do not learn from the transitions per se – they learn from rewards. Specifically, the agent usually learns a value function V, mapping states to their expected cumulative reward. Even if the transitions are momentary, the agent can still maximize their long-term cumulative reward using techniques like bootstrapping (Sutton and Barto 2018). Another exception is when the world model is represented by a recurrent neural network (RNN; Takahashi et al. 2017) as RNNs can encode time series.

system, a given learning situation may be more or less interesting because it relates to the system's ability to reach a specific self-generated goal.

2.2.1 Undirected exploration

Undirected exploration (sometimes random or uniform exploration) refers to a strategy that assigns interest uniformly across the choice space (making all learning situations equally interesting). The effectiveness of this simple strategy is inconsistent across different settings. When applied to action-space exploration, for example, undirected exploration is only effective for simplistic problems (Baranes and Oudeyer 2013; Benureau and Oudeyer 2016), for example, when the environment provides dense rewards or when actions have simple and consistent effects. In more challenging settings, where the mapping between actions and their effects exhibits a combination of non-linearity, stochasticity, and redundancy, motor exploration is not sufficient for effective learning, but goal exploration could be (Moulin-Frier and Oudeyer 2013). This is largely because learning how to reach goals contributes to the agent's competence and thus its ability to control the environment, while learning about various outcomes of all of one's actions in all possible contexts may be worthless in practice (Mirolli and Baldassarre 2013). Besides, trying out random actions from a particular state may be futile for reaching certain hard-to-reach regions of the state space. Think of how hard it would be to learn how to drive from home to work by performing random actions (you would end up crashing your car most of the time). If instead you were to learn how to drive to various places from home (your driveway, a cornershop, a nearby gas station) your chances of finding your way to your office eventually would be much higher.

Despite its simplicity, random goal exploration has proven to be surprisingly efficient, leading to some forms of novelty search as an emergent feature and surpassing directed approaches operating in the action space in the learning of redundant inverse mappings (Benureau and Oudeyer 2016; Colas, Sigaud, and Oudeyer 2018). Because it is simple and computationally cheap, random goal exploration is often combined with other strategies, either to jumpstart the primary mode of directed exploration by collecting initial data, or sometimes as a complementary strategy at a certain level of hierarchical sampling decisions in modular spaces (e.g., Forestier et al. 2017), and sometimes as an epsilon-greedy strategy (see Sutton and Barto 2018b) to balance between random and directed exploration (e.g., Colas, Fournier, et al. 2019). Still, in many situations, undirected exploration may not be as efficient or effective as more sophisticated guided exploration approaches. For example, random (goal) exploration performs poorly when the space of effects has a hierarchical structure, so that certain states are only accessible through reaching some prerequisite states (e.g., Forestier et al. 2017). An agent exploring goals randomly is unlikely to ever get to practice these "out-of-reach" goals, unless there is a mechanism for imagining them that leverages structured representations of goals, such as natural language encodings (e.g., Colas, Karch, Lair et al. 2020).

2.2.2 Knowledge-based exploration

Knowledge-based exploration is perhaps the most diverse family of intrinsically motivated strategies. Not only are there many ways in which one can characterize a relation between learning situations and the agent's knowledge, but there are many kinds of knowledge that agents represent. For example, an agent can maintain a predictive causal model of the effects of its actions and have a metacognitive monitoring system track errors that this predictive model commits. Equipped with such a system, the agent can measure interestingness of actions based on, for example, outcome prediction error of the forward model (e.g., Saegusa et al. 2009). Specifically, the agent can be more

(or less) interested in taking actions for which the forward model does accurately predict the consequences (known as prediction-error strategy). Alternatively, the agent can track changes in prediction accuracy (a strategy known as learning progress; e.g. Schmidhuber 1991b; Oudeyer, Frdric Kaplan, and Hafner 2007; Kim et al. 2020). Here the agent would be more (or less) interested in actions, predictions for which get more accurate with time. While the measure of interestingness in these approaches is based on the predictions from a forward model, exploring agents can also monitor the behavior of an inverse dynamics model (Pathak et al. 2017; Haber et al. 2018).

A distinct subclass of knowledge-based strategies relies on knowledge about frequencies of observed states. In so-called count-based approaches (Bellemare et al. 2016; Tang et al. 2017), this knowledge is represented explicitly, allowing the agent to selectively explore over- or under-visited states. Other systems in this subclass of approaches incorporate different variants of autoencoder networks to learn latent spaces (Twomey and Westermann 2018; Bougie and Ichise 2020) or generative models (Nair et al. 2018; Pong et al. 2020) of states observed by the agent. Specifically, Twomey and Westerman defined several interestingness measures based on backpropagation computation of their category-learning neural network, including measures of weight update, prediction error, and activation-function derivative (which model, respectively, the system's curiosity, novelty, and plasticity; Twomey and Westermann 2018). Bougie and Ichise introduced auxiliary tasks of image reconstruction with context-based autoencoders and defined an intrinsic reward measure derived from reconstruction errors from these tasks (Bougie and Ichise 2020). Others (Nair et al. 2018; Pong et al. 2020) employed variational autoencoders to estimate valid state distributions in order to guide exploration. Although these models do not explicitly encode state visitation counts, the interestingness measures defined on their basis are related to frequencies of the observed states. Autoencoder prediction error, for instance, should decrease with repeated exposure to a given state. Variational autoencoders additionally represent the statistics of the latent space – a feature that can be used to estimate the likelihood of any state (including completely novel states) given what has been observed in the past.

2.2.3 Competence-based exploration

Another family of strategies assigns interestingness based on competence. These approaches do not need to assume any explicit world-dynamics knowledge model (although they can), so they can be readily incorporated in model-free learning systems without the need to introduce any auxiliary tasks. Competence-based exploration strategies are especially well-suited for intrinsically motivated agents that explore self-imposed goals, since the notion of competence corresponds naturally to goal achievement. A simple competence-based heuristic in a goal-oriented context is to sample goals according to the agent's ability to achieve them. For example, Bougie and Ichise reward the agent in a given state based on how incompetent the agent perceives itself to be in that state (Bougie and Ichise 2019). Here the agent generates its data set by taking interest in actions that lead it to states at which it deems itself incompetent, i.e. by choosing actions that maximize incompetence. In a different approach, Florensa and colleagues leveraged the generative power of adversarial networks for generating goals, which the agent evaluates based on the probability of reaching them (Florensa et al. 2018). Thanks to this evaluation, their agent can prioritize goals of intermediate difficulty, thereby avoiding goals that it already knows how to achieve and goals that it knows it cannot achieve. Santucci et al. compared the efficacy of several interestingness measures for autonomous mastering a set of tasks that included unreachable distractor tasks. The best-performing measure that allowed their agent to learn all learnable tasks in the least amount of time was based on

competence prediction-error (Santucci, Baldassarre, and Mirolli 2013). Several other teams have explored modular goal spaces using measures of competence progress (Stout and Barto 2010; Forestier et al. 2017; Colas, Fournier, et al. 2019). Agents in these studies were incentivized to sample goals from the predefined regions (modules) of the goal space where competence was either improving or deteriorating. Oudeyer and colleagues used a similar competence progress-based strategy but in settings where the singular goal-space was progressively modularized (Baranes and Oudeyer 2013; Moulin-Frier, Nguyen, and Oudeyer 2014).

3 FUNCTIONS

The previous section reviewed the diversity of approaches to specifying intrinsically motivated mechanisms in AI. We have mentioned in the introduction that these mechanisms are useful for autonomous learning in the absence of externally assigned objectives, but what specific functional consequences do different intrinsic motivation mechanisms bear? This section focuses on the functional aspects of intrinsically motivated systems addressing the question of how different intrinsic-motivational systems are useful for both artificial and biological agents.

Singh et al. provide a useful connection between the concepts of extrinsic and intrinsic motivation and the concepts of primary and secondary rewards (Singh et al. 2010). While the reception of primary rewards (e.g., related to nutrients, sex, or pain) contributes to an agent's survival and, thus, its reproductive success, secondary reinforcers signal anticipation of primary rewards, but are neutral a priori and have to be learned. In RL, outputs of the reward function can be analogous to primary reward signals, because the reward function is given to the agent. However, value functions evaluate states based on their learned expected cumulative reward, and therefore outputs of the value function are analogous to secondary reinforcers. In this formulation, intrinsic motivators (e.g., novelty, uncertainty, learning progress etc.) are primary reinforcers because their rewardingness is a given. On the other hand, predictors of intrinsic primary reinforcers can acquire rewarding qualities through the learning of secondary reinforcers, in the same way as non-rewarding states gain value due to extrinsic reinforcers. Therefore, intrinsic primary rewards differ from extrinsic primary rewards, mostly due to their causal proximity to evolutionary success. As it is often the case in biology (Dobzhansky 1973), it is sensible to analyze the functional aspect of artificial intrinsic-motivational mechanisms in light of evolution – an exercise that allows us to consider the possible bio-ecological roles of engineered curiosity-driven systems.

In what follows, we build upon Singh et al.'s framework as well as the taxonomy of mechanisms we proposed in the previous section in order to extract key functional aspects of intrinsic motivation in both biological and artificial agents. We start from the more evolutionarily proximal functions (direct procurement of primary rewards) and gradually consider increasingly distal ones (learning of internal models, goal discovery, and cultural innovation).

3.1 PROCUREMENT OF EXTRINSIC PRIMARY REWARDS

In relatively dense primary-reward environments, unstructured random exploration in the action space is sufficient to efficiently elicit extrinsic primary rewards (as it is the case in some standard AI benchmarks based on video games, e.g., Mnih et al. 2015). This points to the most proximal function of intrinsic motivation: the generation of diverse sensorimotor experiences. The direct benefit of generating diverse experiences is to increase the probability of eliciting primary extrinsic rewards.

However, the diversity generated by random exploration is usually not sufficient in sparser reward environments (e.g., Pathak et al. 2017). Knowledge-based exploration can increase the diversity of learning experiences by guiding exploration based on different measures of interestingness (see section Exploration principles). Discovering extrinsic primary rewards such as food, water, or shelter through exploration is clearly linked to the agent's well-being.

3.2 LEARNING INTERNAL MODELS

A more distal function of intrinsic motivation is the learning of internal models of the agent-environment interaction (e.g. forward or inverse models) that can enhance the agent's decision-making abilities. Learning such a model can be formalized as an UL or an SL problem and strongly relies on the information contained in the training dataset. Autonomous agents generate this dataset by interacting with the environment, and the key role of intrinsic motivation is to generate informative training data. Research in intrinsically motivated SL extensively studied two main cases. On the one hand, knowledge-based approaches have proven to be efficient in generating informative data for learning forward models (Oudeyer, Kaplan, and Hafner 2007). On the other hand, competence-based approaches have proven to be more efficient than knowledge-based approaches for learning inverse models (Baranes and Oudeyer 2013). RL agents can also benefit from internal models, be it in the form of a value function and an action policy in model-free RL (Pathak et al. 2017), or in the form of a world-dynamics model (forward or inverse) in model-based RL (Haber et al. 2018). Understanding how the world works per se does not put proverbial food in the agent's mouth, but it allows the agent to act more intelligently in novel situations and plan ahead in order to obtain what it needs more reliably.

3.3 GOAL DISCOVERY

The third level of functions we propose is related to the discovery and learning of novel goals and the associated skills to achieve them. This is the main function of competence-based approaches, where exploration is guided by the pursuit of self-imposed goals. These approaches can automatically organize exploration from simple to more complex skills (Forestier et al. 2017; Pong et al. 2020), as well as discover the full range of the achievable behavioral repertoire, possibly in an open-ended manner (Colas, Fournier, et al. 2019). There are multiple functional advantages to discover and master novel goals that are not extrinsically rewarding. First, in environments where eliciting extrinsic primary rewards require the acquisition of complex skills (e.g., hunting), it is crucial to structure learning in a curriculum from simple (e.g., locomotion) to more complex skills. Complex skill sets often display a hierarchical structure, where mastering easier ones is pre-requisite for acquiring more complex ones. Second, the ability to autonomously explore and discover new goals and skills provides a crucial advantage in changing environments. For example, paleoclimatological data provides evidence for strongly varying climate conditions in the Rift Valley at East Africa approximately 7 million years ago, and it is hypothesized that the ability to rapidly and flexibly reorganize a diverse behavioral repertoire was a key requirement to adapt to such unprecedented conditions (Potts 2013). Thus, the ability to autonomously generate and master novel goals and thereby acquire a diverse repertoire of complex skills provides a crucial advantage for a species' success in such settings of strong environmental variability (see Nisioti and Moulin-Frier 2020, for a recent proposition to apply this principle in AI).

3.4 CULTURAL INNOVATION

Finally, the fourth and most distal level of functions we propose concerns cultural innovation. Several theoretical contributions proposed a potential role of curiosity-driven exploration in both language acquisition (Oller 2000) and evolution (Oudeyer and Smith 2016). From a sensorimotor perspective, active exploration can spontaneously generate diverse behaviors from modality-independent and task-independent internal drives. Such spontaneous behavior can result in vocal activity that may have bootstrapped the emergence of communication. This hypothesis is supported by computational simulations showing a role of curiosity-driven exploration in vocal development (Moulin-Frier, Nguyen, and Oudeyer 2014), social affordance discovery (Oudeyer and 2006), and active control of the emerging conventions in social lexicon (Schueller, Loreto, and Oudeyer 2018). From a cognitive perspective, compositional language itself is a powerful cognitive tool to imagine novel out-of-distribution goals in competence-based intrinsic motivation (Colas, Karch, Lair, et al. 2020). Moreover, recent contributions in multiagent RL have shown how an auto-curriculum of increasingly complex behaviors displaying features of open-ended innovation can emerge from agents' co-adaptation in mixed cooperative-competitive environments (Baker et al. 2020). Such mechanisms are potential precursors of cultural evolution in the human species. Cultural evolution has triggered increasingly complex technological innovation across generations (Fogarty and Creanza 2017). A prime example of this is the industrial revolution, which has resulted in a rapid acceleration of global population growth in the 19th century (Lucas 2004).

4 USAGES

Other than helping artificial agents explore learning situations in abstract task-independent contexts, how can intrinsically-motivated learning algorithms be used? We identify two main directions in which such algorithms can have high impact. On the one hand, they can be a great tool for advancing research in cognitive psychology and neuroscience. Outside cognitive research, these algorithms can be applied directly to problems that require intelligent automated exploration. We review both of these domains of application in the rest of this section.

4.1 COGNITIVE MODELING

The notion of intrinsically-motivated exploration in psychology has been developing – for the most part – independently of AI (Kaplan and Oudeyer 2007). Psychological investigations can be traced back to psychophysiological perspectives on exploration in the early 1940s (Hull 1943). Since then, psychological views on curiosity and information-seeking have undergone multiple changes (Loewenstein 1994; Bazhydai, Twomey, and Westermann 2020) and are now becoming more integrated with the computational perspective (Kaplan and Oudeyer 2007; Gottlieb, Oudeyer, et al. 2013). On the other hand, early formulations of intrinsic motivation in AI were either "discovered by accident" (Andreae and Andreae 1977, p. 5) or influenced by relatively distant research areas, like biological autopoiesis (Maturana and Varela 1980) and aesthetic information theory (Nake 1976, as cited in Schmidhuber 1991a), yet not the aforementioned psychological literature (see Kaplan and Oudeyer 2007, for a historical overview). Over the course of history, psychology and AI have actually been converging on similar ideas for why certain behaviors could be intrinsically rewarding: due to some kind of mismatch between bottom-up observations and top-down predictions (Frederic and Oudeyer 2007).

Evolutionary implications of artificial intrinsic-motivational systems discussed earlier (see **Functions**) reinforce the need to seriously consider them as good candidate models for human curiosity-driven learning. A major advantage that comes naturally with these systems is their precise formulation. Such a formal description unambiguously discloses crucial structural and functional properties of the system in question and thus enables to advance the related theory more efficiently (McClelland 2009).

Cognitive models based on artificial exploring agents are becoming increasingly frequent. For instance, Moulin-Frier et al. explained the progression of human vocal behavior through distinct developmental stages as an intrinsically motivated, competence-progress based goal-exploration process with an emergent curriculum (Moulin-Frier, Nguyen, and Oudeyer 2014). (Such computational accounts of curiosity-driven learning have led to novel hypotheses on the role of curiosity in the evolution of language; see Oudeyer and Smith 2016). More recently, Twomey and Westermann used an actively exploring autoencoder network to hypothesize an algorithmic-level description of visual exploration in infants (Twomey and Westermann 2018); Poli et al. compared several knowledge-based sampling strategies to predict visual-attention control in infants (Poli et al. 2020). Moreover, computational models of intrinsic motivation are invoked to explain self-determined instrumental (Gershman 2019) and non-instrumental (Ten et al. 2021) choices in human adults.

Relating specific mechanistic implementations of curiosity to a common underlying structure can be beneficial for revealing potential theoretical and empirical gaps in the scientific understanding of human information-seeking. For instance, while information-seeking literature is brimming with work investigating episodic curiosity-driven sampling over short time scales (see Bazhydai, Twomey, and Westermann 2020, for a recent review), only a handful of studies have looked at active time-extended curiosity-driven exploration (e.g., Holm, Wadenholt, and Schrater 2019; Ten et al. 2021). Similarly, while most behavioral paradigms study mechanisms with action-based choice spaces (e.g., Twomey and Westermann 2018), hardly any studies investigate intrinsically motivated exploration of goals in humans, specifically of how humans generate and choose new goals in unfamiliar sensorimotor environments.

Of course, successful curiosity-driven systems developed in AI should not be simply adopted as cognitive models of human exploration. Data from carefully designed behavioral studies should bear more weight on the truthfulness of propositions about the mechanisms implemented in biological organisms than the success of these mechanisms in virtual settings. Behavioral studies employing model comparison techniques (e.g., Poli et al. 2020; Ten et al. 2021) can evaluate different AI-inspired proposals by their ability to explain human or animal data. Such studies could also fuel novel ideas to improve existing AI systems, as it was done in instrumental problem settings (Lin et al. 2020). These are exciting areas of future research that can promote synergy between the two fields, allowing us to hone in on a more comprehensive mechanistic understanding of intrinsically-motivated information-seeking in biological organisms.

4.2 PRACTICAL APPLICATIONS

Functional diversity of intrinsic-motivational mechanisms makes them useful for practical applications, such as automated knowledge discovery and education. Mechanisms of intrinsic motivation help autonomous agents learn in settings with complex sensorimotor spaces and sparse or non-existent rewards. They are crucial for building autonomous control systems that can learn

efficiently in open-ended environments. The practical effectiveness of these mechanisms has been recently demonstrated in studies of automated discovery in complex systems, where curious agents learn to control diverse effects in complex non-linear settings, such as smartphone applications (Pan et al. 2020), continuous cellular automata (Etcheverry, Moulin-Frier, and Oudeyer 2021), and real-world chemical systems (Grizou et al. 2020). Automated discovery can have a high societal impact by assisting both scientific research and artistic creation. Moreover, since one of the functions of curiosity-driven systems is knowledge acquisition, directed exploration strategies of such systems can be used to assist learners when they behave suboptimally. Importantly, such intelligent tutoring systems can be tailored to the current levels of knowledge or competence of individual learners (Clement, Oudeyer, and Lopes 2016) and have shown promising results in pedagogical settings (Clement, Roy, et al. 2015; Delmas et al. 2018), where they assist learners in selecting topics and exercises that maximize their individual learning progress.

5 CONCLUSION

The goal of this chapter was to familiarize the reader with the diversity of computational mechanisms and possible evolutionary functions of curiosity-driven exploration. We identified an important problem facing autonomous agents that have control over their learning experiences. Specifically, such agents must decide how to sample learning data in the absence of externally imposed tasks. We briefly reviewed several ways in which artificial agents choose actions, goals, or social peers in order to engage in learning situations. We presented distinct families of exploration strategies, including undirected, knowledge-based, competence-based and socially-influenced approaches. We then discussed how these mechanisms can contribute to evolutionary success at different levels: by helping agents to approach primary rewards, acquire world models, discover goals, and bootstrap a cultural repertoire. Finally, we provided some contemporary examples of how IM algorithms are used in practical applications as well as in cognitive research. We hope that this concise bird's-eye perspective – organized along the proposed mechanistic, functional, and pragmatic dimensions of curiosity-driven exploration – can serve as a stepping stone towards a unified taxonomy of this fascinating and important field.

References

Andreae, P. M., & Andreae, J. H. (1978). A teachable machine in the real world. *International Journal of Man-Machine Studies*, *10*(3), 301-312.

Andrychowicz, M., Wolski, F., Ray, A., Schneider, J., Fong, R., Welinder, P., ... & Zaremba, W. (2017). Hindsight experience replay. *arXiv preprint arXiv:1707.01495*.

Aubret, A., Matignon, L., & Hassas, S. (2019). A survey on intrinsic motivation in reinforcement learning. *arXiv preprint arXiv:1908.06976*.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., & Mordatch, I. (2019). Emergent tool use from multi-agent autocurricula. *arXiv preprint arXiv:1909.07528*.

Baranes, A., & Oudeyer, P. Y. (2009). R-iac: Robust intrinsically motivated exploration and active learning. *IEEE Transactions on Autonomous Mental Development*, *1*(3), 155-169.

Baranes, A., & Oudeyer, P. Y. (2013). Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, *61*(1), 49-73.

Barron, A. B., Hebets, E. A., Cleland, T. A., Fitzpatrick, C. L., Hauber, M. E., & Stevens, J. R. (2015). Embracing multiple definitions of learning. *Trends in neurosciences*, *38*(7), 405-407.

Bazhydai, M., Twomey, K., & Westermann, G. (Accepted/In press). Curiosity and Exploration. In J. B. Benson (Ed.), Encyclopedia of Infant and Early Childhood Development (2nd ed. ed.). Elsevier.

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, *29*, 1471-1479.

Benureau, F. C., & Oudeyer, P. Y. (2016). Behavioral diversity generation in autonomous exploration through reuse of past experience. *Frontiers in Robotics and AI*, *3*, 8.

Berseth, G., Geng, D., Devin, C., Rhinehart, N., Finn, C., Jayaraman, D., & Levine, S. (2019). SMiRL: Surprise Minimizing Reinforcement Learning in Unstable Environments. *arXiv preprint arXiv:1912.05510*.

Bougie, N., & Ichise, R. (2020). Skill-based curiosity for intrinsically motivated reinforcement learning. *Machine Learning*, *109*(3), 493-512.

Bougie, N., & Ichise, R. (2021). Fast and slow curiosity for high-level exploration in reinforcement learning. *Applied Intelligence*, *51*(2), 1086-1107.

Burda, Y., Edwards, H., Pathak, D., Storkey, A., Darrell, T., & Efros, A. A. (2018). Large-scale study of curiosity-driven learning. *arXiv preprint arXiv:1808.04355*.

Caligiore, D., Ferrauto, T., Parisi, D., Accornero, N., Capozza, M., & Baldassarre, G. (2008, April). Using motor babbling and hebb rules for modeling the development of reaching with obstacles and grasping. In *International Conference on Cognitive Systems* (Vol. 13, pp. 22-23).

Chu, J., & Schulz, L. E. (2020). Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, *2*, 317-343.

Clement, B., Oudeyer, P. Y., & Lopes, M. (2016). A Comparison of Automatic Teaching Strategies for Heterogeneous Student Populations. *International Educational Data Mining Society*.

Clement, B., Roy, D., Oudeyer, P. Y., & Lopes, M. (2013). Multi-armed bandits for intelligent tutoring systems. *arXiv preprint arXiv:1310.3174*.

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of artificial intelligence research*, *4*, 129-145.

Colas, C., Fournier, P., Chetouani, M., Sigaud, O., & Oudeyer, P. Y. (2019, May). CURIOUS: intrinsically motivated modular multi-goal reinforcement learning. In *International conference on machine learning* (pp. 1331-1340). PMLR.

Colas, C., Karch, T., Lair, N., Dussoux, J. M., Moulin-Frier, C., Dominey, P. F., & Oudeyer, P. Y. (2020). Language as a cognitive tool to imagine goals in curiosity-driven exploration. *arXiv preprint arXiv:2002.09253*.

Colas, C., Karch, T., Sigaud, O., & Oudeyer, P. Y. (2020). Intrinsically motivated goal-conditioned reinforcement learning: a short survey. *arXiv preprint arXiv:2012.09830*.

Colas, C., Sigaud, O., & Oudeyer, P. Y. (2018, July). Gep-pg: Decoupling exploration and exploitation in deep reinforcement learning algorithms. In *International conference on machine learning* (pp. 1039-1048). PMLR.

Delmas, A., Clement, B., Oudeyer, P. Y., & Sauzéon, H. (2018, November). Fostering health education with a serious game in children with asthma: pilot studies for assessing learning efficacy and automatized learning personalization. In *Frontiers in Education* (Vol. 3, p. 99). Frontiers.

Dobzhansky, T. (2013). Nothing in biology makes sense except in the light of evolution. *The american biology teacher*, *75*(2), 87-91.

Etcheverry, M., Moulin-Frier, C., & Oudeyer, P. Y. (2020). Hierarchically Organized Latent Modules for Exploratory Search in Morphogenetic Systems. *arXiv preprint arXiv:2007.01195*.

Florensa, C., Held, D., Geng, X., & Abbeel, P. (2018, July). Automatic goal generation for reinforcement learning agents. In *International conference on machine learning* (pp. 1515-1528). PMLR.

Fogarty, L., & Creanza, N. (2017). 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.

Forestier, S., Portelas, R., Mollard, Y., & Oudeyer, P. Y. (2017). Intrinsically motivated goal exploration processes with automatic curriculum learning. *arXiv preprint arXiv:1708.02190*.

Gershman, S. J. (2019). Uncertainty and exploration. *Decision*, *6*(3), 277.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society B*, *375*(1803), 20190502.

Gottlieb, J., & Oudeyer, P. Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758-770.

Gottlieb, J., Oudeyer, P. Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585-593.

Grizou, J., Points, L. J., Sharma, A., & Cronin, L. (2020). A curious formulation robot enables the discovery of a novel protocell behavior. *Science advances*, *6*(5), eaay4237.

Gross, M. E., Zedelius, C. M., & Schooler, J. W. (2020). Cultivating an understanding of curiosity as a seed for creativity. *Current Opinion in Behavioral Sciences*, *35*, 77-82.

Haber, N., Mrowca, D., Fei-Fei, L., & Yamins, D. L. (2018). Learning to play with intrinsically-motivated self-aware agents. *arXiv preprint arXiv:1802.07442*.

Harlow, H. F., Harlow, M. K., & Meyer, D. R. (1950). Learning motivated by a manipulation drive. *Journal of experimental psychology*, *40*(2), 228.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational psychologist*, *41*(2), 111-127.

Holm, L., Wadenholt, G., & Schrater, P. (2019). Episodic curiosity for avoiding asteroids: Per-trial information gain for choice outcomes drive information seeking. *Scientific reports*, *9*(1), 1-16.

Hull, C. L. (1943). Principles of behavior: An introduction to behavior theory. Oxford, England: Appleton-Century.

Jaderberg, M., Mnih, V., Czarnecki, W. M., Schaul, T., Leibo, J. Z., Silver, D., & Kavukcuoglu, K. (2016). Reinforcement learning with unsupervised auxiliary tasks. *arXiv preprint arXiv:1611.05397*.

Jaques, N., Lazaridou, A., Hughes, E., Gulcehre, C., Ortega, P., Strouse, D. J., ... & De Freitas, N. (2019, May). Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *International Conference on Machine Learning* (pp. 3040-3049). PMLR.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive science*, *16*(3), 307-354.
Kaplan, F., & Oudeyer, P. Y. (2007). In search of the neural circuits of intrinsic motivation. *Frontiers in neuroscience*, *1*, 17.
Kim, K., Sano, M., De Freitas, J., Haber, N., & Yamins, D. (2020, November). Active world model learning with progress curiosity. In *International conference on machine learning* (pp. 5306-5315). PMLR.
Laversanne-Finot, A., Péré, A., & Oudeyer, P. Y. (2018, October). Curiosity driven exploration of learned disentangled goal spaces. In *Conference on Robot Learning* (pp. 487-504). PMLR.
Laversanne-Finot, A., Péré, A., & Oudeyer, P. Y. (2020). Intrinsically motivated exploration of learned goal spaces. *Frontiers in neurorobotics*, *14*, 109.
Lefort, M., & Gepperth, A. (2015, August). Active learning of local predictable representations with artificial curiosity. In *2015 Joint IEEE Internatiol Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (pp. 228-233). IEEE.
Leibo, J. Z., Hughes, E., Lanctot, M., & Graepel, T. (2019). Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. *arXiv preprint arXiv:1903.00742*.
Lin, B., Cecchi, G., Bouneffouf, D., Reinen, J., & Rish, I. (2019). A story of two streams: Reinforcement learning models from human behavior and neuropsychiatry. *arXiv preprint arXiv:1906.11286*.
Linke, C., Ady, N. M., White, M., Degris, T., & White, A. (2020). Adapting Behavior via Intrinsic Reward: A Survey and Empirical Study. *Journal of Artificial Intelligence Research*, *69*, 1287-1332.
Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, *116*(1), 75.
Lucas, R. E. (2002). The industrial revolution: Past and future. *Lectures on economic growth*, 109-188.
Maturana, H. R., & Varela, F. J. (2012). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.
McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11-38.
Mirolli, M., & Baldassarre, G. (2013). Functions and mechanisms of intrinsic motivations. In *Intrinsically Motivated Learning in Natural and Artificial Systems* (pp. 49-72). Springer, Berlin, Heidelberg.
Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529-533.
Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P. Y. (2014). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in psychology*, *4*, 1006.
Moulin-Frier, C., & Oudeyer, P. Y. (2013, August). Exploration strategies in developmental robotics: a unified probabilistic framework. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (pp. 1-6). IEEE.
Nair, A., Pong, V., Dalal, M., Bahl, S., Lin, S., & Levine, S. (2018). Visual reinforcement learning with imagined goals. *arXiv preprint arXiv:1807.04742*.
Nake, F. (1976). Ästhetik als Informationsverarbeitung: Grundlagen und Anwendungen der Informatik im Bereich ästhetischer Produktion und Kritik. *Journal of Aesthetics and Art Criticism*, *34*(3).
Nguyen, S. M., & Oudeyer, P. Y. (2012). Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn*, *3*(3), 136-146.
Nisioti, E., & Moulin-Frier, C. (2020). Grounding Artificial Intelligence in the Origins of Human Behavior. *arXiv preprint arXiv:2012.08564*.
Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Oudeyer, P. Y. (2018). Computational theories of curiosity-driven learning. *arXiv preprint arXiv:1802.10546*.
Oudeyer, P. Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, *11*(2), 265-286.
Oudeyer, P. Y., & Kaplan, F. (2006). Discovering communication. *Connection Science*, *18*(2), 189-206.
Oudeyer, P. Y., & Kaplan, F. (2009). What is intrinsic motivation? A typology of computational approaches. *Frontiers in neurorobotics*, *1*, 6.
Oudeyer, P. Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492-502.
Pan, M., Huang, A., Wang, G., Zhang, T., & Li, X. (2020, July). Reinforcement learning based curiosity-driven testing of android applications. In *Proceedings of the 29th ACM SIGSOFT International Symposium on Software Testing and Analysis* (pp. 153-164).
Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017, July). Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning* (pp. 2778-2787). PMLR.
Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science advances*, *6*(39), eabb5053.
Pong, V. H., Dalal, M., Lin, S., Nair, A., Bahl, S., & Levine, S. (2019). Skew-fit: State-covering self-supervised reinforcement learning. *arXiv preprint arXiv:1903.03698*.
Potts, R. (2013). Hominin evolution in settings of strong environmental variability. *Quaternary Science Reviews*, *73*, 1-13.
Reinke, C., Etcheverry, M., & Oudeyer, P. Y. (2019). Intrinsically motivated discovery of diverse patterns in self-organizing systems. *arXiv preprint arXiv:1908.06663*.
Rolf, M., Steil, J. J., & Gienger, M. (2010). Goal babbling permits direct learning of inverse kinematics. *IEEE Transactions on Autonomous Mental Development*, *2*(3), 216-229.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68.
Saegusa, R., Metta, G., Sandini, G., & Sakka, S. (2009, February). Active motor babbling for sensorimotor learning. In *2008 IEEE International Conference on Robotics and Biomimetics* (pp. 794-799). IEEE.
Santucci, V. G., Baldassarre, G., & Mirolli, M. (2013). Which is the best intrinsic motivation signal for learning multiple skills?. *Frontiers in neurorobotics*, *7*, 22.
Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2015). Prioritized experience replay. *arXiv preprint arXiv:1511.05952*.
Schmidhuber, J. (1991a). A possibility for implementing curiosity and boredom in model-building neural controllers. In *Proc. of the international conference on simulation of adaptive behavior: From animals to animats* (pp. 222-227).
Schmidhuber, J. (1991b). Curious model-building control systems. In *Proc. international joint conference on neural networks* (pp. 1458-1463).
Schueller, W., Loreto, V., & Oudeyer, P. Y. (2018). Complexity Reduction in the Negotiation of New Lexical Conventions. *arXiv preprint arXiv:1805.05631*.
Singh, S., Lewis, R. L., Barto, A. G., & Sorg, J. (2010). Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development*, *2*(2), 70-82.
Stout, A., & Barto, A. G. (2010, August). Competence progress intrinsic motivation. In *2010 IEEE 9th International Conference on Development and Learning* (pp. 257-262). IEEE.
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
Takahashi, K., Ogata, T., Nakanishi, J., Cheng, G., & Sugano, S. (2017). Dynamic motion learning for multi-dof flexible-joint robots using active–passive motor babbling through deep learning. *Advanced Robotics*, *31*(18), 1002-1015.

Tang, H., Houthooft, R., Foote, D., Stooke, A., Chen, X., Duan, Y., ... & Abbeel, P. (2017). # exploration: A study of count-based exploration for deep reinforcement learning. In *31st Conference on Neural Information Processing Systems (NIPS)* (Vol. 30, pp. 1-18).

Ten, A., Kaushik, P., Oudeyer, P. Y., & Gottlieb, J. (2021). Humans monitor learning progress in curiosity-driven exploration. *Nature communications*, *12*(1), 1-10.

Thrun, S. (1995, January). A lifelong learning perspective for mobile robot control. In *Intelligent robots and systems* (pp. 201-214). Elsevier Science BV.

Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: a neurocomputational approach. *Developmental science*, *21*(4), e12629.