Motivated cognition: Neural and computational mechanisms of curiosity, attention and intrinsic motivation

Jacqueline Gottlieb, Manuel Lopes, Pierre-Yves Oudeyer

Title: **Motivated cognition: Neural and computational mechanisms of curiosity, attention and intrinsic motivation**

Authors: Jacqueline Gottlieb[1,2], Manuel Lopes[3,4] and Pierre-Yves Oudeyer[3,4]

Author affiliation: Department of Neuroscience, Columbia University[1], Kavli Institute for Brain Science, Columbia University[2], Inria, France[3], Ensta-ParisTech, France[4]

Corresponding author:

Jacqueline Gottlieb, PhD
Department of Neuroscience
Columbia University
1051 Riverside Drive,
Kolb Research Annex, Rm. 569
New York, NY 10032
Phone: 212-543-6931, ext. 500
Fax: 212-543-5816
E-mail: jg2141@cumc. columbia.edu

Number of figures: 8

Introduction

Countless studies in neuroscience and psychology have probed the neural basis of cognitive functions such as attention, memory and mental representations. While these studies have traditionally remained independent from studies of decision making and motivation, this separation is beginning to change with the advent of evidence documenting strong effects of motivation on memory (reviewed in this book). These recent results suggest a more integrative view, whereby cognition and motivation are tightly intertwined. This has the strong implication that cognition is not a passive process that is simply "given" to us by the brain, but instead an active, motivated process - a mental act which, much like our physical acts, is proactively oriented toward a goal. A second implication is that cognitive factors – e.g., related to learning, memory or attention - themselves can causally impact motivational states.

This more integrative conception raises fundamental questions about the types of motivation that drive us to *think*. What are the factors that motivate us to learn, memorize or otherwise process new information? How do these intellectual drives serve our biological needs, how do they control our actions and what are their neural substrates?

In this chapter we examine this question with a focus on curiosity – a complex cognitive process that is defined as the intrinsic desire to learn or obtain information. Curiosity reaches its pinnacle in human beings in pursuits such as scientific research, and is arguably a key factor in the considerable success of our species. However, our understanding of curiosity is in its infancy, and its computational and neuroscientific basis are only beginning to be investigated.

We will review recent developments in neuroscience, cognitive psychology and computational modeling and machine learning that pertain to these questions. We will start by reviewing fundamental properties of curiosity and intrinsic motivation, followed by a survey of recent evidence that curiosity recruits motivational systems (including midbrain dopaminergic neurons and dopamine-recipient structures) and systems of selective attention including parietal areas involved in oculomotor control. Finally, we will review a number of factors that contribute to curiosity, including novelty, surprise, uncertainty, rewards and meta-cognitive control, and our current understanding of their neural mechanisms.

A theme that runs throughout the discussion is that curiosity involves a family of mechanisms which, while highly sophisticated in humans, have their roots in more primitive motivational and information sampling systems that are found in many animal species. In addition, we will emphasize the fact that mechanisms of curiosity-driven learning can be computationally modeled, and that such models are highly useful in formulating new hypotheses about the nature and function of curiosity in adult behavior and at a developmental scale.

Curiosity and intrinsic motivation

Curiosity is a specific example of a system of intrinsic motivation - defined as a behavior that is undertaken for no apparent reward except the behavior itself (Ryan and Deci, 2000). As aptly described by Mark Twain in his legendary book *Tom Sawyer, "Work consists of*

whatever a body is obliged to do, and Play consists of whatever a body is not obliged to do. " Tom Sawyer, ch. 2

Intrinsically motivated behaviors include behaviors that subjects are not obliged to do for survival and yet are highly motivated to pursue – such as children's play and adult leisure-time hobbies and creative pursuits. Intrinsically motivated activities are generally pleasurable, and can even cause special states of "flow" characterized by intense feelings of effortless control, concentration, enjoyment and a contraction of the sense of time (Csikszenthmihalyi, 1991).

From a computational perspective, intrinsically motivated behaviors can be characterized as one would any other goal-directed behavior – as actions that seek to maximize an internal goal - formalized mathematically as a reward (value) function. A particular challenge however, is to understand what are the value functions that the agents seek to maximize. Whereas in the vast majority of experiments in neuroscience and psychology, behavior is shaped using easily measurable extrinsic rewards such as money, juice, food or points, intrinsically motivated behaviors depend on internal factors that are much more difficult to characterize and are related to the individual's affective or cognitive structure. For instance, when creating a painting, an individual may be motivated by the desire to please his or her partner (a social reward), a feeling of pleasure in looking at the colors on the canvas (an emotional reward), or a desire to learn more about how to blend colors (a cognitive reward). These factors can be viewed as "rewards" in the broad (and widely accepted) view of the term - as any factor that reinforces behavior and "makes you come back for more" (Thorndike, 1911). However, it remains a formidable challenge to identify which of these internal rewards come into play in any given context and how these motivations are computed by the brain.

Curiosity is a particular system of intrinsic motivation that drives agents to learn. The curious agent shown in **Fig. 1** seems to have satisfied all his material needs - for food, social contact, safety, etc. - and therefore, a perfectly rational action that he may chose to take is to conserve his energy and do nothing at all - wait quietly until new primary needs arise. And yet, the agent is intrinsically motivated to explore – and he expends time and effort to open a closed door (answer a question), and discover new parts of his environment that were not suspected before.

As shown in this example, a hallmark of curiosity is that it generates not a random but a structured pattern of investigation. The agent in our cartoon is not interested indiscriminately in all the information that surrounds him, but becomes curious about specific items. Work in machine learning and robotics clearly shows that, because natural environments contain many possible tasks, including learnable tasks of various levels of complexity, and *unlearnable* or impossible tasks, agents cannot assign "intrinsic" value to all sources of information as is sometimes claimed in the literature. An indiscriminate strategy of examining all the available information would result in collecting disparate pieces of information with nearly no discovery of useful structures, especially given the limited time and energy available over a biological life span. Therefore, a successful curiosity mechanism must assign value to possible endeavors in a way that maximizes the agent's knowledge of, and

ability to predict his environment over vast portions of the learning space and longer time scales.

Computational studies show that , in environments that change quickly and/or continuously, curious individuals can gain an advantage by acquiring new skills and discovering new environmental structures (Singh et al., 2010) (Barto, 2013). However, this long-term (evolutionary) advantage cannot specify the agent's actions on shorter time scale. Therefore, the key question that must be addressed to understand curiosity is how the brain generates interest in *specific items* in a way that maximizes the long-term advantage that can accrue from intrinsically motivated exploration.

Emerging neuroscientific evidence, to which we turn next, suggests that implementing this system requires the concerted action of dopaminergic systems implicated in value and motivation, and cortical systems mediating cognitive processes of memory and attention.

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 during this epoch included the left caudate nucleus, bilateral inferior frontal gyrus (IFG), and loci in the putamen and globus pallidus (**Fig. 2**). 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, experienced as a feeling of curiosity, is encoded in some of the same structures that evaluate material gains.

Two recent studies extend this result in non-human primates by reporting 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 the study on DA cells the subjects 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 (**Fig. 3A**). 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").

The key feature of the behavioral task was that 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), and therefore there was no biological imperative for the monkeys to choose either option. Nevertheless, the monkeys developed a reliable and consistent preference for choosing the informative cue. A subsequent study of area OFC extended showed that the monkeys will chose 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). Therefore the monkeys seem to have a preference for advance reward information that is intrinsically motivated – i.e., it driven by some cognitive or emotional factor that assigned higher value to the predictive cue.

Dopamine neurons encoded both reward prediction errors and the anticipation of reliable information (**Fig. 3B**). The neurons' responses to reward prediction errors arose *after* the monkeys' choice, when the selected target delivered its reward information. At this time, (marked "Cue" in **Fig. 3B**) 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). More interesting, however, were responses to anticipated information gains that arose *before* the monkeys' choice and could contribute to motivating that choice. At the time marked "Target" in **Fig. 3B** 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 independent of the final outcome and seemed to encode enhanced arousal or motivation associated with the informative option. A followup study showed that responses to anticipated information gains are also found in the OFC, and are 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. 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 evaluate the meaning and usefulness of the information.

Eye movements and attention

Although DA neurons and DA-recipient structures are critical for signaling value and motivation, they are not sufficient to explain the full scope of information seeking mechanisms. As shown in **Fig. 3B**, the signals that are conveyed by the cells are not specific for individual objects or locations and thus can only signal a change in motivational state but not the decision to focus on a specific option. Second and most important, making this determination requires input from cognitive processes – such as attention, memory, learning – that can evaluate the informational properties of competing options.

To investigate this question, we adapted the procedure that had been used by Kang et al. by incorporating eye tracking (A. F. Baranes, Oudeyer, & Gottlieb, 2015). We presented subjects with trivia questions, asked them to rate their curiosity and confidence in the answer, and tracked their eye movements while they were waiting for and reading the answer. Questions that elicited higher curiosity were associated with faster anticipatory gaze shifts to the expected location of the answer (**Fig. 4A**). The enhancement of anticipatory gaze was specific to variations in curiosity; ratings of confidence or surprise, despite being partly correlated with ratings of curiosity, had the strongest effects after the answer presentation (**Fig. 4A**). The magnitude of the eye movement effect was correlated with measures of curiosity-related personality traits (see also (Risko, Anderson, Lanthier, & Kingstone, 2012)). Finally, machine learning algorithms could read out curiosity states from the eye movement patterns, generalizing across individual observers and relying primarily on the anticipatory orienting of gaze (**Fig. 4B**).

Additional studies (Gruber, Gelman, & Ranganath, 2014)showed that higher curiosity is associated with better memory performance, possibly through enhanced activation of parahippocampal structures and its dopaminergic projections (Gruber et al., 2014; Kang et al., 2009). (Kang et al., 2009)

In addition to generating global signals of arousal and motivation therefore, curiosity recruits cognitive systems related to memory and attention. This recruitment is clearly beneficial in allowing observers to discriminate, encode and retain the valuable information. In addition, it may be critically important in motivating the observers to sample specific items. In the following question we consider some of the epistemic variables by which the brain may generate interest in sources of information.

What motivates curiosity?

Converging evidence suggests that some of the factors that generate curiosity include surprise, novelty, uncertainty about future rewards, and the probability of rewards for of specific items. We review each factor in turn.

Surprise It has long been recognized that, far from being unbiased, the way in which we sample information from complex visual scenes depends strongly on our knowledge and expectations (Vo & Wolfe, 2015). In the hands of a professional magician, the manipulation of expectations can lead to spectacular misdirection and consequent surprise (Rieiro, Martinez-Conde, & Macknik, 2013). Predictive coding theories suggest that expectations play a key role in orienting attention by predicting away redundant information and freeing resources for detecting significant items (K. Friston et al., 2013). This systematic removal of information through active prediction may be critical for allowing us to see – and indeed, survive – as without it we may be overwhelmed by the sheer amount of information that arrives at our senses.

Studies by Itti and Baldi have shown that surprise, defined in the domain of visual features, attracts human saccades during free-viewing exploration (Baldi & Itti; Itti & Baldi, 2009). Using a Bayesian algorithm combined with computational models of vision, the authors simulated the observers' beliefs about the expected distribution of pixel values at

different visual locations, and defined surprise as the extent to which a visual input changed the observers' beliefs (i.e., the divergence between the prior and posterior beliefs). They showed that this quantitative metric could predict human free-viewing patterns with greater fidelity and flexibility relative to simpler intensity contrast-based predictors. As the authors emphasize, surprise differs from standard measures of Shannon information in that it ascribes central importance to the observers' beliefs rather than being defined purely by the entropy of a stimulus set. This makes it very clear that it is not the mere presence of information that attracts our attention, but the extent to which the information confirms or violates our prior expectations.

Novelty, in contrast with surprise, is not context-specific but is defined by the total amount of exposure that observers had to a given observation. Novelty can be modeled mathematically as the dissimiliary between a stimulus and the representation of familiar stimuli encoded in the observer's memory (Barto, Mirolli, & Baldassare, 2013).

In a classical approach to reinforcement learning (RL), novelty is thought to act as an internal reward that is equivalent to extrinsic rewards. Consistent with this view, novel stimuli activate midbrain dopaminergic structures in humans and other animals (Horvitz, 2000; Laurent, 2008; Bianca C Wittmann, Bunzeck, Dolan, & Düzel, 2007; B. C. Wittmann, Daw, Seymour, & Dolan, 2008), and provides a bonus for organizing reward-based exploration (Barto et al., 2013; Brafman & Tennenholtz, 2003; Kakade & Dayan, 2002; Laurent, 2008; Manuel Lopes & Oudeyer, 2012).

However, a study in our laboratory suggests that novelty also recruits attentional resources through reward-independent effects (Foley, Jangraw, Peck, & Gottlieb, 2014; Peck, Suzuki, Efem, & Gottlieb, 2009).

To examine the impact of novelty and reward in guiding attention, we trained monkeys on a task in which they received visual cues that could be highly familiar or novel and could bring "good" or "bad" news – i.e., signaled whether the trial will end with a reward or a lack of reward (**Fig. 5A**). After presentation of the cue at a peripheral location, the monkeys maintained fixation for a brief delay and then made a saccade to a *separate* target that could appear either at the same or at the opposite location as the cue. In this task therefore, the cues did not allow the monkeys to choose the trial's outcome. Instead, they only brought information and could automatically bias attention toward or away from their visual field location.

We recorded the activity of visually responsive neurons in the lateral intraparietal area (LIP), a cortical area which, together with the frontal eye field (FEF), is implicated in the selection of targets for attention and gaze (Bisley & Goldberg, 2010) (**Fig. 5B**). LIP neurons have visual receptive fields (RF), selectively encode the locations of attention-worthy items, and are thought to provide top-down signals for orienting attention and eye movements (saccades) toward those locations (ibid).

In our task using reward cues, LIP neurons had sharp visual responses if a reward cue appeared in their RF, suggesting that both positive and negative cues transiently

attracted attention (**Fig. 5C**) (Foley et al., 2014; Peck et al., 2009). However at slightly longer delays, the orienting response in LIP changed according to the reward signaled by the cue. The neurons maintained slight excitation for a familiar cue that brought good news (Fam+), but developed sustained *inhibition* for a familiar cue that signaled bad news (Fam-). Consistent with these neuronal responses, saccades were facilitated if they were directed toward the location of a positive cue (which was excited in the LIP representation) and impaired if they were directed toward the location of a negative cue (which was suppressed in LIP). These neural and saccadic effects were spatially specific, occurring at the cue location but not at the opposite visual field location. That is, beyond producing global changes in arousal or motivation, the reward cues facilitated or impaired attentional processing of *specific* sources of information.

Comparison of novel and familiar cues showed that these reward-dependent attentional effects were much weaker or absent for newly-learnt items. Instead LIP neurons showed enhanced responses to novel visual cues, and this novelty enhancement persisted for dozens of presentations for cues that signaled negative or positive outcomes (**Fig. 5D, right**). When they were first confronted with a novel cue, the monkeys showed anticipatory licking, indicating that they expected to receive a reward following the cue, but this licking quickly extinguished if a cue turned out to signal a negative outcome (Foley et al., 2014) (**Fig. 5D**, left). Strikingly, however, the Nov- cue continued to produce enhanced salience and LIP responses, suggesting that it continued to attract attention even after the monkeys had learnt its negative reward associations (**Fig. 5D, right**).

The results suggest that novelty exerts its effects through multiple pathways, both by activating motivational systems and through reward-independent visual/attentional effects. Understanding how these processes work in concert will be an important topic for future investigations.

Reward and uncertainty While the factors of novelty and surprise that we discussed above can engage attention independently of the observer's task, attention can also be controlled in a top-down fashion – i.e., tightly focused on achieving a goal. Since the early studies of Yarbus in the 1950s it has been appreciated that, when observers are engaged in a task, their eye movements are directed very selectively to task-relevant stimuli with very few glances to salient distractors, revealing the strength and importance of task-related control (Tatler, Hayhoe, Land, & Ballard, 2011).

Some insight into the computational basis of task-related control comes from studies of naturalistic behaviors where subjects perform tasks such as driving in virtual reality settings (Gottlieb, Hayhoe, Hikosaka, & Rangel, 2014; Hayhoe & Ballard, 2014; Sullivan, Johnson, Rothkopf, Ballard, & Hayhoe, 2012{Hayhoe, 2014 #9266{Gottlieb, 2014 #9306). Behavior in such contexts was computationally analyzed using reinforcement-learning models that partition the subjects' actions into discrete sub-tasks; for instance, while driving, one may have to coordinate between the sub task of monitoring the speed and that of monitoring the road. These models suggest that gaze is allocated to competing sub-tasks based on two factors: the rewards and informational demands (uncertainty) of each individual task. This dual control mechanism allows subjects to direct gaze efficiently - to

inform those actions that are not only valuable for achieving a goal but also have uncertainty and need for information (Hayhoe & Ballard, 2014; Sullivan et al., 2012; Tatler et al., 2011).

Remarkably, whereas these models were developed to explain gaze deployment based on extrinsic motivation, a recent study in our laboratory suggests that they may also be applicable to the intrinsically motivated – curiosity based - sampling of information (Daddoua, Lopes, & Gottlieb, 2016). To demonstrate this link we used a variant of an observing paradigm, where monkeys received juice rewards probabilistically on each trial and could search for cues that provided precise information about the reward (**Fig. 6A**). To examine the logic of intrinsic motivation, we did not oblige the monkeys to engage in the search; instead, the rewards of each trial arrived at a fixed time according to a pre-ordained probability, regardless of whether or not the monkeys had searched for the cue.

We reasoned that, if the monkeys' intrinsic motivation was sensitive to uncertainty and reward likelihood, it should depend on their beliefs about the trial's reward probability. If the monkeys were motivated to reduce uncertainty, their search should be most vigorous when they believed that the trial had an uncertain outcome (e.g., 50% chance of a reward or a lack of reward). If on the other hand, they were motivated by higher reward likelihood, they should search most vigorously when they believed the trial had a high reward probability (see the legend to **Figure 6A** for the task details).

Consistent with findings from instrumental behaviors, we found that both factors affected the monkeys' intrinsically motivated search (**Fig. 6B**). The monkeys searched more vigorously for the 2nd cue when they had a 50% relative to 100% prior reward expectation, consistent with a desire to reduce uncertainty. However, the monkeys also searched more vigorously for the 2nd cue when they had *no uncertainty* but *high reward expectations* (100% versus 0% likelihoods). Control analyses established that this behavior could not be explained by spurious factors, such as residual uncertainty due to incomplete learning of the cues or changes in arousal or motivation.

These findings suggest a remarkable parallel between information sampling in operant (task-related) and non-operant settings, and also resonate with dual-process psychological theories proposing that curiosity arises both from a desire to close "information gaps" (reduce uncertainty, or harvest information), and as a mere feeling of "interest" or "liking" of pleasurable items (Litman, 2007; Lowenstein, 1994). The "liking" component of these theories is consistent with the reward-based effect in **Fig. 6B**, and with a rich literature showing that animals automatically approach and attend to positively conditioned Pavlovian cues (Castro & Berridge, 2014; Dayan, Niv, Seymour, & Daw, 2006; Flagel et al., 2011; Hickey, Chelazzi, & Theeuwes, 2010a, 2010b; Peck et al., 2009). Our findings extend this literature by showing that Pavlovian cues remain effective even when they bring no new information, and that they not only elicit reactive orienting actions when they appear in a visual display, but act as a source of intrinsic motivation that drive animals to work for finding the cues. Therefore, a description of curiosity as being motivated both by information gains and conditioned reinforcement from pleasurable cues may have broad validity in diverse settings.

While fully acknowledging the parallels between curiosity and instrumental settings, it is important to consider how the underlying computations differ between the two settings. In an instrumental, task-related context, the value functions guiding the subjects' actions are based on the material gains that are expected to accrue from the action. In a non-instrumental setting by contrast, the actions bring no material gains and the value functions must be based on the expected *epistemic or emotional consequences* of performing that action.

Consistent with this view, reinforcement learning (RL) simulations of the search task in **Fig. 6A** show that internal value functions based only on extrinsic gains produced no visual search - as the monkeys could not increase their extrinsic rewards by observing a cue (**Fig. 6C**, solid gray trace in the upper left panel). Internal value functions that assigned intrinsic value to reducing uncertainty (an epistemic gain) *or* to viewing a positive cue (an emotional gain) only partially explained the results (**Fig. 6C**, upper right panels), while functions that combined both factors replicated the monkeys' search pattern (**Fig. 6C**, lower two panels).

In sum, our results suggest that systems of intrinsic motivation represent elaborations upon systems of instrumental sampling, which assign intrinsic value to epistemic or emotional factors and motivate subjects to act independently of material gains(Castro & Berridge, 2014; Dayan et al., 2006; Flagel et al., 2011; Hickey et al., 2010a, 2010b; Peck et al., 2009)

Learning progress and meta cognition The four factors we reviewed above - novelty, surprise, reward and uncertainty – can in principle act in combination and produce a rich range of exploratory actions. However, several considerations suggest that these factors are still not sufficient to explain the full range of curiosity based exploration.

Novelty and surprise are important heuristics for arousing curiosity, but they have the limitation that they do not necessarily signal significant or learnable environmental properties. A curiosity system that is based only on searching for novelty and surprise would only produce what early researchers called "diversive curiosity" (Lowenstein, 1994) – the type of transient curiosity we may show when we browse the internet with no specific aim – but cannot explain more deliberate, sustained investigative actions.

Reward and uncertainty can produce longer-lasting effects, but they have the critical limitation that they are defined in a narrow range of conditions and are typically *not known* to agents before the exploration. Whereas in the non-instrumental tasks we described above (**Fig. 3** and **Fig. 6**) the monkeys were extensively trained and explicitly informed about the reward and uncertainty involved in a task, in more realistic settings subjects begin exploring with only vague estimates of these quantities. As illustrated in the example shown in **Fig. 1**, an agent cannot know the payoffs or uncertainty associated with opening the door and critically important, cannot even expect that opening the door will *reduce his uncertainty*. Formal computational studies show that heuristic mechanisms that motivate agents to explore situations of high uncertainty do not guarantee any learning at all and conversely, heuristics that motivate agents to *minimize* uncertainty will cause them to focus exclusively

on well-learnt, predictable tasks. In more general terms, such studies confirm that searching for novelty, high or low uncertainty, or high or low entropy, may be useful in well-delimited contexts, but are inefficient in acquiring knowledge and skills in unbounded large spaces or spaces with unlearnable tasks (Baranès & Oudeyer, 2009; M. Lopes & Montesano, 2014; Manuel Lopes & Oudeyer, 2010). To understand the full range of our curiosity therefore, we must account for the coexistence of two conflicting drives: on one hand, the desire to reduce uncertainty on a short time scale and on the other hand, the *intellectual risk taking* that motivates us to increase uncertainty in order to learn on longer time scales.

To address these shortcomings, studies of artificial curiosity have developed computational strategies based on a meta-cognitive mechanism that assigns value to competing tasks based on the empirical learning progress (LP) related to each task (A. Baranes & Oudeyer, 2013; Manuel Lopes, Lang, Toussaint, & Oudeyer, 2012; Moulin-Frier, Nguyen, & Oudeyer, 2013; P. Y. Oudeyer, F. Kaplan, & V. V. Hafner, 2007; Schmidhuber, 1991; Srivastava, Steunebrink, & Schmidhuber, 2013). Manuel Lopes, Lang, Toussaint, & Oudeyer, 2012 proved that exploration based on learning progress is equivalent to methods based solely on the number of visits (e.g. Brafman, R. I., & Tennenholtz, M. (2003)) but becomes more robust when encountering changing situations or having the wrong expectations.

LP-driven mechanisms prioritize competing tasks based on the rate of improvement - derivative - of the cost function that the learner is trying to maximize. LP can be defined based on the rate of improvement of predictions of a sensorimotor outcome or of the reward/success rate in a task. Compared to heuristics that search for high uncertainty, LP-driven mechanisms will motivate the learner to investigate situations that are initially uncertain and *keep exploring* them *only if* these situations lead to learning in practice. This can be formulated as an operational implementation of the information-theoretic framework of the free-energy principle (Karl Friston et al., 2015).

An example of algorithmic architecture implementing an LP-driven curiosity process is the R-IAC architecture, detailed in **Fig. 8B**. In this architecture, a robot learns to predict the consequence of its actions. Such predictive learning is made with statistical inference over the data collected when the robot carries out « experiments », i.e. tries an action and observes the results. The robot then chooses which task to perform based on a meta-cognitive module that monitors the evolution of prediction errors in various regions of the sensorimotor space: it selects regions to explore with a probability that is proportional to the rate of improvement in the past (such probabilistic scheme allows to continually search for new niches of progress).

In one study we showed that such an architecture allows a robot to master hand-eye coordination much faster relative to strategies based on random exploration or a search for maximal uncertainty (Baranès & Oudeyer, 2009). Similar results were shown for the acquisition of other skills such as omnidirectional legged locomotion (A. Baranes & Oudeyer, 2013) or the manipulation of flexible objects (Nguyen & Oudeyer, 2013)

Interestingly, these analyses showed that, in addition to providing very efficient for acquiring new skills in large task spaces, LP-based algorithms produce exploration strategies that spontaneously progress from simple to more complex tasks in the absence of external instructions. For example, in the Playground Experiment (**Fig. 7A**) several behavioral and cognitive phases spontaneously formed during learning. After a phase of random body babbling, the robot focused on moving only certain body parts, and then focused on increasingly complex action-object affordances - beginning by learning how its leg can push or grasp objects, and ending up exploring how its vocalizations could produce reactions in another robot. Repeated runs of this experiment showed that in many cases similar developmental milestones appeared in a similar order while other robots showed deviations from these milestones or went through them in a different order, similar to the dual properties of universal tendencies and diversity seen in the development of infants (P.-Y. Oudeyer, F. Kaplan, & V. V. Hafner, 2007; Oudeyer & Smith, in press).

In a related experiment on vocal development, robots used an LP-based algorithm to discover how to communicate with peers (Moulin-Frier et al., 2013). This experiment relied on a physical model of the vocal tract, its motor control and the auditory system, and showed how such a mechanism can explain the adaptive transition from vocal self-exploration with little sensitivity to 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-organized, and shared properties with infant data: the vocal learner first discovered how to control phonation, then vocal variations of unarticulated sounds, and finally articulated proto-syllables. As the vocal learner becomes more proficient at producing complex sounds, the imitating vocalizations of the teacher provide high learning progress resulting in a shift from self-exploration to vocal imitation.

One can apply such automatic organization of learning in intelligent tutoring systems. These systems' goal is to provide automatic assistance to learners based on their skills. As each student will have their own background, particular strengths and weakness, a general model equal for all students will not accurately predict the behavior of any particular student. Clement, Roy, Oudeyer and Lopes, 2015 proposed the use of learning progress measures to allow an intelligent tutoring system to adapt to the particular learning progression of each individual student and showed that different paths provide a faster learning.

In addition to being a hallmark of developmental progression, the self-organization of behavior from simple to more complex tasks is a cornerstone of theories of intrinsic motivation in personality research in adults (Ryan & Deci, 2000a, 2000b). In a recent study we replicated the latter effect in a laboratory setting by using a task where subjects were given a set of computer games of variable complexity and could freely choose the games they wished to play (A. F. Baranes, Oudeyer, & Gottlieb, 2014). A game lasted several seconds and required subjects to press a key as accurately as possible to intercept a series of dots that streamed past the center of the screen (**Fig. 8A**). Even though the subjects received no instruction about which game to select, they spontaneously organized their exploration in consistent patterns. Subjects did an initial survey of the entire space of the available games – including the most difficult games where dot speed were very high and performance was low – and then focused their exploration on games of intermediate complexity, where performance was 70-80% correct (**Fig. 8B**). This general trend was

modulated by factors such as how much novelty could be found in games of a given complexity, and how the difficulty of tasks was spread along the game distribution.

Therefore, behavioral evidence is consistent with the idea that a self-organizing pattern based on task complexity shapes intrinsically motivated behaviors in a variety of contexts. However, more evidence is needed to establish whether this pattern indicates an LP-based mechanism. Many forms of learning are nonlinear in time (showing effects such as savings and consolidation) and it is unclear whether subjects can accurately track their learning progress or which aspects of progress determine intrinsic motivation. Addressing these questions will be critical for a better understanding of our most elaborate curiosity based forms of exploration.

Conclusions

We reviewed a theory of curiosity that emerges from a synthesis of findings from neuroscience and machine learning fields. A central theme that we stressed throughout the review is the fact that curiosity implies a tight interaction between cognitive and motivational systems. Rather than being an external process that acts *on* learning and cognition, curiosity arises organically in conjunction with these processes. When it is engaged in cognitive processing, the brain does more than simply discriminate, encode, and remember information – it also evaluates the epistemic and emotional qualities of these cognitive operations and uses them to generate the "interest" and intrinsic motivation that determine its future engagement with a specific task.

We have also stressed the fact that the mechanisms that generate curiosity would be ideally adapted to allow agents to discover new and useful regularities in large open-ended spaces that contain unlearnable tasks. This is a formidable challenge for which agents may have no optimal solution, and we proposed that organisms meet this challenge by combining a variety of strategies. These strategies include simple heuristics such as exploration based on novelty, surprise, reward and uncertainty that may have their roots in simpler active sensing behaviors. In addition, they may include more complex targeted investigations potentially based on meta-cognitive estimates of learning progress and information gain. These mechanisms may act in concert to autonomously organize exploration of vast unbounded spaces, steering agents away from overlearnt (low uncertainty) tasks and away from unlearnable (high uncertainty) tasks, toward a middle range where the agent can make learning progress and discover new structures. While many of the views we outlined remain to be refined and substantiated through future research, we hope that they provide a useful road map for identifying the important questions to be addressed in that research.

Figure legends

Figure 1. **Cartoon** depicting the essential qualities of curiosity.

Figure 2. Brain regions that showed differential activity in high- versus low-curiosity trials during the first question presentation in (Kang et al., 2009). Colored areas showed greater

anticipatory activation on high-curiosity trials in experiment 1 ($p<.001$ uncorrected, prep >.99, extent threshold 5) using a median-split analysis (red), the modulator analysis (yellow), and the analysis of residual curiosity (green). The illustration at the right is a close-up view of the overlapping caudate activations. Ant, anterior; Pos, posterior; L, left; R , right; IFG, inferior frontal gyrus. Reproduced with permission from (Kang et al., 2009).

Figure 3 Dopamine neuron responses in an information choice task A The task sequence. On each trial after achieving central fixation monkeys viewed a target prefacing an informative (green) or uninformative (orange) cue. Single target trials (top and bottom) were interleaved with 2 -target trials where monkeys were free to select the target they wished to view. If monkeys shifted gaze to the informative target (green) they were shown two subsequent cues that were consistently associated with, respectively, a large or small water reward. If monkeys shifted gaze to the uninformative target (ora nge) they were shown two other cues that were inconsistently associated with the large or small rewards (50% predictive validity). The large and small rewards were equally likely to occur, so that the informative and uninformative targets had equal expected rewards. **B Neural responses of DA cells on the information choice task** The traces show average activity in a population of DA cells, aligned on the time of target presentation, appearance of the reward cues and delivery of the final reward. At the time of target presentation the neurons had stronger responses when the display contained an informative target (dark and light red traces) than when it only contained the uninformative target (blue). After the information was revealed (cue) DA neurons had the expected reward prediction response. At the time of cue presentation they had excitatory and inhibitory responses to, respectively, the high and low reward predictive pattern, and small excitatory responses to the uncertain pattern announcing a 50% probability of reward. At the time of the reward the neurons had excitatory and inhibitory responses upon receipt of, respectively, the large and small reward, but only if this reward was unpredicted (i.e., upon selection of the uninformative cue). Reproduced with permission from (Bromberg-Martin & Hikosaka, 2009).

Figure 4: Curiosity affects eye movements A. Ocular anticipation in relation to curiosity For each trial with high or low ratings we computed the distance between the eye position and the left edge of the answer box every 2 ms. Distances were averaged for each subject, and we display the mean and SEM across subjects. Average distances before and after answer onset were compared with a 1-way ANOVA; stars show ** $p<10^{-45}$, ***$p<10^{-75}$. B **Classification accuracy for different implementations of a machine learning algorithm. Left:** Classification across the entire data set. **Middle.** Classification with across-subject cross-validation. **Right.** Same as **the middle panel** but using only the 15 most informative features. In **the middle and right panels** the open points show individual subject predictions and the black points and bars show average and SEM. Reproduced with permission from (A. F. Baranes et al., 2015).

Figure 5. **Independent effects on reward and novelty on visual salience** (**A) Task design** A trial began when the monkeys fixated a central fixation point (small black dot). A reward cue was then presented for 300ms at a randomly selected location that could fall inside the RF of an LIP cell (gray oval) or at the opposite location (for simplicity, only the RF

location is illustrated). The cue could fall into one of 4 categories depending on whether it was familiar (Fam) or novel (Nov) and signaled a positive (+) or a negative (-) outcome. The cue presentation was followed by a 600ms delay period during which the monkeys had to maintain fixation ("Delay"), and then by the presentation of a saccade target at the same or opposite location relative to the cue. If the monkeys made a correct saccade to the target they received the outcome predicted by the cue – a reward on Nov+ and Fam+ trials, but no reward on Nov- and Fam- trials. Trials with incorrect saccades were immediately repeated. (**B) Cortical oculomotor areas** Lateral view of the macaque brain showing the approximate locations of the FEF and LIP. **(C) LIP neurons are modulated by reward and novelty** Normalized activity (mean and standard error (SEM)) in a population of LIP cells, elicited by cues that appeared in the RF and which could be familiar or newly-learntand bring "good news" (predicting a reward; Nov+ and Fam+), or bring "bad news" (predicting a lack of reward; Nov- and Fam-). The cues appeared for 300 ms (thick horizontal bar) and were followed by a 600 ms delay period during which the monkeys maintained fixation. The familiar cues showed strong reward modulations, with Fam- cues evoking a lower visual responses and sustained delay period suppression that was not seen for Fam+ cues. However, newly-learnt cues elicited stronger overall responses and weaker reward modulations. In particular, Nov- cues did not elicit the sustained suppression seen for the Fam- cues. **(D) Learning of cue-reward associations as a function of the number of cue exposures during a session**. The points show the duration of anticipatory licking and the normalized visual response (during the visual epoch, 150-300ms after cue onset) as a function of the number of cue exposures during the session. Error bars show SEM. Anticipatory licking for the Nov- cues declined rapidly but the visual response elicited by the Nov- cue remained high throughout the session. Although the monkeys rapidly learn negative cue-reward associations, they are slower to reduce the salience of a "bad news" cue. Reproduced with permission from Peck et al., 2009 (panel B) and Foley et al. 2014 (panels C,D).

Figure 6. A The *Active search task* The monkey initiated each trial by fixating a central point and maintaining gaze on it while cue1 was shown in the periphery for 0.3 s. After an additional 1 second of fixation, the fixation point was replaced with a search display containing 3 white placeholders. The monkeys could freely examine during a 2 second "free search" period, when maintaining gaze within a 2 degree window centered on a placeholder for 300 ms caused it to reveal the underlying pattern (a gray square or an additional cue). In the example illustrated, the monkey first uncovered an uninformative gray square and later found cue 2 at the middle location. The search display then disappeared, and after an additional 1.3 s delay (blank screen presentation) , the trial ended with a tone that was accompanied by the outcome (a reward or a lack of reward according to the probability signaled by cue 1). The search behavior was entirely unconstrained and had no bearing on the final outcome. **B Behavior as a function of the reward probability signaled by cue 1** In the top panel the points show the mean and standard errors across all testing sessions, after z-scoring across cue type within individual sessions. In the bottom panel we computed the probability of finding cue 2 in each session, and show the mean and standard errors of these probabilities, z-scored across all sessions. Stars indicate $p < 0.025$ (Wilcoxon test). The **insets** in each panel show the average of the raw data per session. The dotted red trace indicates 0% cue 1, the solid red trace shows 100% cue 1 and the solid blue trace

shows 50% cue 1. **C RL model simulations. (Top left)** Search behavior (the number of revealed placeholders) shown by a model that incorporates only an operant reward (gray solid trace), only an information reward (black solid trace), or information reward combined with an operant reward (black dotted). **(Top right)** Left panel: Search behavior shown by models that incorporate only an Pavlovian reward (solid trace) or a Pavlovian and operant reward (dotted trace). Right panel: the effect of increasing the weight of the Pavlovian reward. **(Bottom left)** Left panel: Search behavior shown by models that incorporate all 3 reward components, when the information component depends only on the entropy of the reward distribution (solid trace), or on both the reward and visual distributions (dotted trace). Right panel: the effect of increasing α_p, the weight of the Pavlovian reward. **(Bottom right)** Deep or shallow search curves (similar to the behaviors shown by the two monkeys) produced with different parameters of the 3-component model. In all the panels, each point shows the mean and SEM over 100 action selection iterations.

Figure 7: **Spontaneous curiosity-driven exploration can be efficiently driven by searching for situations that improve current predictions (LP), and self-organize a learning curriculum of increasing complexiy. A The Playground Experiment :** a quadruped robot placed on an infant play mat with a set of nearby objects, as well as an 'adult' robot peer. The robot is equipped with a repertoire of motor primitives parameterized by several continuous numbers, which can be combined to form a large continuous space of possible actions. The robot learns how to use and tune them to affect various aspects of its surrounding environment, and exploration is driven by maximization of learning progress using the R-IAC architecture. We observe the self- organization of structured developmental trajectories, whereby 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. The robot also discovers primitive vocal interaction as a result of the same process. **B** The R-IAC architecture implements this curiosity-driven process with several modules. A prediction machine (M) learns to predict the consequences of actions taken by the robot in given sensory states. A meta-cognitive module (metaM) estimates the evolution of errors in prediction of M in various subregions of the sensorimotor space, which in turn is used to compute learning progress as an intrinsic reward. Because 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 incrementally model the creation and refining of activities/tasks. Then an action selection system chooses activities to explore for which estimated learning progress is high. This choice is stochastic in order to monitor other activities for which learning progress might increase. **C** Confronted with four sensorimotor activities characterized by different learning profiles (i.e., evolution of prediction errors), exploration driven by maximization of learning progress results in avoidance of activities already predictable (curve 4) or too difficult to learn to predict (curve 1) to focus first on the activity with the fastest learning rate (curve 3) and eventually, when the latter starts to reach a plateau, to switch to the second most promising learning situation (curve 2). This allows the creation of an organized exploratory strategy necessary to engage in open-ended development. Adapted with permission from (REF).

Figure 8: Intrinsically motivated exploration in a laboratory game. A: Task design. The top panel shows an individual game lasting ~30 seconds, in which subjects pressed a key to intercept a stream of moving dots (arrow) as they crossed the screen center. **The bottom panel shows the selection screen** with 64 games from which subjects could freely choose which game they wished to sample. **B The speed (difficulty) of the selected games during a session.** The colormap indicates the probability of selection of a given speed, measured across all subjects in a sliding window over the session**.** Adapted with permission from (A. F. Baranes et al., 2014).

References

Baldi, P., & Itti, L. Of bits and wows: A Bayesian theory of surprise with applications to attention. *Neural Netw, 23*(5), 649-666.

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

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

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*(Oct. 14).

Baranes, A. F., Oudeyer, P. Y., & Gottlieb, J. (2015). Eye movements encode epistemic curiosity in human observers. *Vis Res, in press*.

Barto, A., Mirolli, M., & Baldassare, G. (2013). Novelty or surprise? *Frontiers in Psychology, 11 December*.

Bisley, J., & Goldberg, M. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33*, 1-21.

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

Brafman, R. I., & Tennenholtz, M. (2003). R-max-a general polynomial time algorithm for near-optimal reinforcement learning. *The Journal of Machine Learning Research, 3*, 213-231.

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

Castro, D. C., & Berridge, K. C. (2014). Advances in the neurobiological basis for food "liking" versus "wanting". *Physiology and Behavior, 136*, 22-30.

Daddoua, N., Lopes, M., & Gottlieb, J. (2016). Intrinsically motivated visual exploration is sensitive to reward and uncertainty in non-human primates. *Scientific Reports, in press*.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw, 19*(8), 1153-1160.

Flagel, S. B., Clark, J. J., Robinson, T. E., Mayo, L., Czul, A., Willuhn, I., et al. (2011). A selective role for dopamine in stimulus-reward learning. *Nature, 469*(7328), 53-57.

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.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive neuroscience*, 1-28.

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front Hum Neurosci, 7*, 598.

Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, reward and information seeking. *Journal of Neuroscience, 34*(46), 15497-154504.

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron, 84*(2), 486-496.

Hayhoe, M., & Ballard, D. (2014). Modeling task control of eye movements. . *Current Biology, 24*(13), 622-628.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010a). Reward changes salience in human vision via the anterior cingulate. *Journal of Neuroscience, 30*(33), 11096-11103.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010b). Reward guides vision when it's your thing: trait reward-seeking in reward-mediated visual priming. *PLoS One, 5*(11), e14087.

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience, 96*(4), 651-656.

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

Kakade, S., & Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Netw, 15*(4-6), 549-559.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., et al. (2009). The wick in the candle of learning: epistemic curiosity activates reward circuitry and enhances memory. *Psychol Sci, 20*(8), 963-973.

Laurent, P. A. (2008). The emergence of saliency and novelty responses from Reinforcement Learning principles. *Neural networks : the official journal of the International Neural Network Society, 21*(10), 1493-1499.

Litman, J. A. (Ed.). (2007). *Curiosity as a feeling of interest and feeling of deprivation: the I/D model of curiosity*: Nova Science Publishers, Inc.

Lopes, M., Lang, T., Toussaint, M., & Oudeyer, P.-Y. (2012, 2012). *Exploration in model-based reinforcement learning by empirically estimating learning progress.* Paper presented at the Neural Information Processing Systems (NIPS 2012).

Lopes, M., & Montesano, L. (2014). Active learning for autonomous intelligent agents: exploration, curiosity and interaction *arXiv: 1403.1497 [ca.AI]*.

Lopes, M., & Oudeyer, P.-Y. (2010). Guest editorial active learning and intrinsically motivated exploration in robots: Advances and challenges. *IEEE Transactions on Autonomous Mental Development, 2*(2), 65-69.

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

Lowenstein, G. (1994). The psychology of curiosity: a review and reinterpretation. *Psychological Bulletin, 116*(1), 75-98.

Moulin-Frier, C., Nguyen, S. M., & Oudeyer, P.-Y. (2013). Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Frontiers in psychology, 4*.

Nguyen, S. M., & Oudeyer, P.-Y. (2013). Socially guided intrinsic motivation for robot learning of motor skills. *Autonomous Robots*.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *Evolutionary Computation, IEEE Transactions on, 11*(2), 265-286.

Oudeyer, P. Y., Kaplan, F., & Hafner, V. V. (2007). Instrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computations, 11*(2), 265-286.

Oudeyer, P. Y., & Smith, L. (in press). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*.

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.

Rieiro, H., Martinez-Conde, S., & Macknik, S. L. (2013). Perceptual elements in Penn & Teller's "Cups and Balls" magic trick. . *PeerJ 1:e19*.

Risko, E. F., Anderson, N. C., Lanthier, S., & Kingstone, A. (2012). Curious eyes: Individual differences in personality predict eye movement behavior in scene-viewing. *Cognition, 122*, 86-90.

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp Educ Psychol, 25*(1), 54-67.

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychology*

55(1), 68-78.

Schmidhuber, J. (1991). *Curious model-building control systems.* Paper presented at the IEEE International Joint Conference on Neural Networks.

Srivastava, R. K., Steunebrink, B. R., & Schmidhuber, J. (2013). First experiments with PowerPlay. *Neural Networks, 41*, 130–136.

Sullivan, B. T., Johnson, L., Rothkopf, C. A., Ballard, D., & Hayhoe, M. (2012). The role of uncertainty and reward on eye movements in a virtual driving task. *J. Vis., 12*(13).

Tatler, B. W., Hayhoe, M. N., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: reinterpreting salience. *J Vis, 11*(5), 5-25.

Vo, M. L., & Wolfe, J. M. (2015). The role of memory for visual search in scenes. *Ann NY Acad Sci, 1339*, 72-81.

Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Düzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *NeuroImage, 38*(1), 194-202.

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron, 58*(6), 967-973.

Benjamin Clement, Didier Roy, Pierre-Yves Oudeyer, Manuel Lopes, Multi-Armed Bandits for Intelligent Tutoring Systems, Journal of Educational Data Mining (JEDM), 2015, 7 (2), pp.20--48