31 Curiosity-Driven Learning in Development

Computational Theories and Educational Applications

Pierre-Yves Oudeyer

31.1 Curiosity-Driven Exploration and Learning in Child Development

The timing and ordering of learning experience play a critical role in developmental processes. To make sense and learn skills out of the initial "buzzy blooming confusion," multiple mechanisms interact to canalize the infant through such ordered and progressively more complex learning experiences, harnessing the complexity of the world.

Which activities can be practiced, and what can be learned at a given point in development, is constrained both by the learning opportunities provided by the physical and social environment (caretakers in particular) and by current physical and cognitive abilities. For example, early in life infants cannot locomote due to both insufficient muscular power and lack of adequate sensorimotor coordination. At this stage, they can only observe or physically interact with objects and events in the environment which their parents set for them. Such constrained experience has been hypothesized to constrain and canalize species-typical development (Gottlieb, 1991). West and King (1987) proposed the idea of "ontogenetic niches" to capture the fact that developing organisms face an ordered series of environments that may be exploited by evolutionary processes to ensure adaptive outcomes.

In addition, learning experiences do not passively "happen" to infants. Rather, they play an active role in creating and selecting these experiences. In particular, infants spontaneously explore their body and how it can interact with the environment, physically experimenting with the effects of their arm or vocal tract movements, or the effects of touching, mouthing, grasping, or throwing all kinds of objects, often for the intrinsic pleasure of practicing these activities, without a separate distal goal. Such spontaneous exploration, often called "play" or "curiosity" in colloquial terms, is not random but rather organized, and partly results from brain mechanisms of intrinsic motivation (Gottlieb et al., 2013; Lowenstein, 1994) selecting sensorimotor activities which are "interesting."

Mechanisms of intrinsically motivated exploration and learning and the notion of "interestingness" have long remained studied at an intuitive level in psychology, where concepts like cognitive dissonance (Kagan, 1972), optimal incongruity (Hunt, 1965), intermediate novelty (Berlyne, 1960), or optimal challenge (Csikszenthmihalyi, 1991) were discussed. Furthermore, the relation between intrinsic motivation on one hand, and learning and development on the other hand has been little considered until recently. Yet, important advances and novel theoretical approaches have been achieved in the last decade, with a whole series of operational models in developmental robotics (Baldassarre & Mirolli, 2013; Oudeyer & Smith, 2016; Oudeyer et al., 2007; Twomey & Westermann, 2018), arguments

Large parts of this chapter are adapted from Oudeyer and Smith (2016) and Oudeyer et al. (2016), with permission.

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

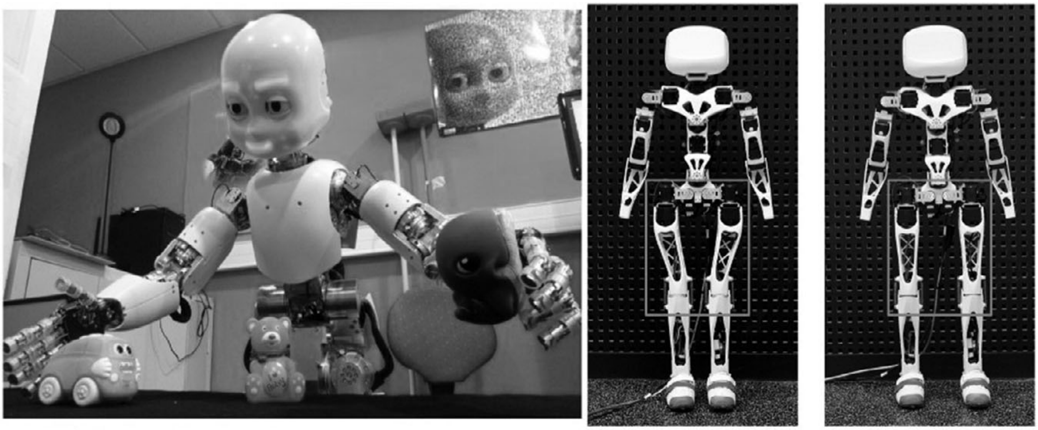
and models of the evolutionary origins of intrinsic motivation systems (Barto, 2013; Singh et al., 2010), and recent findings in neuroscience linking intrinsic motivation with attention (Gottlieb & Oudeyer, 2018), as well as new formal models of infant visual attention (Kidd et al., 2012).

A key idea in these recent approaches is that *learning progress in and for itself* can generate intrinsic rewards driving such spontaneous exploration, leading learners to avoid learning situations that are either too easy or too difficult at a given point of their development. Learning progress refers to the infant's *improvement* of its predictions or control over an activity they practice (Kaplan & Oudeyer, 2007a, 2007b), which can also be described as *reduction* of uncertainty (Friston et al., 2012). Such intrinsically motivating activities have been called "progress niches" (Oudeyer et al., 2007). Thus, learning progress is not simply a consequence of intrinsically motivated exploration, but a primary driver (and accordingly, intrinsic rewards for learning progress/uncertainty reduction may be *primary* rewards). From a machine learning perspective, such a mechanism of information seeking is called "active learning," where the learner probabilistically selects experiments according to their potential for reducing uncertainty.

These advances lead to a definition of curiosity as an epistemic motivational mechanism, which pushes an organism to explore activities for the primary sake of *gaining information* (as opposed to searching for information in service of achieving an external goal, like finding food or shelter). Such a motivational mechanism of curiosity will often be only one of several motivational mechanisms operating in any living being, and, at any given time, curiosity may interact, complement, or conflict with other motivations. From a machine learning perspective, mechanisms of information seeking are called *active learning*, where the learner probabilistically selects experiences according to their potential for reducing uncertainty (Cohn et al., 1996). They have been used either as an "exploration bonus" mechanism in service of efficient maximization of a task-specific reward, or as primary rewards driving models of curiosity-driven learning (Gottlieb et al., 2013).

As such, active self-exploration bi-directionally interacts with learning, strongly influencing what skills the infant will practice and eventually acquire, and it is bound to have a significant impact on the ordering and organization of development. In fact, through the analysis of robotic experiments in Section 31.3, we argue that such mechanism can be an essential force in the self-organization of developmental structures, and be a pillar of a principled dynamic systems approach to developmental change. Indeed, as we will show, active search for learning progress automatically leads a system to first explore simple activities, and progressively shift to more complex learning experiences, effectively self-generating a learning curriculum adapted to the current constraints, and at the same time itself constraining learning.

The key theoretical idea instantiated in these robotic models is epigenesis (see Chapter 2), in the sense proposed by Gottlieb (1991). Developmental structures in these models are neither learned from "tabula rasa" nor a predetermined result of an innate "program": they self-organize out of the dynamic interaction between constrained cognitive mechanisms (including curiosity, learning, and abstraction), the morphological properties of the body, and the physical and social environment that itself is constrained and ordered by the developmental level of the organism (Oudeyer, 2011; Thelen & Smith, 1996). This self-organization includes the dynamic and automatic formation of behavioral and cognitive stages of progressively increasing complexity, sharing many properties with infant development (Piaget, 1952; Houdé, 2015, for a more recent review).



(a) bended thighs (b) straight thighs

Figure 31.1 Open-source baby robots. Robots can help us model and study the complex interaction between the brain, the body, and the environment during cognitive development. Here we see two open-source robotic platforms used in laboratories. Being open-source allows open science through revealing all details in the experiments as well as replicability. Based on 3D printing, the Poppy platform allows fast and efficient exploration of various body morphologies (Lapeyre et al., 2014), such as leg shape (see alternatives on the right (a) and (b)), and how this can affect development of skills. Left: ICub; www.icub.org, right: Poppy www.poppy-project.org.[1]
A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.

This complex dynamical systems perspective on child development has fueled the use of computational and robotic modeling tools in the field of developmental robotics to design formal mechanistic theories of cognitive development, as well as enable scientists to sharpen their intuitions of these processes and propose new hypotheses (Cangelosi & Schlesinger, 2015, see Figure 31.1). Models of curiosity-driven learning discussed in this category have been designed from this perspective.

31.2 Models of Curiosity-Driven Learning

Diverse theoretical approaches to intrinsic motivation and to the properties that shall make certain activities intrinsically interesting/motivating have been proposed and published by diverse research communities within psychology (e.g., Csikszentmihalyi, 1991; Festinger, 1957; Harlow, 1950; Hull, 1943; Hunt, 1965; Kagan, 1972; Lowenstein, 1994; Montgomery, 1954; Ryan & Deci, 2000; White, 1959) and neuroscience (see Gottlieb & Oudeyer, 2018 for a review), in such a way that still today there is no consensus among these communities on a unified or integrated view of intrinsic motivation and curiosity. Yet, a convincing integrated view was actually proposed in the 1960s by Daniel Berlyne (Berlyne, 1965), and has been used as a fruitful theoretical reference for developing formal mathematical models of curiosity. The central concept of this integrated approach to intrinsic motivation is that of "collative variables," as explained in the following quotations:

[1] Both websites last accessed August 23, 2021.

> The probability and direction of specific exploratory responses can apparently be influenced by many properties of external stimulation, as well as by many intraorganism variables. They can, no doubt, be influenced by stimulus intensity, color, pitch, and association with biological gratification and punishment, ... [but] the paramount determinants of specific exploration are, however, a group of stimulus properties to which we commonly refer by such words as "novelty," "change," "surprisingness," "incongruity," "complexity," "ambiguity," and "indistinctiveness." (Berlyne, 1965, p. 245)

More precisely, Berlyne developed the notion that the most rewarding situations were those with an intermediate level of novelty, between already familiar and completely new situations (Berlyne, 1960). This perspective was echoed by Csikszentmihalyi's flow theory arguing that interesting situations were those of optimal challenge, recently echoed by Kidd et al. (2012), who showed an experiment where infants preferred stimuli of intermediate complexity.

31.2.1 The Learning Progress Hypothesis

Berlyne's concept of intermediate novelty, as well as the related concept of intermediate challenge of Csikszentmihalyi, have the advantage of allowing intuitive explanations of many behavioral manifestations of curiosity and intrinsic motivation. However, recent developments in theory of curiosity, and in particular its computational theory, have questioned its applicability as an operant concept capable of generating an actual mechanism for curiosity. A first reason is that the concept of "intermediate" appears difficult to define precisely, as it implies the use of a relatively arbitrary frame of reference to assess levels of novelty/complexity. A second reason is that while novelty or complexity in themselves may be the basis of useful exploration heuristics for organisms in some particular contexts, there is in general no guarantee that observing a novel of complex stimulus (be it of intermediate level) provides information that can improve the organism's prediction and control in the world. Indeed, as computational theory of learning and exploration has shown, our environment is full of novel and complex stimuli of all levels, and among them only a few may be associated with actual useful and learnable patterns. As curiosity-driven spontaneous exploration may have evolved as a means to acquire information and skills in rapidly changing environments (Barto, 2013), it appears that heuristics based on maximizing novelty and complexity are suboptimal (Oudeyer et al., 2007; Schmidhuber, 1991).

For these reasons, computational learning theory has explored an alternative mechanism, in which learning progress generates itself intrinsic reward (Oudeyer et al., 2007; Schmidhuber, 1991), and it was hypothesized that this mechanism could be at play in humans and animals (Kaplan & Oudeyer, 2007a; Oudeyer & Smith, 2016). This hypothesis proposes that the brain, seen as a predictive machine constantly trying to anticipate what will happen next, is intrinsically motivated by exploring activities in which predictions are improving (i.e., where uncertainty is decreasing, and learning is actually happening). This means that what is of interest is neither activities which are too easy or too difficult to predict (i.e., where uncertainty is low or where uncertainty is high but not reducible), but activities just beyond the current predictive capacities. So, for example, an infant will be more interested in exploring how its arm motor commands can allow it to predict the movement of its hand in its visual field (initially difficult but learnable) rather than predicting the movement of walls (too easy) or the color of the next car passing through the window (always novel, but not learnable).

As shown by computational studies, this leads in practice systems to the exploration of activities and stimuli of apparently "intermediate complexity." However, this notion of intermediate level is not directly represented in the mechanism: rather, it is a side-effect of selecting actions and stimuli that maximize the derivative of errors in prediction. Furthermore, this concept allows us to bridge several hypotheses related to curiosity and intrinsic motivation, but remains conceptually separated so far.

First, within the learning progress hypothesis, the central concept of prediction errors (and the associated measure of improvement) applies to multiple kinds of predictions. It applies to predicting the properties of external perceptual stimuli (and thus relates to the notion of perceptual curiosity, Berlyne, 1960), as well as the conceptual relations among symbolic items of knowledge (and this relates to the notion of epistemic curiosity, and to the subjective notion of information gap proposed by Lowenstein, 1994). Here the maximization of learning progress leads to behaviors that were previously understood through Berlyne's concept of intermediate novelty/complexity. It also applies to predicting the consequences of one's own actions in particular activities, or to predicting how well one's current skills are capable to solve a given goal/problem: here the maximization of learning progress, measuring a form of progress in competences related to an activity or a goal, allows to reframe Csikszentmihalyi's concept of intermediate challenge in the flow theory as well as related theories of intrinsic motivation based on self-measures of competences (Csikszentmihalyi, 1991; White, 1959).

Second, the learning progress hypothesis allows us to create a new causal link between memory retention and curiosity. As argued, experimental work showed that state curiosity could facilitate memory retention (Kang et al., 2009; Stahl & Feigenson, 2015). These results showed that curiosity and prediction errors had an influence on learning, but curiosity and learning were considered as two separate mechanisms (and indeed, seeing curiosity as search for pure novelty makes it separate from actual learning mechanisms). Furthermore, the experimental protocols were such that novelty/surprise was always imposed by experimenters, with little possibility for subjects to actively seek and explore their environment (and thus limiting the possibility to study the processes by which organisms would encounter such novelty/complexity in a more ecological situation).

The learning progress hypothesis provides a strong complement to this view: it proposes that experiencing learning in a given activity (rather than just intermediate novelty) triggers an intrinsic reward, and thus that learning in itself causally participates to establish state curiosity. Thus, this hypothesis argues that there is a closed self-reinforcing feedback loop between learning and curiosity-driven intrinsic motivation. Here the learner becomes fundamentally active, searching for niches of learning progress, in which in turn memory retention is facilitated. In Section 31.3, we will outline computational experiments that have shown that such an active learning mechanisms can self-organize a progression in learning, with automatically generated developmental phases that have strong similarities with infant developmental trajectories.

31.3 The Playground Experiment: A Developmental Robotics Model of Curiosity-Driven Self-organization of Developmental Trajectories

In this section, we describe a series of robot experiments that illustrate how mechanisms of curiosity-driven exploration, dynamically interacting with learning, physical, and social constraints, can self-organize developmental trajectories and in particular lead a learner to discover successively object affordances and vocal interaction with its peers.

In these Playground Experiments, a quadruped "learning" robot is placed on an infant play mat with a set of nearby objects and is joined by an "adult" robot peer, see Figure 31.2(a) (Kaplan & Oudeyer, 2007b; Oudeyer & Kaplan, 2006; Oudeyer et al., 2007). On the mat and near the learning robot are objects for discovery: an elephant (which can be bitten or "grasped" by the mouth), and a hanging toy (which can be "bashed" or pushed with the leg). The adult robot peer is preprogramed to imitate the sound made by the learning robot when the learning robot looks to the adult while vocalizing at the same time.

The learning robot is equipped with a repertoire of motor primitives parameterized by several continuous numbers that control movements of its legs, head, and a simulated vocal tract. Each motor primitive is a dynamical system controlling various forms of actions: (a) turning the head in various

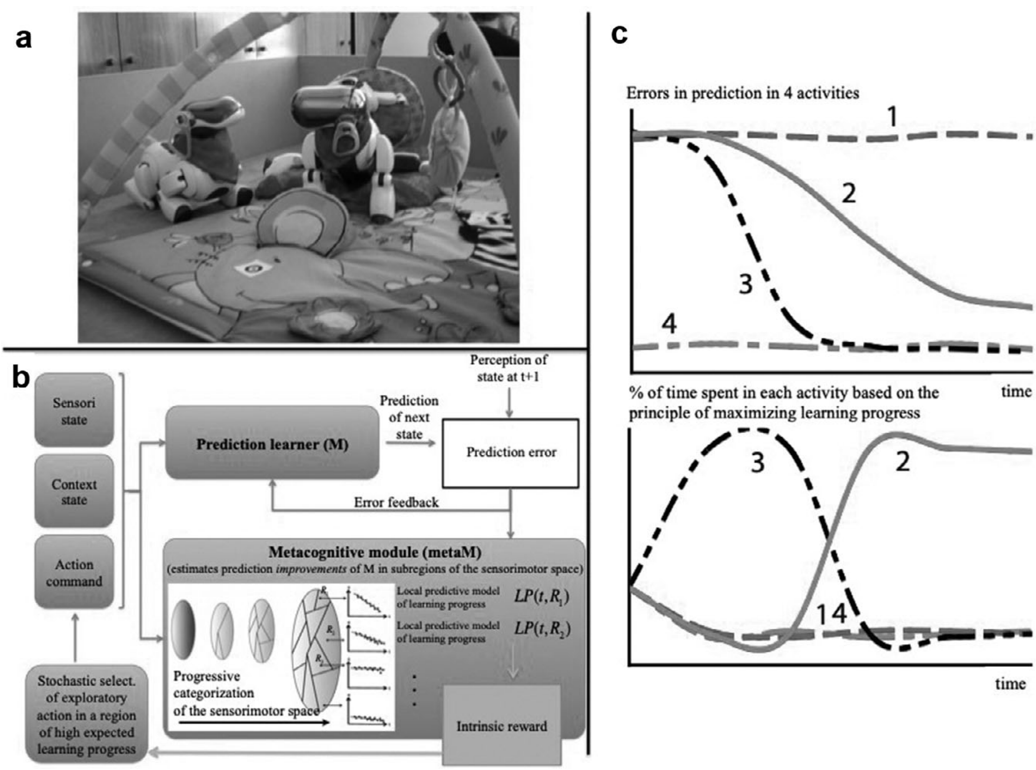


Figure 31.2 The playground experiment (Oudeyer & Kaplan, 2006; Oudeyer et al., 2007). (a) The learning context; (b) The computational architecture for curiosity-driven exploration in which the robot learner probabilistically selects experiences according to their potential for reducing uncertainty, that is, for learning progress; (c) Illustration of a self-organized developmental sequence where the robot automatically identifies, categorizes, and shifts from simple to more complex learning experiences. Figure adapted with permission from Gottlieb et al. (2013)

A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

directions, (b) opening and closing the mouth while crouching with various strengths and timing, (c) rocking the leg with various angles and speed, and (d) vocalizing with various pitches and lengths. These primitives can be combined to form a large continuous space of possible actions. Similarly, several kinds of sensory primitives allow the robot to detect visual movement, salient visual properties, proprioceptive touch in the mouth, and pitch and length of perceived sounds. For the robot, these motor and sensory primitives are initially black boxes and he has no knowledge about their semantics, effects, or relations.

The robot learns how to use and tune these primitives to produce various effects on its surrounding environment, and exploration is driven by the maximization of learning *progress*, that is, by choosing physical experiences ("experiments") that improve the quality of predictions of the consequences of its actions.

Figure 31.2(b) outlines a computational architecture, called Robust Intelligent Adaptive Curiosity (R-IAC), that makes possible this curiosity-driven exploration and learning process (Moulin-Frier et al., 2014; Oudeyer et al., 2007). It is composed of several modules. A prediction machine (M) learns to predict the consequences of actions taken by the robot in given sensory contexts. For example, this module might learn to predict (with a neural network) which visual movements or proprioceptive perceptions result from using the leg bashing motor primitive with certain parameters. A meta-cognitive module estimates the evolution of errors in prediction of M in various subregions of the sensorimotor space. This module estimates how much errors decrease in predicting an action, for example, in predicting the consequence of a leg bashing movement when this action is applied toward a particular area of the environment. These estimates of error reduction are used to compute learning progress as an intrinsic reward. This reward is an internal quantity that is proportional to the decrease of prediction errors. The maximization of such reward is the aim of action selection in a computational reinforcement learning architecture (Oudeyer et al., 2007). Importantly, the action selection system chooses most often to explore activities where estimated learning progress is high. However, this choice is probabilistic, which leaves the system open to learning in new areas and open to discovering other activities which might yield learning progress.[2] Since the sensorimotor flow does not come pre-segmented into activities and tasks, a system that seeks to maximize differences in learnability is also used to progressively categorize the sensorimotor space into regions, which models the incremental creation and refining of cognitive categories differentiating activities/tasks.

To illustrate how such an exploration mechanism can automatically generate ordered learning stages, let us first imagine a robot confronted with four categories of activities, as shown in Figure 31.2(c). The practice of each of these four activities, which can be of varying difficulty, leads to different learning rates at different points in time (see the top curves, which show the evolution of prediction errors in each activity if the robot would focus full-time on each). If the robot uses curiosity-driven exploration to decide what and when to practice by focusing on progress niches, it will avoid activities which are already predictable (curve 4) or too difficult to learn to predict (curve 1), in order to focus first on the activity with the fastest learning rate (curve 3) and, eventually, when the latter starts to reach a

[2] Technically, the decision on how much time to spend on high learning progress activities and other activities is achieved using Multi-Armed Bandit algorithms for the so-called exploration/exploitation dilemma (Audibert et al., 2009).

"plateau" to switch to the second most promising learning situation (curve 2). Thus, such robots will show a regular developmental course – one that will be "universal" for learners with similar internal processes learning in similar environments. Embodied exploration driven by learning progress creates an organized exploratory strategy: the system systematically achieves these learning experiences in an order and does so because they yield (given the propensities of the learner and the physical world) different patterns of uncertainty reduction.

In the Playground experiment described, multiple experimental runs lead to two general categories of results: self-organization and a mixture of regularities and diversities in the developmental patterns (Oudeyer & Kaplan, 2006; Oudeyer et al., 2007).

31.3.1 Self-organization

In all of the runs, one observes the self-organization of structured developmental trajectories, where the robot explores objects and actions in a progressively more complex stage-like manner while acquiring autonomously diverse affordances and skills that can be reused later on and that change the learning progress in more complicated tasks. The following developmental sequence is typically observed:

1. In a first phase, the robot achieves unorganized body babbling.
2. In a second phase, after learning a first rough model and meta-model, the robot stops combining motor primitives, exploring them one by one, but each primitive is explored itself in a random manner.
3. In a third phase, the robot now begins to experiment with actions toward zones of its environment where the external observer knows there are objects (the robot is not provided with a representation of the concept of "object"), but in a non-affordant manner (e.g., it vocalizes at the non-responding elephant or tries to bash the adult robot which is too far to be touched).
4. In a fourth phase, the robot now explores affordant experiments: it first focuses on grasping movements with the elephant, then shifts to bashing movements with the hanging toy, and finally shifts to exploring vocalizing toward the imitating adult robot.
5. In the end, the robot has learnt sensorimotor affordances with several objects, as well as social affordances with a peer, and mastered multiple skills, yet none of these specific objectives were preprogramed in the beginning. They self-organize through the dynamic interaction between curiosity-driven exploration, statistical inference, the properties of the body, and the properties of the environment.

These playground experiments do not simply simulate particular skills (such as batting at toys to make them swing or vocalizations) but simulate an ordered and systematic developmental trajectory, with a universality and stage-like structure that may be mistakenly taken to indicate an internally-driven process of maturation. However, the trajectory is created through activity and through the general principle that sensorimotor experiences that reduce uncertainty in prediction are rewarding. In this way, developmental achievements can build on themselves without specific pre-programmed dependencies, but nonetheless – like evolution itself –create structure (see Smith & Breazeal, 2007 and Smith, 2013, for related findings and arguments).

31.3.2 Regularities and Diversity

Because these are self-organizing developmental processes, they generate not only strong

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

regularities but also diversity across individual developmental trajectories. For example, in most runs one observes successively unorganized body babbling, then focused exploration of head movements, then exploration of touching an object, then grasping an object, and finally vocalizing toward a peer robot (preprogrammed to imitate). This can be explained as gradual exploration of new progress niches, and those stages and their ordering can be viewed as a form of attractor in the space of developmental trajectories. Yet, with the same mechanism and same initial parameters, individual trajectories may invert stages, or even generate qualitatively different behaviors. This is due, stochasticity, to even small variability in the physical realities and to the fact that this developmental dynamic system has several attractors with more or less extended and strong domains of attraction (characterized by amplitude of learning progress). We see this diversity as a positive outcome since individual development is not identical but always unique in its own ways. This kind of approach, then, offers a way to understanding individual differences as emergent in the developmental process itself and makes clear how the developmental process might vary across contexts, even with an identical mechanism.

A further result to be highlighted is the early development of vocal interaction. With a single generic mechanism, the robot both explores and learns how to manipulate objects and how to vocalize to trigger specific responses from a more mature partner. While vocal babbling, and more language play and games, have been shown to be key in infant language development, the interest of infants to engage in babbling and such language games has often been associated with an ad hoc language-specific motivation. The Playground Experiment makes it possible to see how the exploration and learning of communicative behavior might be at least partially explained by general curiosity-driven exploration of the body affordances, as also suggested by Oller (2000).

Further models have explored more specifically how social guidance provided by social peers can be leveraged by an intrinsically motivated active learner and dynamically interact with curiosity to structure developmental trajectories (Nguyen & Oudeyer, 2013; Thomaz & Breazeal, 2008). Focusing on vocal development, Moulin-Frier et al. (2014) conducted experiments where a robot explores the control of a realistic model of the vocal tract in interaction with vocal peers, and driven to maximize learning progress. This model relies on a physical model of the vocal tract, its motor control, and of the auditory system. Experiments showed how such a mechanism can explain the adaptive transition from vocal self-exploration with little influence from the speech environment, to a later stage where vocal exploration becomes influenced by vocalizations of peers. Within the initial self-exploration phase, a sequence of vocal production stages self-organizes, and shares properties with infant data: the vocal learner first discovers how to control phonation, then focuses on vocal variations of unarticulated sounds, and finally automatically discovers and focuses on babbling with articulated proto-syllables. As the vocal learner becomes more proficient at producing complex sounds, imitating vocalizations of peers starts to provide high learning progress explaining an automatic shift from self-exploration to vocal imitation.

31.4 Curiosity-Driven Learning in Artificial Intelligence (AI) and Machine Learning

The AI and machine learning literature has also shown how various forms of intrinsically motivated exploration and learning could

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

guide efficiently the autonomous acquisition of repertoires of skills in large and difficult spaces, providing a perspective in which to interpret the evolution of curiosity-driven learning in living organisms.

A first reason is that intrinsically motivated exploration can be used as an active learning algorithm that learns efficient forward and inverse models of the world dynamics through efficient selection of experiences. Indeed, such models can be reused either directly (Baranes & Oudeyer, 2013; Oudeyer et al., 2007), or through model-based planning mechanisms (Lopes et al., 2012; Schmidhuber, 1991; Singh et al., 2004), to solve repertoires of tasks that were not specified during exploration (hence without the need for long re-experiencing of the world for each new task). For example, Baranes and Oudeyer (2013) have shown how intrinsically motivated goal exploration could allow robots to sample sensorimotor spaces by actively controlling the complexity of explored sensorimotor goals and avoiding goals which were either too easy or unreachable. This allowed the robots to learn fast repertoires of high-dimensional continuous action skills to solve distributions of sensorimotor problems such as omnidirectional-legged locomotion or how to manipulate flexible objects. Lopes et al. (2012) showed how intrinsically motivated model-based reinforcement learning, driven by the maximization of empirical learning progress, allows efficient learning of world models when this dynamics is non-stationary, and how this accelerates the learning of a policy that targets to maximize an extrinsic reward (task predefined by experimenters).

A second reason for the efficiency of intrinsic motivation is that by fostering spontaneous exploration of novel skills, and leveraging opportunistically potential synergies among skills, it can create learning pathways toward certain skills that would have remained difficult to reach if they had been the sole target of the learning system. Indeed, in many contexts, learning a single pre-defined skill can be difficult as it amounts to searching (the parameters of) a solution with very rare feedback until one is very close to the solution, or with deceptive feedback due to the phenomenon of local minima. A strategy to address these issues is to direct exploration with intrinsic rewards, leading the system to explore a diversity of skills and contingencies which often result in the discovery of new sub-spaces/areas in the problem space, or in mutual skill improvement when exploring one goal/skill provides data that can be used to improve other goals/skills, such as in goal babbling (Baranes & Oudeyer, 2013; Benureau & Oudeyer, 2016) or off-policy reinforcement learning (see the Horde architecture, Sutton et al., 2011). For example, Lehman and Stanley (2011) showed that searching for pure novelty in the behavioral space a robot finds a reward in a maze more efficiently than if it had been searching for behavioral parameters that directly optimized the reward. In another model, Forestier and Oudeyer (2016) showed that intrinsically motivated exploration of a hierarchy of sensorimotor models allowed a simulated robot to scaffold the successive acquisition of object reaching, tool grasping, and tool use (note that behaviors aiming at direct search tool use were less efficient).

A third related reason for the efficiency of intrinsically motivated exploration is that it can drive the acquisition of macro-actions, or sensorimotor primitives, which can be combinatorially reused as building blocks to accelerate the search for complex solutions in structured reinforcement learning problems. For example, Singh et al. (2004) showed how intrinsic rewards based on measures of saliency could guide a reinforcement learner to progressively learn "options," which are temporally extended macro-actions, reshaping the structure of the search space and finally learning action policies that solve an extrinsic (abstract) task that is very difficult to solve through standard reinforcement learning exploration. Related uses of intrinsic motivation with a hierarchical reinforcement learning framework were demonstrated in Bakker and Schmidhuber (2004) and Kulkarni et al. (2016).

In a related line of research studying the function and origins of intrinsic motivation, Singh et al. (2010) have shown through computational modeling the potential evolutionary usefulness of intrinsic motivation systems for maximizing extrinsic rewards (e.g., quantity of food collected) in a distribution of changing environments. In such changing environments, it could be more robust for reinforcement learning agents to represent and use an internal reward function that does not directly correspond to this extrinsic reward, but rather includes a component of intrinsic motivation that pushes the system to explore its environment beyond the direct search for the extrinsic reward.

31.5 Applications of Models of Curiosity-Driven Learning in Educational Technologies

Given the strong causal interactions between curiosity-driven exploration and learning that we just reviewed, these topics have attracted the attention of theorists and experimenters on the application domain of education. Long before recent controlled experimental results showing how intrinsic motivation and curiosity could enhance learning, educational experimenters like Montessori (1948) and Froebel (1885) studied how open-ended learning environments could foster individual child development, where learners are active, and where the tutor's role is to scaffold challenges of increasing complexity and provide feedback (rather than instruction). Such experimental approaches have more recently influenced the development of hands on educational practices, such as the pioneering LOGO experiments of Papert (1980), where children learn advanced concepts of mathematics, computer science, and robotics, and are now disseminating on large scales in several countries (Resnick et al., 2009; Roy et al., 2015).

In parallel, philosophers and psychologists like Dewey, Vygotski, Piaget, and Bruner developed theories of cognitive and social constructivist learning which directly pointed toward the importance of fostering curiosity, free play, and exploration in the classroom. Recently, the large body of research in educational psychology has led others to study systematically how states of intrinsic motivation can be fostered, or on the contrary weakened, in the classroom, for example when the educational context provides strong extrinsic rewards (Deci et al., 2001).

As educational technologies are now thriving, in particular with the wide spreading of Massive Open Online Courses (MOOCs) and educational applications on tablets and smartphones, it has become natural to enquire how fundamental understanding of curiosity, intrinsic motivation, and learning could be leveraged and incorporated in these educational tools to increase their efficiency.

A first line of investigation has been to embed educational training within motivating and playful video games. In a pioneer study, Malone (1980) used and refined theories of intrinsic motivation as proposed by Berlyne, White and psychologists of the 1950–1970s period, to evaluate which properties of video games could make them intrinsically motivating, and to study how such contexts could be used to distill elements of scholarly knowledge to children. In particular, he showed that video games were more intrinsically motivating when including clear goals of progressively increasing complexity, when the system

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

provided clear feedback on the performance of users, and when outcomes were uncertain to entertain curiosity. For example, he showed how arithmetic concepts could be taught in an intrinsically motivating scenarized dart video game. As an outcome of their studies, they could generate a set of guidelines for the design of education-oriented video games.

In a similar study, studying the impact of several of the factors identified by Malone, Cordova and Lepper (1996) presented a study of a population of elementary school children using a game targeting the acquisition of arithmetic order-of-operation rules, scenarized in a "space quest" story. In this specific experimental context, they showed that embedding personalization in the math exercises (based on preferences expressed through a pre-questionnaire) significantly improved intrinsic motivation, task engagement, and learning efficiency, and that this effect was heightened if in addition the software offered personalization of visual displays and a variety of exercise levels children could choose from.

Beyond explicitly including educational elements in video games, it was also shown that "pure" entertainment games such as certain types of action games can enhance attentional control, cognitive flexibility, and learning capabilities by exercising them in an intrinsically motivating playful context (Cardoso-Leite & Bavelier, 2014). Within this perspective, Merrick and Maher (2009) suggested that implementing artificial curiosity in non-player characters in video games could enhance the "interestingness" of video games.

A second line of investigation has considered how formal and computational models of curiosity and intrinsic motivation could be applied to Intelligent Tutoring Systems (ITS) (Nkambou et al., 2010), as well as Massive Open Online Courses (MOOCS) (Liyanagunawardena et al., 2013). ITS, and more recently MOOCs, have targeted the design of software systems that could help students acquire new knowledge and skills, using artificial intelligence techniques to personalize teaching sequences, or the way teaching material is presented, and in particular proposing exercises that match the particular difficulties or talents of each individual learner. In this context, several approaches were designed and experimented on, so as to promote intrinsic motivation and learning.

Clement et al. (2015) have presented and evaluated an ITS system that directly reused computational models of curiosity-driven learning based on the learning progress hypothesis described in Section 31.3 (Oudeyer et al., 2007). This study considered teaching arithmetic decomposition of integer and decimal numbers, in a scenarized context of money handling, to a population of seven-to-eight years old children (see Figure 31.3). To design the ITS system, a human teacher first provided pedagogical material in the form of exercises grouped along coarsely defined levels and coarsely defined types. Then, an algorithm called ZPDES (Zone of Proximal Development and Empirical Success) was used to automatically personalize the sequence of exercises for each student, and this personalization was made incrementally during the course of interaction with each student. This personalization was achieved by probabilistically proposing to students exercises that maximized learning progress at their current level, that is, the exercises where their errors decrease fastest. In order to dynamically identify these exercises, and shift automatically to new ones when learning progress becomes low, the system used a multi-armed bandit algorithm that balanced exploring new exercises to assess their potential for learning progress, and exploiting exercises that recently lead the student to learning progress. During this process, the coarse structure organizing exercises that was provided by a human teacher is used

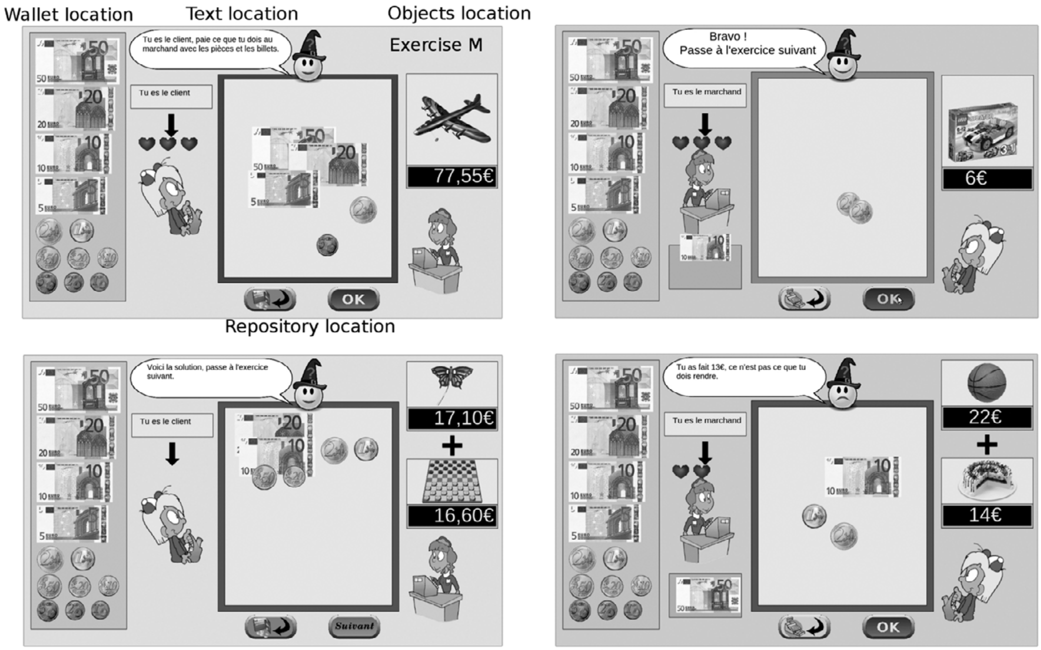


Figure 31.3 Educational game used in Clement et al. (2015): a scenario where elementary school children have to learn to manipulate money is used to teach them the decomposition of integer and decimal numbers. Four principal regions are defined in the graphical interface. The first is the wallet location where users can pick and drag the money items and drop them on the repository location to compose the correct price. The object and the price are present in the object location. Four different types of exercises exist: M (customer/one object), R (merchant/one object), MM (customer/two objects), RM (merchant/two objects). The intelligent tutoring system then dynamically proposes to students the exercises in which they are currently making maximal learning progress, targeting to maximize intrinsic motivation and learning efficiency

A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.

to guide the algorithm toward finding fast which exercises provide maximal learning progress: the system starts with exercise types that are at the bottom of the difficulty hierarchy, and when some of them show a plateau in the learning curve, they are deactivated and new exercises higher in the hierarchy are made available to the student (see Figure 31.4). The use of learning progress as a measure to drive the selection of exercises had two interacting purposes, relying on the bidirectional interaction described above. First, it aimed to propose exercises that could stimulate the intrinsic motivation of students by dynamically and continuously setting them challenges that were neither too difficult nor too easy. Second, by doing this using learning progress, it aimed at generating exercise sequences that are highly efficient for maximizing the average scores over all types of exercises at the end of the training session. Indeed, Lopes and Oudeyer (2012) showed in a theoretical study that, when faced with the problem of strategically choosing which topic/exercise type to work on, selecting topics/exercises that maximize learning progress is quasi-optimal for important

https://doi.org/10.1017/9781108399838.036 Published online by Cambridge University Press

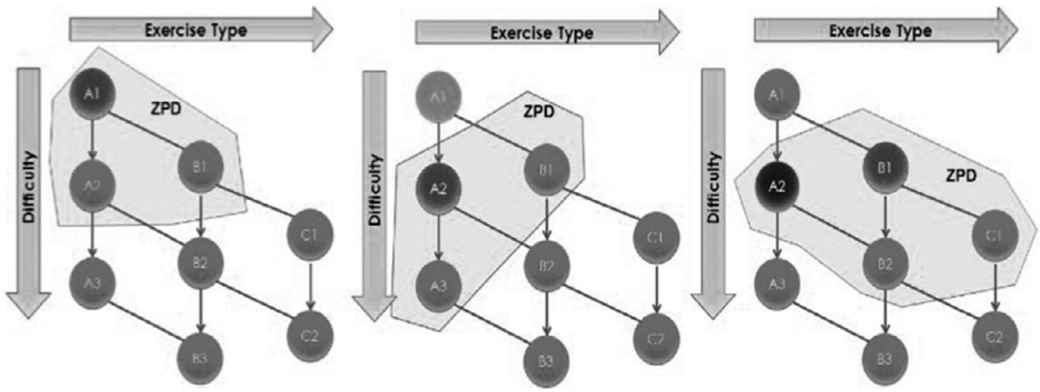


Figure 31.4 Example of the evolution of the zone-of-proximal development (ZPD) based on the empirical results of the student. The ZPD is the set of all activities that can be selected by the algorithm. The expert defines a set of preconditions between some of the activities (A1 → A2 → A3 . . .), and activities that are qualitatively equal (A == B). Upon successfully solving A1 the ZPD is increased to include A3. When A2 does not achieve any progress, the ZPD is enlarged to include another exercise type C, not necessarily of higher or lower difficulty, for example, using a different modality, and A3 is temporarily removed from the ZPD. Adapted from Clement et al. (2015)
A black-and-white version of this figure will appear in some formats. For the color version, please refer to the plate section.

classes of learner models. Experiments with 400 children from 11 schools were performed, and the impact of this algorithm selecting exercises that maximize learning progress was compared to the impact of a sequence of exercises hand-defined by an expert teacher (that included sophisticated branching structures based on the errors-repair strategies the teacher could imagine). Results showed that the ZPDES algorithm, maximizing learning progress, allowed students of all levels to reach higher levels of exercises. Also, an analysis of the degree of personalization showed that ZPDES proposed a higher diversity of exercises earlier in the training sessions. Finally, a pre- and post-test comparison showed that students who were trained by ZPDES progressed better than students who used a hand-defined teaching sequence.

Several related ITS systems were developed and tested. For example, Beuls (2013) described a system targeting the acquisition of Spanish verb conjugation, where the ITS attempts to propose exercises that are just above the current capabilities of the learner. Recently, a variation of this system was designed to foster the learning of musical counterpoint (Beuls & Loeckx, 2015). In another earlier study, Pachet (2004) presented a computer system aiming to help children discover and learn how to play musical instruments, but also capable of supporting creativity in experienced musicians, through fostering the psychological experience of Flow (Csikszenthmihalyi, 1991). This system, called the Continuator (Pachet, 2004), continuously learnt the style of the player (be it a child beginner or expert) and used an automatic improvisation algorithm to respond to the user's musical phrases with musical phrases of the same style and complexity, but different from those actually played by users. Pachet (2004) observed that both children and expert musicians most often experience an "Eureka moment" using this system. Their interest and attention appeared to be strongly attracted by playing with the system, leading children to try and discover different modes of play and to increase the complexity of what they could do. Expert musicians also reported that the system allowed them to discover novel musical ideas and to support creation interactively.

31.6 Conclusion

Computational and robotic models of curiosity-driven exploration and learning have enabled us to formalize various forms of spontaneous exploration mechanisms that had remained elusive in psychology and neuroscience so far. These models have considered curiosity mechanisms from a perspective where the child is viewed as a sense-making organism that explores for the sake of building good predictive models of its world. This perspective has been instrumental in identifying which curiosity mechanisms could guide efficient exploration and learning in large uncontrolled real-world environments. For example, this has allowed us to see the limits of mechanisms pushing organisms to search for high uncertainty or high entropy states (doing this they could be trapped exploring parts of the environments where there is nothing useful to learn). This has also led us to formulate the learning progress (LP) hypothesis, stating that organisms explore situations where they empirically improve various aspects of their world models. Computational and robotic experiments of this LP mechanism have shown that beyond enabling efficient learning of world models in large sensorimotor and cognitive spaces, it also enabled us to self-organize ordered trajectories of development. These self-organized trajectories emerge out of the dynamic interaction between the brain, the body, and the environment, and show similar regularities and diversity to human infant developmental trajectories. Finally, these new theoretical insights provide a new view on educational perspectives, and have been adapted in educational technology applications, aiming at personalizing the sequences of exercises provided to each learner, in order to maximize both the efficiency of learning and intrinsic motivation.

References

Audibert, J.-Y., Munos, R., & Szepesvari, C. (2009). Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, *410*, 1876–1902.

Bakker, B., & Schmidhuber, J. (2004). Hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization. *Proceedings of the 8th Conference on Intelligent Autonomous Systems* (pp. 438–445). Amsterdam, Netherlands.

Baldassare, G., Mirolli, M. (2013). *Intrinsically Motivated Learning in Natural and Artificial Systems*. Berlin: Springer-Verlag.

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

Barto, A. G. (2013). Intrinsic motivation and reinforcement learning. In G. Baldassarre, & M. Mirolli (eds.), *Intrinsically Motivated Learning in Natural and Artificial Systems* (pp. 17–47). Berlin: Springer.

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

Berlyne, D. (1960). *Conflict, Arousal and Curiosity*. New York: McGraw-Hill.

Berlyne, D. (1965). *Structure and Direction in Thinking*. New York: John Wiley and Sons, Inc.

Beuls, K. (2013). Towards an Agent-based Tutoring System for Spanish Verb Conjugation. PhD thesis, Vrije Universiteit, Brussel.

Beuls, K., & Loeckx, J. (2015). Steps towards intelligent MOOCs: A case study for learning counterpoint. In L. Steels (ed.), *Music Learning with Massive Open Online Courses* (pp. 119–144). Amsterdam: IOS Press.

Cangelosi, A., & Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press.

Cardoso-Leite, P., & Bavelier, D. (2014). Video game play, attention, and learning: How to shape the development of attention and influence learning? *Current Opinion in Neurology*, *27*, 185–191.

Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. *Journal of Educational Data Mining (JEDM)*, *7*, 20–48.

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, *88*, 715.

Csikszenthmihalyi, M. (1991). *Flow-the Psychology of Optimal Experience*. New York: Harper Perennial.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*, 1–27.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Forestier, S., & Oudeyer, P.-Y. (2016). Curiosity-driven development of tool use precursors: A computational model. *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (CogSci 2016) (pp. 1859–1864). August 2016, Philadelphia, PA.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, *3*, 151.

Froebel, F. (1885). *The Education of Man*. New York: A. Lovell & Company.

Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. *Developmental Psychology*, *27*, 4–13.

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

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and neural mechanisms. *Trends in Cognitive Science*, *17*, 585–596.

Harlow, H. (1950). Learning and satiation of response in intrinsically motivated complex puzzle performances by monkeys. *Journal of Comparative and Physiological Psychology*, *43*, 289–294.

Houdé, O. (2015). Cognitive development during infancy and early childhood across cultures. In J. D. Wright (ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., pp. 43–50). Oxford: Elsevier Science.

Hull, C. L. (1943). *Principles of Behavior: An Introduction to Behavior Theory*. New-York: Appleton-Century-Croft.

Hunt, J. M. (1965). Intrinsic motivation and its role in psychological development. *Nebraska Symposium on Motivation*, *13*, 189–282.

Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, *22*, 51–66.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*, 963–973.

Kaplan, F., & Oudeyer, P.-Y. (2007a). The progress-drive hypothesis: An interpretation of early imitation. In K. Dautenhahn, & C. Nehaniv (eds.), *Models and Mechanisms of Imitation and Social Learning: Behavioural, Social and Communication Dimensions* (pp. 361–377). Cambridge: Cambridge University Press.

Kaplan, F., & Oudeyer, P.-Y. (2007b). In search of the neural circuits of intrinsic motivation. *Frontiers in Neuroscience*, *1*, 225–236.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*, e36399.

Kulkarni, T. D., Narasimhan, K., Saeedi, A., & Tenenbaum, J. (2016). Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In M. I. Jordan, Y. LeCun, & S. A. Solla (eds.), *Advances in Neural Information Processing Systems* (pp. 3675–3683). Cambridge, MA: MIT Press.

Lapeyre, M., N'Guyen, S., Le Falher, A., & Oudeyer, P. Y. (2014). Rapid morphological exploration with the Poppy humanoid platform. 2014 IEEE-RAS International Conference on Humanoid Robots (pp. 959–966). IEEE. November 18–20, 2014, Madrid, Spain.

Lehman, J., & Stanley, K. O. (2011). Abandoning objectives: Evolution through the search for novelty alone. *Evolutionary Computation*, *19*, 189–223.

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning*, *14*, 202–227.

Lopes, M., Lang, T., Toussaint, M., & Oudeyer, P.-Y. (2012). Exploration in model-based reinforcement learning by empirically estimating learning progress. *Neural Information Processing Systems (NIPS 2012)*. December 3–8, 2012, Tahoe, USA.

Lopes, M., & Oudeyer, P. Y. (2012). The strategic student approach for life-long exploration and learning. 2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL) (pp. 1–8). IEEE. November 7–9, 2012, San Diego, CA.

Lowenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*, 75–98.

Malone, T. W. (1980). What makes things fun to learn? A study of intrinsically motivating computer games. *Technical report*. Xerox Palo Alto Research Center, Palo Alto, CA.

Merrick, K. E., & Maher, M. L. (2009). *Motivated Reinforcement Learning: Curious Characters for Multiuser Games*. Berlin: Springer Science & Business Media.

Montessori, M. (1948/2004). *The Discovery of the Child*. New Delhi: Aakar Books.

Montgomery, K. (1954). The role of exploratory drive in learning. *Journal of Comparative and Physiological Psychology*, *47*, 60–64.

Moulin-Frier, C., Nguyen, M., & Oudeyer, P.-Y. (2014). Self organization of early vocal development in infants and machines: The role of intrinsic motivation. *Frontiers in Cognitive Science*, *4*, 1006.

Nguyen, M., & Oudeyer, P.-Y. (2013). Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn, Journal of Behavioural Robotics*, *3*, 136–146.

Nkambou, R., Mizoguchi, R., & Bourdeay, J. (2010). *Advances in Intelligent Tutoring Systems* (Vol. 308). Berlin: Springer.

Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum and Associates, Inc.

Oudeyer, P.-Y. (2011). *Developmental Robotics, Encyclopedia of the Sciences of Learning*, ed. N. M. Seel. Springer Reference Series. Berlin: Springer.

Oudeyer, P.-Y., Gottlieb, J., & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In S. Waxman, D. G. Stein, D. Swaab, & H. Fields (eds.), *Progress in Brain Research* (Vol. 229, pp. 257–284). Amsterdam: Elsevier.

Oudeyer, P.-Y., & Kaplan, F. (2006). Discovering communication. *Connection Science*, *18*, 189–206.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, *11*, 265–286.

Oudeyer, P.-Y., & Smith, L. (2016). How evolution can work through curiosity-driven developmental process, *Topics in Cognitive Science*, *8*, 492–502.

Pachet, F. (2004). On the design of a musical flow machine. In M. Tokoro, & L. Steels (eds.), *A Learning Zone of One's Own* (pp. 111–134). Amsterdam: IOS Press.

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books, Inc.

Piaget, J. (1952) *The Origins of Intelligence in Children*. New York: International University Press.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., & Kafai, Y. (2009). Scratch: Programming for all. *Communications of the ACM*, *52*, 60–67.

Roy, D., Gerber, G., Magnenat, S., Riedo, F., Chevalier, M., Oudeyer, P. Y., & Mondada, F. (2015). IniRobot: A pedagogical kit to initiate children to concepts of robotics and computer science. *Proceedings of the RIE 2015*. May 20–22, 2015, Yverdon les bains, Switzerland.

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.

Schmidhuber, J. (1991). Curious model-building control systems. *IEEE International Joint Conference on Neural Networks* (pp. 1458–1463). 18-November 21, 1991, Singapore.

Singh, S., Barto, A. G., & Chentanez, N. (2004). Intrinsically motivated reinforcement learning. *NIPS'04: Proceedings of the 17th International Conference on Neural Information Processing Systems* (pp. 1281–1288). December 2004, Vancouver, BC, Canada.

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*, 70–82.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*, 618.

Smith, L. B., & Breazeal, C. (2007). The dynamic lift of developmental process. *Developmental Science*, *10*, 61–68.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*, 91–94.

Sutton, R. S., Modayil, J., Delp, M., Degris, T., Pilarski, P. M., White, A., & Precup, D. (2011). Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. *Proceedings of t*he 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011) (Vol. 2, pp. 761–768). May 2–6, 2011, Taipei, Taiwan.

Thelen, E. S., & Smith, L. B. (1996). *Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thomaz, A. L., & Breazeal, C. (2008) Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers. *Connection Science*, *20*, 91–110.

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

West, M. J., & King, A. P. (1987). Settling nature and nurture into an ontogenetic niche. *Developmental Psychobiology*, *20*, 549–562.

White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*, 297–333.