Intrinsic motivation, curiosity and learning: theory and applications in educational technologies

Pierre-Yves Oudeyer, Jacqueline Gottlieb, Manuel Lopes

Intrinsic motivation, curiosity and learning: theory and applications in educational technologies

Pierre-Yves Oudeyer (1), Jacqueline Gottlieb (2), Manuel Lopes (1)
(1) Inria and Ensta ParisTech, France
(2) Department of neuroscience, Kavli institute for brain science,
Columbia University, NY, US and

Abstract: *This article studies the bi-directional causal interactions between curiosity and learning, and discusses how understanding these interactions can be leveraged in educational technology applications. First, we review recent results showing how state curiosity, and more generally the experience of novelty and surprise, can enhance learning and memory retention. Then, we discuss how psychology and neuroscience have conceptualized curiosity and intrinsic motivation, studying how the brain can be intrinsically rewarded by novelty, complexity or other measures of information. We explain how the framework of computational reinforcement learning can be used to model such mechanisms of curiosity. Then, we discuss the learning progress (LP) hypothesis, which posits a positive feedback loop between curiosity and learning. We outline experiments with robots that show how LP-driven attention and exploration can self-organize a developmental learning curriculum scaffolding efficient acquisition of multiple skills/tasks.. Finally, we discuss recent work exploiting these conceptual and computational models in educational technologies, showing in particular how Intelligent Tutoring Systems can be designed to foster curiosity and learning.*

Keywords: curiosity; intrinsic motivation; learning; education; active learning; active teaching; neuroscience; computational modeling; artificial intelligence; educational technology.

Curiosity fosters learning and memory retention

Curiosity is a form of intrinsic motivation that is key in fostering active learning and spontaneous exploration. For this reason, curiosity-driven learning and intrinsic motivation have been argued to be fundamental ingredients for efficient education (Freeman et al., 2014). Thus, elaborating a fundamental understanding of the mechanisms of curiosity, and of which features of educational activities can make them "fun" and foster motivation, is of high-importance with regards to the educational challenges of the 21st century.

While there is not yet a scientific consensus on how to define curiosity operationally (Oudeyer and Kaplan, 2007; Gottlieb et al., 2013; Kidd and Hayden, 2015), states of curiosity are often associated with a psychological interest for activities or stimuli that are surprising, novel, of intermediate complexity, or characterized by a knowledge gap or by errors in prediction, which are features that can themselves be quantified mathematically (Schmidhuber, 1991; Oudeyer and Kaplan, 2007; Barto et al., 2013). Such informational features that attract the brain's attention have been called "collative variables" by Berlyne (1965).

Recent experimental studies in psychology and neuroscience have shown that experiencing these features improved memory retention and learning in human children and adults, in other animals and in a variety of tasks. In a famous series of experiments with monkeys, Waelti et al. (2001) showed that monkeys could learn the predictive association between a stimuli and a reward only in situations where prediction errors happened: if the reward was anticipated by other means, then learning was blocked. This experiment complied with formal models of reinforcement learning, and in particular TD learning (Sutton and Barto, 1981), predicting that "organisms only learn when events violate their expectations" (Rescorla and Wagner, 1972, p. 75). In a study mixing behavioral analysis and brain imaging, Kang et al. (2009) showed that human adults show greater long-term memory retention for verbal material for which they had expressed high curiosity than for low-curiosity questions. They observed that before the presentation of answers to high-curiosity questions, curiosity states were correlated with higher activity in the striatum and inferior frontal cortex. When subjects observed answers that did not match their predictions (i.e. an error was experienced), then an increase in activation of putamen and left inferior frontal cortex was observed. The modulation of hippocampus-dependent learning by curiosity states was confirmed in (Gruber et al., 2014). Recently, Stahl and Feigenson (2015) showed that a similar phenomenon happens in infants, observing that the infants created stronger associations between sounds/words and visual objects in a context where object movements violated the expected laws of physics.

Novelty, surprise, intermediate complexity and other related features that characterize informational properties of stimuli have not only been shown to enhance memory retention, but they have also been argued to be intrinsically rewarding, motivating organisms to actively search for them. Three strands of research developed arguments and experimental evidence in this direction. First, psychologists proposed that forms of intrinsic motivation motivate the organism to search for information and competence gain. Second, neuroscientists have shown that reward-related dopaminergic circuits can be activated by information independently of extrinsic reward, and behavioral preference for novelty can be observed in various animals (as along with the apparently inconsistent observation of neophobia). Third, theoretical computational models and their experimental tests in robots have shown how such mechanisms could function and how they can improve learning efficiency by self-organizing developmental learning trajectories.. In what follows, we discuss these advances in turn, and then study how this perspective on curiosity and learning opens new directionsin educational technologies.

Curiosity and intrinsic-motivation in psychology[1]

---

[1] Parts of the text in this section is adapted with permission from (Oudeyer and Kaplan, 2007).

2 In this instantiation of the LP hypothesis, an internal module metaM monitors how learning progresses to generate intrinsic rewards. However, the LP hypothesis in general does not require such an internal capacity for measuring learning progress : such information may also be provided by the environment, either directly by objects or games children play with, or by adults/social peers.

3 Here, action selection is made within a simplified form of reinforcement learning: learning progress is maximized only on the short term, and the environment is configured so that it returns to a rest position after each sensorimotor experiment. This corresponds to what is

In psychology, curiosity can be approached within the conceptual framework of intrinsic motivation (Ryan and Deci, 2000; Berlyne, 1960). Ryan and Deci (2000) proposed a distinction of intrinsic and extrinsic motivation based on the concept of instrumentalization (pp. 56):

"Intrinsic motivation is defined as the doing of an activity for its inherent satisfaction rather than for some separable consequence. When intrinsically motivated, a person is moved to act for the fun or challenge entailed rather than because of external products, pressures or reward."

Intrinsic motivation is clearly visible in young infants, who consistently try to grasp, throw, bite, squash or shout at new objects they encounter, without any clear external pressure to do it. Although the importance of intrinsic motivation declines during development, human adults are still often intrinsically motivated to engage in activities such as crossword puzzles, painting, gardening, read novels or watch movies. Accordingly, Ryan and Deci define extrinsic motivation as:

"Extrinsic motivation is a construct that pertains whenever an activity is done in order to attain some separable outcome. Extrinsic motivation thus contrasts with intrinsic motivation, which refers to doing an activity simply for the enjoyment of the activity itself, rather than its instrumental value." (Ryan and Deci, 2000)

Given this broad distinction between intrinsic and extrinsic motivation, psychologists have proposed theories about which properties of activities make them intrinsically motivating, and in particular foster curiosity as one particular form of intrinsically motivated exploration (Oudeyer and Kaplan, 2007).

Drives to manipulate, drives to explore. In the 1950s, psychologists attempted to give an account of intrinsic motivation and exploratory activities on the basis of the theory of drives (Hull, 1943), defined as specific tissue deficits that the organisms try to reduce, like hunger or pain. Montgomery, 1954 proposed a drive for exploration and Harlow, 1950 proposed that subjects have a drive to manipulate. This drive naming approach had shortcomings which were criticized by White in 1959 (1959): intrinsically motivated exploratory activities have a fundamentally different dynamics. Indeed, they are not homeostatic: the general tendency to explore is not a consummatory response to a stressful perturbation of the organism's body.

Reduction of cognitive dissonance. An alternative conceptualization was proposed by Festinger's theory of cognitive dissonance (Festinger, 1957), which asserted that organisms are motivated to reduce dissonance, defined as an incompatibility between internal cognitive structures and the situations currently perceived. Fifteen years later, a related view was articulated by Kagan stating that a primary motivation for humans is the reduction of uncertainty in the sense of the "incompatibility between (two or more) cognitive structures, between cognitive structure and experience, or between structures and behavior" (Kagan, 1972). More recently, the related concept of "knowledge gap" was argued to be a driver for curiosity-driven exploration (Lowenstein, 1994). However, these theories do not provide an

account of certain spontaneous exploration behaviours which increase uncertainty (Gottlieb et al., 2013). Also, they do not specify whether the brain values differently or similarly different degrees of knowledge gaps.

Optimal incongruity. People seem to look for situations between completely uncertain and completely certain. In 1965, Hunt developed the idea that children and adults look for optimal incongruity (Hunt, 1965). He regarded children as information-processing systems and stated that interesting stimuli were those where there was a discrepancy between the perceived and standard levels of the stimuli. For Dember and Earl, the incongruity or discrepancy in intrinsically-motivated behaviors was between a person's expectations and the properties of the stimulus (Dember and Earl, 1957). Berlyne developed similar notions as he observed 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 recently echoed by Kidd et al. (2012) who showed an experiment where infants prefered stimuli of intermediate complexity.

Motivation for competence. A last group of researchers preferred the concept of challenge to the notion of optimal incongruity. These researchers stated that what was driving human behavior was a motivation for "effectance" (White, 1959), personal causation (De Charms, 1968), competence and self-determination (Deci and Ryan, 1985). Basically, these approaches argue that what motivates people is the degree of control they can have on other people, external objects and themselves. An analogous concept is that of optimal challenge as put forward in the theory of "Flow" (Csikszentmihalyi, 1991).

Berlyne's informational approach to curiosity and intrinsic motivation. These diverse theoretical approaches to intrinsic motivation and the properties that render certain activities intrinsically interesting/motivating have been proposed by diverse research communities within psychology, but so far there is no consensus on a unified view of intrinsic motivation. Even more, it could be argued that distinguishing intrinsic and extrinsic motivation based on instrumentalization can be circular (Oudeyer and Kaplan, 2007). Yet, a convincing integrated non-circular view has actually been proposed in the 60's by Daniel Berlyne (Berlyne, 1965), and has been used as a fruitful theoretical reference for developing formal mathematical models of curiosity, as described below. The central concept of this integrated approach to intrinsic motivation is that of "collative variables", as explained in the following quotations:

"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, pp. 245),

« ... these properties possess close links with the concepts of information theory, and they can, in fact, all be discussed in information-theoretic terminology. In the case of "ambiguity" and "indistinctiveness", there is uncertainty due to a gap in available information. In some forms

of "novelty" and "complexity", there is uncertainty about how a pattern should be categorized, that is, what labeling responses should be attached to it and what overt response is appropriate to it. When one portion of a "complex" pattern or of a sequence of "novel" stimuli is perceived, there is uncertainty about what will be perceived next. In the case of "surprisingness" and "incongruity", there is discrepancy between information embodied in expectations and information embodied in what is perceived. For these reasons, the term "collative" is proposed as an epithet to denote all these stimulus properties collectively, since they all depend on collation or comparison of information from different stimulus elements, whether they be elements belonging to the present, past or elements that are simultaneously present in different parts of one stimulus field".

It should be pointed out that the uncertainty we are discussing here is "subjective uncertainty", which is a function of subjective probabilities, analogous to the "objective" uncertainty (that is, the standard information-theoretic concept of uncertainty) that is a function of objective probabilities." (Berlyne, 1965), pp. 245-246.

As these psychological theories of curiosity and intrinsic motivation hypothesize that the brain could be intrinsically rewarded by experiencing information gain, novelty or complexity, a natural question that follows is whether one could identify actual neural circuitry linking the detection of novelty with the brain reward system. We now review several strands of research that identified several dimensions of this connection.

Information as a reward in neuroscience

Dopaminergic systems that process primary rewards are activated by curiosity. To examine the motivational systems that are recruited by curiosity, Kang et al. used functional magnetic resonance imaging (fMRI) to monitor brain activity in human observers who pondered trivia questions (Kang et al., 2009). After reading a question subjects rated their curiosity and confidence regarding the question and, after a brief delay, were given the answer. The key analyses focused on activations during the *anticipatory* period – after the subjects had received the question but before they were given the answer.

Areas that showed activity related to curiosity ratings during this epoch included the left caudate nucleus, bilateral inferior frontal gyrus (IFG), and loci in the putamen and globus pallidus. In an additional behavioral task, the authors showed that subjects were willing to pay a higher price to obtain the answers to questions that they were more curious about – i.e., could compare money and information on a common scale. They concluded that the value of the information, reported by subjects as a feeling of curiosity, is encoded in some of the same structures that evaluate material gains.

Two recent studies extend this result, and report that midbrain dopaminergic (DA) cells and cells in the orbitofrontal cortex (OFC), a pre-frontal area that receives DA innervation, encode the anticipation of obtaining reliable information from visual cues (Blanchard, Hayden, & Bromberg-Martin, 2015; Bromberg-Martin & Hikosaka, 2009).In that study on DA cells, monkeys were trained on so-called "observing paradigms", where they had to choose between observing two cues that had equal physical rewards but differed in their offers of information

(Bromberg-Martin & Hikosaka, 2009). Monkeys began each trial with a 50% probability of obtaining a large or a small reward and, before receiving the reward, had to choose to observe one of two visual items. If the monkeys chose the informative target, this target changed to one of two patterns that reliably predicted whether the trial will yield a large or small reward ("Info"). If the monkeys chose the uninformative item, this target also changed to produce one of two patterns, but the patterns had only a random relation to the reward size ("Rand").

After a relatively brief experience with the task, the monkeys developed a reliable and consistent preference for choosing the informative cue. Because the extrinsic rewards that the monkeys received were equal for the two options (both targets had a 50% probability of delivering a large or small reward), this showed that monkeys were motivated by some cognitive or emotional factor that assigned intrinsic value to the predictive/informational cue.

Dopamine neurons encoded both reward prediction errors and the anticipation of reliable information. The neurons' responses to reward prediction errors confirmed previous results and arose *after* the monkeys' choice, when the selected target delivered its reward information. At this time, the neurons gave a burst of excitation if the cue signaled a large reward (a better than the average outcome) but were transiently inhibited if the cue signaled a small reward (an outcome that was worse than expected).

Responses to anticipated information gains, by contrast, arose *before* the monkeys' choice and thus could contribute to motivating that choice. Just before viewing the cue, the neurons emitted a slightly stronger excitatory response if the monkeys expected to view an informative cue and a weaker response if they expected only the random cue (red vs. blue traces). This early response was clearly independent of the final outcome and seemed to encode enhanced arousal or motivation associated with the informative option.

A subsequent study of area OFC extended the behavioral results by showing that the monkeys will choose the informative option even if its payoff is slightly lower than that of the uninformative option – that is, monkeys are willing to sacrifice juice reward to view predictive cues (Blanchard et al., 2015). In addition, the study showed that responses to anticipated information gains in the OFC are carried by a neural population that is different from those that encode the value of primary rewards, suggesting differences in the underlying neural computations.

Together, these investigations show that, in both humans and monkeys, the motivational systems that signal the value of primary rewards are also activated by the desire to obtain information. This conclusion is consistent with earlier reports that DA neurons respond to novel or surprising events that are critical for learning environmental contingencies (Bromberg-Martin ES, Matsumoto M, Hikosaka, 2010). The convergence of responses related to rewards and information gains is highly beneficial in allowing subjects to compare different types of currencies – e.g., knowledge and money – on a common value scale when selecting actions. At the same time, the separation between the neural representations of information value and biological value in OFC cells, highlights the fact these two types of values require distinct computations. While the value of a primary reward depends on its biological properties (e.g., its caloric content) the value of a source of information depends on

semantic and epistemic factors that establish the meaning of the information.

Seeking information for itself: liking and wanting novelty, surprise and intermediate complexity. Many animal studies have shown phenomena of neophilia. Rats prefer novel environments and objects to familiar ones (Bardo and Bevins, 2000) and learn motor strategies that allow them to trigger the appearance of novel items (Myers and Miller, 1954). In certain contexts, rats have also been shown to prefer obtaining novel stimuli over obtaining food or drug or at the cost of crossing electrifying grids (see Hugues, 2007 for a review). Moreover, brain responses to novelty in rats have strong similarities with brain responses to drug rewards (Bevins, 2001). In human adults, studies by Itti and Baldi have shown that surprise, defined in the domain of visual features, attracts human saccades during free-viewing exploration (Itti & Baldi, 2009). Baranes et al extended this result to the epistemic domain, by showing that curiosity about trivia questions elicits faster anticipatory eye movements to the expected location of the answer, suggesting that eye movements are influenced by expected gains in semantic information (Baranes et al., 2015). Kidd et al. (2012) showed that human infants had a preference for looking at stimuli of intermediate complexity in the visual or auditory domain (Kidd et al., 2014).

Another recent study suggests that novelty also recruits attentional resources through reward-independent effects (Foley, Jangraw, Peck, & Gottlieb, 2014; Peck, Suzuki, Efem, & Gottlieb, 2009). In this experiment, monkeys were trained on a task in which they had initial uncertainty about the trial's outcome, and were given cues that resolved this uncertainty, by signaling whether the trial will end in a reward or a lack of reward. When the reward contingencies were signaled by novel visual cues (abstract patterns that the monkeys had never seen before), these cues evoked enhanced visual and orienting responses in the parietal lobe. If a novel cue signaled "bad news" (a lack of reward) the monkeys quickly learned this contingency and extinguished their anticipatory licking in response to the cues. Strikingly however, the newly-learnt cues continued to produce enhanced visual and saccadic responses for dozens of presentations after the extinction of the licking response. This suggests that novelty attracts attention through reward-independent mechanisms, allowing the brain to prioritize and learn about novel items for an extended period even if these items signal negative outcomes.

The puzzle of neophobia. Gershman and Niv (2015) recently discussed a puzzling observation. Alongside a large experimental corpus showing neophilia in several animal species, an equally large corpus demonstrates neophobia - the avoidance of novelty (Hughes, 2007). Neophobia has been observed in rats (Blanchard et al., 1974), in adult humans (Berlyne, 1960) in infants (Weizmann et al., 1971) and in non-human primates (Weiskrantz and Cowey, 1963). To explain the apparent contradiction between these results Gershman and Niv (2015) studied the hypothesis that certain kinds of novelty (characterized by their cues) can be selectively and aversively reinforced. That is, an individual may learn and generalize that, in different families of situations, novelty may be associated with positive or with negative outcomes, and thus learn to avoid novelty when their associated outcome is negative.

However, another complementary hypothesis to explain this apparent contradiction is the "intermediate novelty" hypothesis proposed by Berlyne (1960). Following this hypothesis, approach or avoidance of novelty would depend on the degree of novelty, i.e. the degree of distance/similarity between the perceived stimuli and existing internal representations in the brain.

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 theories of curiosity, and in particular its computational theories, have questioned its applicability as an operant concept capable to generate 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 or intermediate complexity stimulus 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 convey useful or learnable patterns. As curiosity-driven spontaneous exploration may have evolved as a mean to acquire information and skills in rapidly changing environments (Barto, 2013), it appears that heuristics based on searching for novelty and complexity can be inefficient in large or non-stationary environments (Schmidhuber, 2001; Oudeyer et al., 2007).

For these reasons, computational learning theory has explored an alternative mechanism, in which *learning progress* generates intrinsic reward (Schmidhuber, 2001; Oudeyer et al., 2007), and it was hypothesized that this mechanism could be at play in humans and animals (Kaplan and Oudeyer, 2007; Oudeyer and Smith, 2016). This hypothesis proposes that the brain, seen as a predictive machine constantly trying to anticipate what will happen next, is intrinsically motivated to pursue activities in which predictions are *improving*, i.e. where uncertainty is decreasing and learning is actually happening. This means that the organism loses interest in activities that are too easy or too difficult to predict (i.e. where uncertainty is low or where uncertainty is high but not reducible), and focuses specifically on learnable activities that are just beyond its current predictive capacities. So for example, an infant will be more interested in exploring how its arm motor commands can allow her to predict the movement of her hand in the 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 (novel but not learnable). As shown by the computational studies we discuss below, a practical consequence of behaviors driven by the search for learning progress is the targeted exploration of activities and stimuli of "intermediate complexity". Yet, an explicit measure of intermediate complexity is not computed by this mechanism: it is an emergent property of selecting actions and stimuli that maximize the derivative of errors in prediction.

The LP hypothesis posits a positive feedback loop between curiosity and learning

The learning progress hypothesis posits a new causal link between learning and curiosity. As described in the first two sections of the article, previous work in neuroscience and psychology considered a unidirectional causal chain: the brain would be motivated to search for (intermediate) novelty or complexity, and then when finding it would be in a curiosity state that would foster learning and memory retention (see figure 1 (A)). In this view (Stahl and Feigenson, 2015; Kang et al., 2009), learning in itself does not have consequences on state curiosity and motivation. On the contrary, the learning progress hypothesis proposes that experiencing learning in a given activity (rather than just intermediate novelty) triggers an intrinsic reward, and thus that learning in itself causally influences state curiosity and intrinsic motivation (see figure 1 (B)). **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.** As shown by computational experiments outlined below, this feedback loop has important consequences on the organization of learning experiences on the long term: as learners actively seek for situations and activities which maximize learning progress, they will first focus on simple learnable activities before shifting to more complex ones (see figure 2), and the activities they select shape their knowledge and skills, which will in turn change the potential progress in other activities and thus shape their future exploratory trajectories. As a consequence, the learning progress hypothesis does not only introduce a causal link between learning and curiosity, but also introduces the idea that curiosity may be a key mechanism in shaping developmental organization. Below, 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.

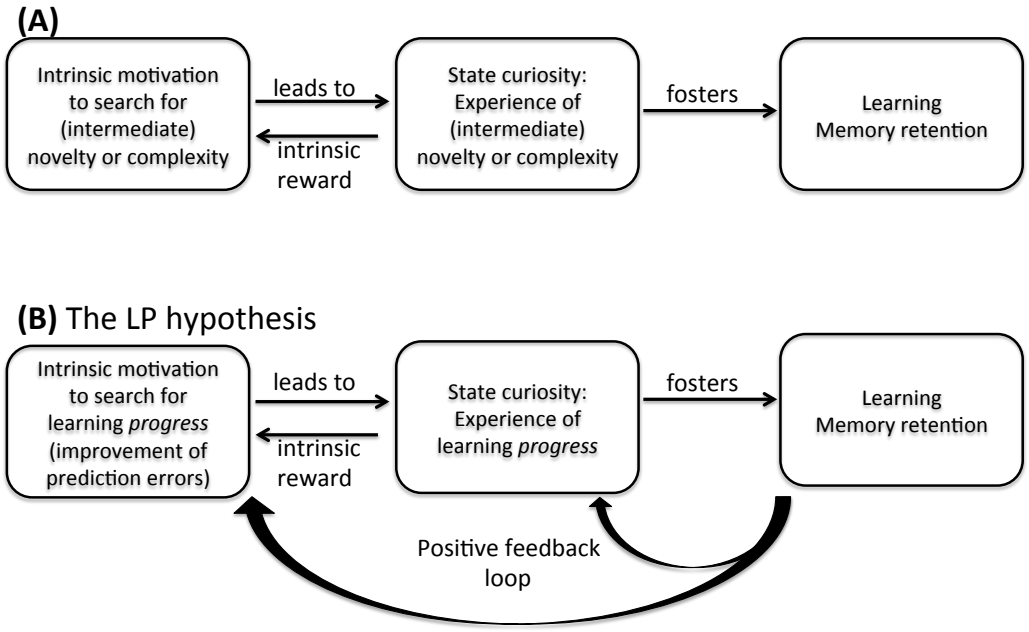

Figure 1

Figure 1 Many studies of curiosity and learning have considered a one directional causal relationship between state curiosity and learning (A). The Learning Progress hypothesis suggest that learning progress itself, measured as the improvement of prediction errors, can be intrinsically rewarding: this introduces a positive feedback loop between state curiosity and learning (B). This positive feedback loop in turn introduces a complex learning dynamics self-organizing learning curriculum with phases of increasing complexity, such as in the Playground Experiment (Oudeyer et al., 2007; see Figures 2 and 3).

The LP hypothesis unifies various qualitative theories of curiosity.

The LP hypothesis is also associated to a mathematical formalism (outlined in the next section) that allows to bridge several hypotheses related to curiosity and intrinsic motivation, that had been so far conceptually separated (Oudeyer and Kaplan, 2007). 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, and such mechanisms correspond to a class of intrinsic motivation that has been called "knowledge-based intrinsic motivation" (Oudeyer and Kaplan, 2007; Mirolli

and Baldassarre, 2013). It also applies to predicting the consequences of one's own actions in particular situations, 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, can be used to model Csikszentmihalyi's concept of intermediate challenge in the flow theory as well as related theories of intrinsic motivation based on self-measures of competences (White, 1959 ; Csikszentmihalyi, 1991). This second form of the LP hypothesis, where learning progress is measured in terms of how much competences improve with experience, correspond to a class of intrinsic motivation mechanisms that has been called "competence-based intrinsic motivation" (Oudeyer and Kaplan, 2007 ; Mirolli and Baldassarre, 2013).

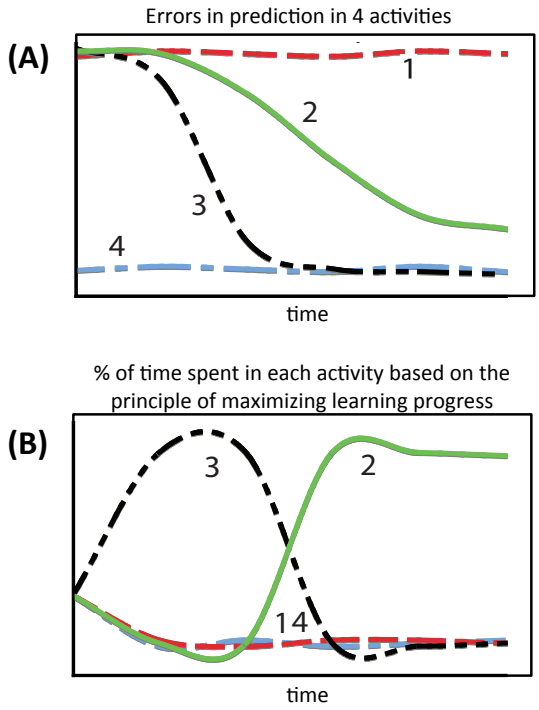

Figure 2

Figure 2 The LP hypothesis proposes that active spontaneous exploration will favor exploring activities which are providing maximum improvement of prediction errors. If one imagines four activities with different learning rate profiles (A), then LP-driven exploration will avoid activities that are either too easy (4) or too difficult (1) as they do not provide learning progress, then first focus on an activity which initially provides maximal learning progress (3), before reaching a learning plateau in this activity and shifting to another one (2) which at this point in the curriculum provides maximum progress (potentially thanks to skills acquired in activity (3)). As a consequence, an ordering of exploration phases forms spontaneously, generating a structured developmental trajectory (adapted from Kaplan and Oudeyer, 2007).

Computational models: curiosity-driven reinforcement learning

Computational and robotic models have recently thrived in order to conceptualize more precisely theories of curiosity-driven learning and intrinsic motivation, as well as to study the associated learning dynamics and make experimental predictions (Baldassarre and Mirolli, 2013; Gottlieb et al., 2013). A general formal framework that has been used most often to model learning and motivational systems is computational reinforcement learning (Sutton and Barto, 1998). In reinforcement learning, one considers a set of states S (characterizing the state of the world as sensed by sensors as well as the state of internal memory); a set of actions A that the organism can make; a reward function R(s,a) that provides a number r(s,a) that depends on states and actions and that should be maximized; an action policy P(a|s) which determines which actions should be made in each state so as to maximize future expected reward; and finally a learning mechanism L that allows to update the action policy in order to improve rewards in the future. Many works in computational neuroscience and psychology have focused on the details of the learning mechanism, for example to explain differences in model-based versus model-free learning (Gershman, in press). However, the same framework can be used to model motivational mechanisms, through modeling the structure and semantics of the reward function. For example, extrinsic motivational mechanisms associated to food/energy search can be modeled through a reward function that measures the quantity of food gathered (Arkin, 2005). A motivation for mating can be modeled similarly, and as each motivational mechanism is modeled as a real number that should be maximized, such numbers can be used as a common motivational currency to make tradeoffs among competing motivations (Konidaris and Barto, 2006).

Similarly, it is possible to use this framework to provide formal models of intrinsic motivation and curiosity as formulated by most theories mentioned above, in architecture called "intrinsically motivated reinforcement learning" (Singh et al., 2004) and as reviewed in (Baldassarre and Mirolli, 2013; Oudeyer and Kaplan, 2007). In this context, an intrinsic motivation system that pushes organisms to search for novelty can be formalized for example by considering a mechanism which counts how often each state of the environment has already been visited, and then using a reward function that is inversely proportional to these counts. This corresponds to the concept of exploration bonus studied by Dayan et al. (1996) and Sutton (1990). If one considers a model-based RL system that learns to predict which states will be observed upon a series of actions, as well as measures of uncertainty of these predictions, one can formalize surprise (and automatically derive an associated reward) as situations in which the subject makes an unexpected high error in predictions.

To understand how the learning progress hypothesis can be formally modeled in this framework, let us consider the model used in the Playground Experiment (see figure 3 (A)). In this experiment, a quadruped "learning" robot (the learner) is placed on an infant play mat with a set of nearby objects and is joined by an "adult" robot (the teacher), see Figure 3 (A) (Oudeyer and Kaplan, 2006; Kaplan and Oudeyer, 2007b; Oudeyer et al., 2007). On the mat and near the learner are objects for discovery: an elephant (which can be bitten or "grasped" by the mouth), a hanging toy (which can be "bashed" or pushed with the leg). The teacher is pre-programmed to imitate the sounds made by the learner when the learning robot looks to

the teacher while vocalizing at the same time.

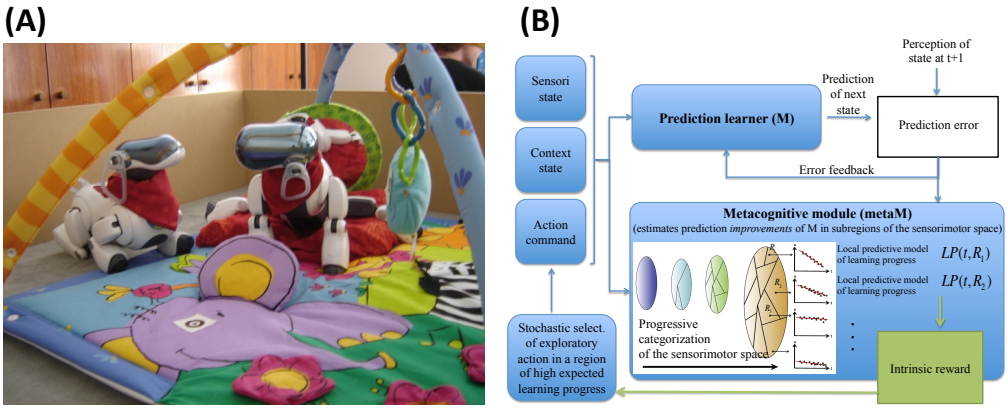

Figure 3

Figure 3 (A) The Playground Experiment: a robot explores and learns the contingencies between its movement and the effect they produce on surrounding objects. To drive its exploration, it uses the active learning architecture described in (B). In this architecture, a meta-learning module tracks the evolution of errors in predictions that the robot makes using various kinds of movements in various situations. Then, an action selection module selects probabilistically actions and situations which have recently provided high improvement of predictions (learning progress), using this measure to heuristically expect further learning progress in similar situations. Adapted from (Oudeyer et al., 2007).

The learner is equipped with a repertoire of motor primitives parameterized by several continuous numbers that control movements of its legs, head and a simulated vocal production system. Each motor primitive is a dynamical system controlling various forms of actions: (a) turning the head in different directions; (b) opening and closing the mouth while crouching with varying strengths and timing; (c) rocking the leg with varying angles and speed; (d) vocalizing with varying pitches and lengths. These primitives are parameterized by real numbers and can be combined to form a large continuous space of possible actions. Similarly, 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*, by choosing physical experiences ("experiments") that improve the quality of predictions of the consequences of its actions. As data is collected though this exploration process, the robot builds a model of the world dynamics that can be reused later on for new tasks that were not known at the time of exploration (for example using model-based reinforcement learning mechanisms).

Figure 3 (B) outlines a computational architecture, called R-IAC (Oudeyer, Kaplan et al. 2007; Moulin-Frier et al., 2014). 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 which visual movements or proprioceptive perceptions result from using a leg motor primitive with certain parameters (this model learning can be done with a neural network or any other statistical m`:ùxc nklninference algorithm). Another module (metaM) estimates the evolution of errors in prediction of M in various regions of the sensorimotor space[2]. This module estimates how much errors decrease in predicting an action in certain situations, for example, in predicting the consequence of a leg movement when this action is applied towards a particular area of the environment. These estimates of error reduction are used to compute the intrinsic reward from progress in learning. This reward is an internal quantity that is proportional to the decrease of prediction errors, and the maximization of this quantity is the goal of action selection within a computational reinforcement-learning architecture (Kaplan and Oudeyer, 2003; Oudeyer and Kaplan, 2007; Oudeyer et al., 2007). Importantly, the action selection system chooses most often to explore activities where the estimated reward from 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 that may also yield progress in learning[3]. 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. This categorization thereby models the incremental creation and refining of cognitive categories differentiating activities/tasks.

In all of the runs of the experiment, one observes the self-organization of structured

2 In this instantiation of the LP hypothesis, an internal module metaM monitors how learning progresses to generate intrinsic rewards. However, the LP hypothesis in general does not require such an internal capacity for measuring learning progress : such information may also be provided by the environment, either directly by objects or games children play with, or by adults/social peers.

3 Here, action selection is made within a simplified form of reinforcement learning: learning progress is maximized only on the short term, and the environment is configured so that it returns to a rest position after each sensorimotor experiment. This corresponds to what is called episodic reinforcement learning, and action selection can be handled efficiently in this case using multi-armed bandit algorithms (Audibert et al., 2009). Other related computational models have considered maximizing forms of LP over the long term through RL planning techniques in environments which dynamics is state dependent (Schmidhuber, 1991; Kaplan and Oudeyer, 2003) and non-stationary (Lopes et al., 2012).).

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. Typically, after a phase of random body babbling, the robot focuses on performing various kinds of actions towards objects, and then focuses on some objects with particular actions that it discovers are relevant for the object. In the end, the robot is able to acquire sensorimotor skills such as how to push or grasp objects, as well as how to perform simple vocal interactions with another robot, as a side effect of its general drive to maximize learning progress. This typical trajectory can be explained as gradual exploration of new progress niches (zones of the sensorimotor space where it progresses in learning new skills), and those stages and their ordering can be viewed as a form of attractor in the space of developmental trajectories. Yet, one also observes diversity in the developmental trajectories observed in the experiment. With the same mechanism and same initial parameters, individual trajectories may generate qualitatively different behaviors or even invert stages.. This is due to the stochasticity on the policy, 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). This diversity can be seen as an interesting modeling outcome since individual development is not identical across different individuals but is always, for each individual, unique in its own ways. This kind of approach, then, offers a way to understand individual differences as emergent in developmental process itself and makes clear how developmental process might vary across contexts, even with an identical learning mechanism.

How LP-driven curiosity generates developmental trajectories that reproduce infant development sequences and can act in synergy with social learning

Focusing on vocal development, Moulin-Frier et al. conducted experiments where a robot explored the control of a realistic model of the vocal tract in interaction with vocal peers through a drive to maximize learning progress (Moulin-Frier et al., 2014). This model relied on a physical model of the vocal tract, its motor control and the auditory system. It also included an additional mechanism allowing the active learner to take into account social signals provided by peers. As a simulated caretaker would himself produce vocalizations organized around the systematic reuse of certain phonemes, the curiosity-driven learning system could decide whether it should try to reproduce these external speech sounds (imitation) using its current know-how, or whether it should self-explore other kinds of speech sounds. The choice was made hierarchically: first, it decided to imitate or self-explore based on how much each strategy provided learning progress in the past. Second, if self-exploration was selected, it decided which part of the sensorimotor space to explore based on how much learning progress could be expected. The experiments showed how such a mechanism generated automatically 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 vocal variations of unarticulated sounds, and finally articulated proto-syllables. In this initial phase, imitation is rarely tryied by the

learner as the sounds produced by caretakers are too complicated to make any progress. But as the vocal learner becomes more proficient at producing complex sounds through self-exploration, the imitating vocalizations of the teacher begin to provide high learning progress, resulting in a shift from self-exploration to vocal imitation. This also illustrates how intrinsically motivated self-exploration can guide the system to efficiently and autonomously acquire basic sensorimotor skills that are instrumental to learn faster other more complicated skills.

Intrinsically motivated exploration scaffolds efficient multitask learning

Computational models in the literature have shown how various forms of intrinsically motivated exploration and learning could guide efficiently the autonomous acquisition of repertoires of skills in large and difficult spaces.

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 and Oudeyer, 2013; Oudeyer et al., 2007), or through model-based planning mechanisms (Schmidhuber, 1991; Singh and Barto, 2004; Lopes et al., 2012), 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 towards 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 and Oudeyer, 2013; Benureau and 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 behavioural space a robot to find a reward in a maze more

efficiently than if it had been searching for behavioural parameters that optimized directly the reward. In another model, Forestier and Oudeyer (2016 REF) 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 (and where direct search for tool use behaviours was vastly 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 block 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 RL exploration. Related uses of intrinsic motivation with a hierarchical reinforcement learning framework were demonstrated in (Bakker and Schmidhuber, 2004; 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 evolutionary computational modelling that given a distribution of changing environments and an extrinsic reward that organisms need to maximize, it could be more robust for RL agents to represent and use a surrogate 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.

Applications in educational technologies and video games

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) have 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 now disseminating at 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 constructivist learning which directly pointed towards the importance of fostering curiosity and free play and exploration in the classroom. Recently, the large body of research in educational psychology has began 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 50-70's 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 provided clear feedback on the performance of users, and when outcomes where 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 et al. (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 and Bavelier, 2014). Within this perspective, Merrick and Mahler (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 so as to promote intrinsic motivation and learning.

Clement et al. (2015) have presented and evaluated and ITS system that directly reused computational models of curiosity-driven learning based on the learning progress hypothesis

described above (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 7-8 years old children (see figure 4). 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 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, i.e. the exercises where their errors decrease fastest. In order to identify dynamically 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 to guide the algorithm towards 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 upper in the hierarchy are made available to the student (see figure 5). 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 targeted to propose exercises that could stimulate the intrinsic motivation of students by dynamically and continuously proposing them challenges that were neither too difficult nor too easy. Second, by doing this using learning progress, it targeted to generate 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 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.

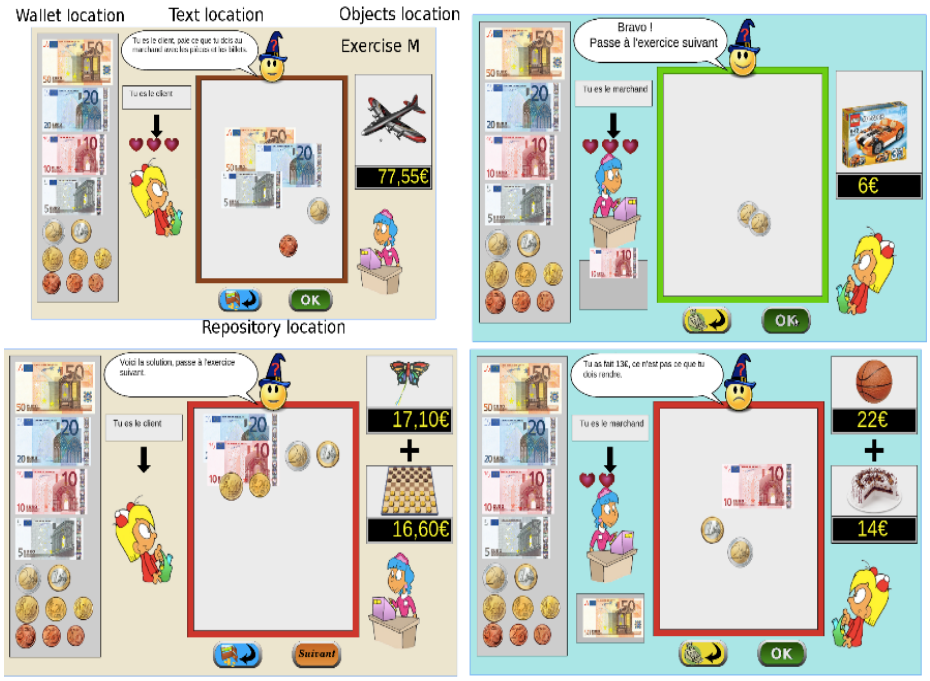

Figure 4

Figure 4 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 ITS 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.

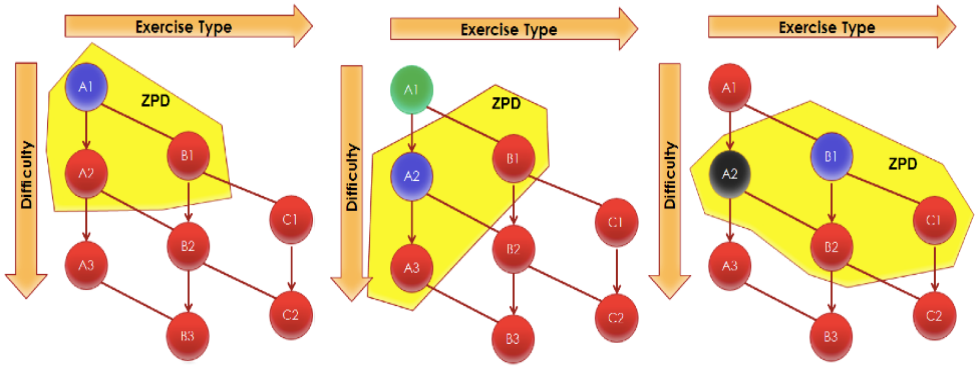

Figure 5

Figure 5 Example of the evolution of the zone-of-proximal development 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 pre-conditions 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, e.g. using a different modality, and A3 is temporarily removed from the ZPD. (Adapted from Clement et al., 2015).

Several related ITS systems were developed and experimented. 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 and Loekx, 2015). In another earlier study, Pachet (2004) presented a computer system targeting to help children discover and learn how to play musical instruments, but also capable to support creativity in experienced musicians, through fostering the 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 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 observed that both children and expert musicians most often experience an "Eureka moment". 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.

Discussion: convergences, open questions and educational design

Converging research strands in psychology, neuroscience and computational learning theory indicate that curiosity and learning are strongly connected along several dimensions, and that these connections have wide implications for education.

As often informally observed by many education practioners, recently developed experimental protocols showed that experiencing situations with novelty, complexity and prediction errors fostered memory retention. Furthermore, several lines of evidence showed that the brain is equipped with neural circuits which consider information as an intrinsic reward, and thus actively searches for these situations featuring novelty and/or prediction errors.

At the same time, there are several open scientific questions. One of them is to characterize precisely which informational features are intrinsically rewarding. As mathematical formalization shows, novelty, complexity and prediction errors can be computed in many different ways, and then used potentially in equally different ways to determine intrinsic reward in curiosity-driven exploration. For example, some hypotheses approach curiosity as a mechanism maximizing surprise (Itti and Baldi, 2009), or intermediate complexity (Kidd et al., 2012), or learning progress (Oudeyer et al., 2007). While on some situations these formal variations may be equivalent, they can also generate vastly different learning dynamics. For example, the LP hypothesis introduces a positive feedback loop between learning and state curiosity, and in turn this involves deep consequences on the long-term formation of learning and developmental trajectories (Oudeyer and Smith, 2016). A related question is whether the brain includes a unified mechanism for spontaneous exploration, or whether it combines several of these heuristics (and how this combination happens). Designing experimental protocols to disentangle these various hypotheses is the subject of active current research (Baranes et al., 2014; Medder and Nelson, 2012; Taffoni et al., 2014; Markant et al., 2015).

However, even if these questions are still unresolved, existing results suggest several guidelines for educational practice and the design of educational technologies. First, they highlight the importance of providing students with learning materials that are informationally engaging (surprising or with the right level of complexity/learnability) in order to foster memory retention. Second, they suggest the importance of personalization, active learning and active teaching. Indeed, features like (intermediate) novelty or learning progress are fundamentally subjective in the sense that they are a measure of the relation between a particular educational material and the state of knowledge of each particular student at a given time of its learning trajectory. As a consequence, what triggers curiosity and learning will be different for different students. Human or computational teachers can address this issue by tracking the errors and behaviors of each student in order to present sequences of items that are personalized to maximize their experience of features associated to states of curiosity and motivation. Learners have also a fundamental capability that should be leveraged: as their

brain is intrinsically rewarded by features like novelty or learning progress, they will spontaneously and actively search for these features and select adequate learning materials if the environment/teacher provides sufficient choices. While most existing studies have focused on either active learning or active teaching, the study of the dynamic interaction between active learners and teachers is still a largely open question that should be addressed to understand how this dynamics could scaffold mutual guidance towards efficient curiosity-driven learning.

References

Arkin, R., 2005. Moving up the food chain: motivation and emotion in behavior based robots. In Who Needs Emotions: The Brain Meets the Robot, J. Fellous and M. Arbib, eds (Oxford University Press), pp. 245–270.

Audibert, J-Y., Munos, R. and C. Szepesvari, 2009. Exploration-exploitation tradeoff using variance estimates in multi-armed bandits. Theor. Comput. Sci., 410(19):1876–1902.

Bakker, B., Schmidhuber, J., 2004. Hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization. In Proc. of the 8-th Conf. on Intelligent Autonomous Systems (pp. 438-445).

Baldassare, G. and M. Mirolli, 2013. Intrinsically motivated learning in natural and artificial systems, Berlin: Springer-Verlag.

Baranes, A., Oudeyer, P. Y. and Gottlieb, J., 2015. Eye movements reveal epistemic curiosity in human observers. *Vision research*, *117*, 81-90.

Baranes, A. F., Oudeyer, P. Y., & Gottlieb, J., 2014. The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in neuroscience*, *8*.

Bardo, M., Bevins, R., 2000. Conditioned place preference: what does it add to our preclinical understanding of drug reward? Psychopharmacology 153 (1), 31–43.

Barto, A., 2013. Intrinsic motivation and reinforcement learning In G. Baldassarre and M. Mirolli, editors, Intrinsically Motivated Learning in Natural and Artificial Systems, pp. 17-47, Springer.

Benureau FCY and Oudeyer P-Y., 2016. Behavioral Diversity Generation in Autonomous Exploration through Reuse of Past Experience. Front. Robot. AI 3:8. doi: 10.3389/frobt.2016.00008

Berlyne, D., 1960. Conflict, Arousal and Curiosity (McGraw-Hill).

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

K. Beuls, 2013. *Towards an agent-based tutoring system for Spanish verb conjugation*. PhD thesis, Vrije Universiteit Brussel.

Beuls, K. and Loeckx, J., 2015. **Steps towards intelligent MOOCs : A case study for learning counterpoint.** Music Learning With Massive Open Online Courses. ed. / Luc Steels. Amsterdam : IOS Press, 2015. p. 119-144 9 (The Future of Learning; Vol. 6).

Blanchard, T. C., Hayden, B. Y., and Bromberg-Martin, E. S., 2015. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron, 85*(3), 602-614.

Blanchard, R., Kelley, M., Blanchard, D., 1974. Defensive reactions and exploratory behavior in rats. Journal of Comparative and Physiological Psychology 87 (6), 1129–1133.

Bromberg-Martin, E. S., & Hikosaka, O., 2009. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron, 63*(1), 119-126.

Bromberg-Martin ES, Matsumoto M, Hikosaka O., 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron. 2010 Dec 9;68(5):815-34. doi: 10.1016/j.neuron.2010.11.022.

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, Special Issue on Developmental Disorders, 27(2), 185-191.

Dayan, P. and Sejnowski, T.J., 1996. Exploration bonuses and dual control. Mach. Learn. 25, 5–22

Barto, A., 2013. Intrinsic motivation and reinforcement learning In G. Baldassarre and M. Mirolli, editors, Intrinsically Motivated Learning in Natural and Artificial Systems, pp. 17-47, Springer.

Barto, A., Mirolli, M., and Baldasarre, G., 2013. **Novelty or surprise?** *Frontiers in Cognitive Science*, 11, doi: 10.3389/fpsyg.2013.00907

Clement B., Roy D., Oudeyer P-Y., Lopes M., 2015. Multi-Armed Bandits for Intelligent Tutoring Systems, ***Journal of Educational Data Mining (JEDM)**, Vol 7, No 2*.

Bevins, R., 2001. Novelty seeking and reward: Implications for the study of high-risk behaviors. Current Directions in Psychological Science 10 (6), 189.

Cordova, D.I. and Lepper, M.R., 1996. Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of educational psychology*, *88*(4), p.715.

Csikszenthmihalyi, M. (1991). Flow-the Psychology of Optimal Experience (Harper Peren- nial).

De Charms, R., 1968. Personal Causation: *The Internal Affective Determinants of Behav- ior* (New York, Academic Press).

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), 1-27.

Dember, W. N., and Earl, R. W., 1957. Analysis of exploratory, manipulatory and curiosity behaviors. *Psychol. Rev.* 64, 91–96.

Foley, N. C., Jangraw, D. C., Peck, C., & Gottlieb, J., 2014. Novelty enhances visual salience independently of reward in the parietal lobe. *J neurosci, 34*(23), 7947-7957.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P., 2014. Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111*(23), 8410–8415.

Froebel, F., 1885. *The education of man*. A. Lovell & Company.

Gershman, S.J., in press. Reinforcement learning and causal models. In M. Waldmann, Ed, *Oxford Handbook*

of Causal Reasoning. Oxford University Press.

Gershman, S.J. and Niv, Y., 2015. Novelty and inductive generalization in human reinforcement learning. *Topics in Cognitive Science*, 1-25.

Gottlieb, J., Oudeyer, P-Y., Lopes, M., Baranes, A., 2013. Information Seeking, Curiosity and Attention: Computational and Neural Mechanisms Trends in Cognitive Science, , 17(11), pp. 585-596.

Gruber, M. J., Gelman, B. D., & Ranganath, C., 2014. States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit. *Neuron*.

Festinger, L., 1957. *A theory of Cognitive Dissonance* (Evanston, 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.

Harlow, H., 1950. Learning and satiation of response in intrinsically motivated complex puzzle performances by monkeys. *J. Comp. Physiolo. Psychology* 43, 289–294.
Hughes, R., 2007. Neotic preferences in laboratory rodents: Issues, assessment and substrates. Neuroscience & Biobehavioral Reviews 31 (3), 441–464.

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.

Itti, L., & Baldi, P., 2009. Bayesian surprise attracts human attention. *Vision research*, *49*(10), 1295-1306.

Kagan, J., 1972. Motives and development. J. Pers. Soc. Psychol. 22, 51–66.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F., 2009. The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol Sci*, *20*(8), 963–73.

Kaplan, F., Oudeyer, P-Y, 2003. Motivational principles for visual know-how development. In Prince, C.G. and Berthouze, L. and Kozima, H. and Bullock, D. and Stojanov, G. and Balkenius, C., editors, Proceedings of the 3rd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems, no. 101, pages 73-80, 2003. Lund University Cognitive Studies.

Kaplan, F. and Oudeyer, P-Y., 2007a. The progress-drive hypothesis: an interpretation of early imitation. In Dautenhahn, K. and Nehaniv, C., editor, Models and mechanisms of imitation and social learning: Behavioural, social and communication dimensions, pp. 361--377, Cambridge University Press.

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

Kidd, C., Piantadosi, S. T., and 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(5), e36399.

Kidd, C., Piantadosi, S.T., & Aslin, R.N., 2014. The Goldilocks effect in infant auditory cognition. Child Development, 85(5):1795-804.

Kidd, C. and Hayden, B. Y., 2015. The psychology and neuroscience of curiosity. *Neuron*, vol. 88, no 3, p.

449-460.

Konidaris, G.D., and Barto, A.G., 2006. **An adaptive robot motivational system** *Animals to Animats 9: Proceedings of the 9th International Conference on Simulation of Adaptive Behavior (SAB-06)*, CNR, Roma, Italy

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J., 2016. Building machines that learn and think like people. CBMM Memo 46

Law, E., Yin, M., Joslin Goh, K. C., Terry, M., and Gajos, K. Z., 2016. Curiosity Killed the Cat, but Makes Crowdwork Better, In *Proceedings of CHI'16*, 2016.

Lehman, J. and Stanley, K. O., 2011. Abandoning objectives: Evolution through the search for novelty alone. Evolutionary computation, 19(2), 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*(3), 202-227.

Lopes, M., & Oudeyer, P. Y., 2012. The strategic student approach for life-long exploration and learning. In *Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference on* (pp. 1-8). IEEE.

Lopes M., Lang T., Toussaint M. and Oudeyer P-Y., 2012. Exploration in Model-based Reinforcement Learning by Empirically Estimating Learning Progress Neural Information Processing Systems (NIPS 2012), Tahoe, USA.

Lowenstein, G., 1994. The psychology of curiosity: a review and reinterpretation, Psychological Bulletin 116(1): 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, Calif., forthcoming.

Markant, D.B., Settles, B., and Gureckis, T.M., 2015. Self-directed learning favors local, rather than global, uncertainty. Cognitive Science, 40(1), pp. 100-120.

Meder, B., & Nelson, J. D., 2012. Information search with situation-specific reward functions. Judgment and Decision Making, 7, 119-148.

Merrick, K. E., & Maher, M. L., 2009. *Motivated reinforcement learning: curious characters for multiuser games*. Springer Science & Business Media.

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.

Montessori, M., 1948/2004. *The discovery of the child*. Aakar Books.

Montgomery, K., 1954. The role of exploratory drive in learning. *J. Comp. Physiol. Psychol.* 47, 60–64.

Moulin-Frier, C., Nguyen, M. and Oudeyer, P-Y., 2014. Self-Organization of Early Vocal Development in Infants and Machines: The Role of Intrinsic Motivation, Frontiers in Cognitive Science, doi: 10.3389/fpsyg.2013.01006

Myers, A., Miller, N., 1954. Failure to find a learned drive based on hunger; evidence for learning motivated by exploration. Journal of Comparative and Physiological Psychology 47 (6), 428.

Nkambou, R., Mizoguchi, R., and Bourdeay, J., 2010. *Advances in intelligent tutoring systems*. Vol. 308. Springer.

Oller, D.K., 2000. The Emergence of the Speech Capacity. Lawrence Erlbaum and Associates, Inc.

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

Oudeyer P-Y. and Kaplan F., 2007. What is intrinsic motivation? A typology of computational approaches Frontiers in Neurorobotics, 1:6, doi: 10.3389/neuro.12.006.2007

Oudeyer P-Y., Kaplan F., 2006. Discovering Communication, Connection Science, 18(2), pp. 189--206.

Oudeyer, P-Y. and Smith, L., 2016. How evolution can work through curiosity-driven developmental process, Topics in Cognitive Science, 8(2), pp. 492-502.

Papert, S., 1980. *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc..

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., and Kafai, Y., 2009. Scratch: programming for all. *Communications of the ACM*, *52*(11), 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. In proceedings of *RIE 2015*.

Singh, S.P., Barto, A.G. and N. Chentanez, 2004. Intrinsically motivated reinforcement learning. In Advances in neural information processing systems, pages 1281–1288.

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

Pachet, F., 2004. On the Design of a Musical Flow Machine. In: Tokoro, M. and L. Steels (eds.) A learning zone of one's own. IOS Press, Amsterdam.

Peck, C. J., Jangraw, D. C., Suzuki, M., Efem, R., & Gottlieb, J., 2009. Reward modulates attention independently of action value in posterior parietal cortex. *J Neurosci, 29*(36), 11182-11191.

Rescorla, R. A., & Wagner, A. R., 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, *2*, 64-99.

Ryan, R., and Deci, E., 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Educ. Psychol.* 25, 54–67.

Schmidhuber, J., 1991. Curious model-building control systems. IEEE International Joint Conference on Neural Networks.

Singh, S., Barto, A.G., and Chentanez, N., 2004. **Intrinsically motivated reinforcement learning** *18th Annual Conference on Neural Information Processing Systems* (NIPS), Vancouver, B.C., Canada

Stahl, A. E., & Feigenson, L., 2015. Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91-94.

Steels,L. 2015. Social Flow in Social MOOCs, In L. Steels, *Music Learning with Massive Open Online Courses* (pp. 119-144). Amsterdam: IOS Press.

L. Steels, 2015. *Music Learning with Massive Open Online Courses* (pp. 119-144). Amsterdam: IOS Press.

Sutton, R. S., and Barto, A. G., 1981. Toward a modern theory of adaptive networks: expectation and prediction. *Psychological review*, *88*(2), 135.

Sutton, R. S., and Barto, A. G., 1998. *Reinforcement learning: An introduction*. MIT press.

Sutton, R.S., 1990. Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In Proceedings of the 7th International Conference on Machine Learning, pp. 216–224, ICML

Sutton, R.S., Modayil, J., Delp, M., Degris, T., Pilarski, P. M., White, A., and D. Precup, 2011. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2, pages 761–768. International Foundation for Autonomous Agents and Multiagent Systems.

Taffoni, F., Tamilia, E., Focaroli, V., Formica, D., Ricci, L., Di Pino, G., Baldassarre, G., Mirolli, M., Guglielmelli, E. and F. Keller, 2014. Development of goal-directed action selection guided by intrinsic motivations: an experiment with children. *Experimental brain research* 232, no. 7: 2167-2177.

Weiskrantz, L., Cowey, A., 1963. The aetiology of food reward in monkeys. Animal Behaviour 11 (2-3), 225–234.
Weizmann, F., Cohen, L., Pratt, R., 1971. Novelty, familiarity, and the development of infant attention. Developmental Psychology 4 (2), 149–154.

White, R., 1959. Motivation reconsidered: The concept of competence. *Psychol. Rev.* 66, 297–333.

Waelti, P., Dickinson, A., and Schultz, W., 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48.