1

From hardware and software to kernels and envelopes: a concept shift for robotics, developmental psychology and brain sciences.

Frédéric Kaplan (1) and Pierre-Yves Oudeyer (2)

(1) EPFL-CRAFT - CE 1 628 Station 1 CH - 1015 Lausanne SWITZERLAND
(2) INRIA-Futur Bordeaux 351, cours de la Liberation Batiment A29 33405 Talence FRANCE

1.1 From hardware and software to kernels and envelopes

At the beginning of robotics research, robots were seen as physical platforms on which different behavioral programs could be run, like the hardware and software parts of a computer. However, recent advances in developmental robotics have permit to consider a reversed paradigm in which a single software, called a kernel[1], is capable of exploring and controlling many different sensorimotor spaces, called envelopes. In this chapter, we come back on studies we previously published about kernels and envelopes to retrace the history of this concept shift and discuss its consequences for robotic designs but also for developmental psychology and brain sciences.

This chapter is based of previous other studies we published on various aspects of this subject (Kaplan and Oudeyer, 2007b,a, 2008, 2009). Its aim is to reframe these works into a coherent framework in order to give a more global overview of this concept shift. The first section

[1] The term *kernel* is currently used with different meanings in computer science. The term is used here in a different way than in the machine learning community (e.g. kernels of Support Vector Machines)

of this chapter discusses in more details the epistemological transition from the classical dualism that views a robot as fixed body on which different programs can be plugged to the new dualism based on kernel and envelopes. Each important conceptual steps in this evolution is illustrated with concrete examples and experiments. The main point of this first section is to introduce this concept shift and not to define precisely what kind of systems can be considered a kernel and what kind of systems cannot. The kernel is simply generally defined as what is stable across applications and independent of the particular trajectory of one agent. Typically "metalearning" algorithms are good candidate to be part of a kernel as long as they can be considered task and embodiment independent. On the contrary, memory used by the learning systems (weight of neural networks, list of prototypes, data in general) would typically not be part of the kernel. The main goal of this first section is to articulate how the kernel / envelop dichotomy opens the way to new kind of experimental studies. In particular, we discuss the case of generic algorithms capable of learning to control a robotic body without knowing its characteristics beforehand. With such kind of algorithms you can perform experiences where you can precisely characterize the importance of the embodiment in the final behavior obtained, simply by changing the embodiment and keeping the kernel stable.

This new kind of experiments opens different perspective on data obtained by research in developmental science and neuroscience. In the following section, we argue why children development can indeed be seen as a succession of temporary embodiment corresponding not necessarily to physical changes but to the acquisition of new skills. A child learning how to walk or how to play the piano discovers whole new spaces to explore. As he learns, he experiences a kind of metamorphosis. Likewise, to perform basic tasks, his body envelope literately extends itself to include objects, clothes, tools or even vehicles. What stays the same in this developmental and behavioral process is the kernel, origin of the motivation and action of the child. We discuss how this view permits to reinterpret developmental psychology data from the development of sensorimotor dexterity to the acquisition of language.

Eventually, we review different hypotheses about the possible underlying neural substrate for kernels and envelopes. In particular we discuss the putative role of subcortical systems in the process of envelope creation, the possible importance of tonic dopamine as a learning progress signal and the kind of computation that could be performed by microcortical circuits. The chapter ends with the discussion of an evolutionary

scenario illustrating how an old brain circuitry optimized for specific extrinsic needs could have evolved into a subcortical kernel, possibly at the origins of the formidable cortical extension that characterizes the human brain.

1.2 A concept shift for robotics

1.2.1 The reunited body

Between the 1950s and the 1980s, the classical gap between the builders of robotic bodies and the researcher trying to model "intelligence" has some direct consequences on the performances of the machines produced. The AI algorithms, designed to manipulate predefined unambiguous symbols show clearly their inadequacy when it comes to deal with the complexity and the unpredictability of the real world. Consider for instance the problem of programming the walking behavior of a four-legged robot using a classical AI algorithm. The set of joints of a robotic body are not a set of abstract symbols but rather a complex system that can easily end up being in out of equilibrium positions especially if it is made of rigid parts, like most robots are. The type of ground and the degrees of friction have a direct influence on the behavior of the machine. With a symbolic AI approach, but also with many approaches in control theory, it is important that the system is equipped with precise model of the robot body but also on the environment in which the robot evolves. It many cases this is just impossible. Viewed from this angle, walking on four legs can reveal itself to be a harder problem than demonstrating mathematical theorems.

To go out of this dead-end, a new school of thought emerged at the end of the 1980s, with the work of researchers like Rodney Brooks, Luc Steels and Rolf Pfeifer. The so-called embodied artificial intelligence, or new AI, strongly criticized the disembodied and symbolic approach of the "classical" artificial intelligence, claiming that intelligence could not be considered without reference to the body and the environment (Pfeifer and Scheier, 1999). Rodney Brooks added that bodies and environments are impossible to model and that therefore research should not try to build models of external reality but on the contrary concentrate on direct situated interaction: "the world is its own best model" (Brooks, 1999; Steels, 1994).

This change of perspective introduced a renewal of robotic exper-

iments and in someway a return to conception and experimentation methods that were characteristics of robotics *before* the advent of the digital computer. Grey Walter's cybernetic "tortoises" built in 1948 are taken as canonical example of what a good conception is, integrating seamlessly the physical design of the machine to the targeted behavior. These entirely analogical robots were capable of complex behavior, without the need of any internal "representation" (Grey Walter, 1953).Their design was taking into account that they were physical machines, on which many kinds of "forces" had an influence, from gravity to frictions and that perception itself was primarily the result of their own movement and behavior (a concept later known as "enaction" (Varela et al., 1991)). The nature and positioning of their sensors enabled them to solve complex tasks, like returning to their charging station, without the need to make any kinds of complex "reasoning".

Inspired by von Uexkull's writings (von Uexkull, 1909), research of the new AI defined the behavior of their robot taking into account their "Umwelts: the very nature and structure of their body immersed them in a specific ecological niche where certain stimuli are meaningful and others not. This research was also supported by the reappraisal of a non-dualistic philosophical trend which in the tradition of Merleau-Ponty views cognition as being situated and embodied in the world (Merleau-Ponty, 1942, 1945; Varela et al., 1991).

To try to convince the cognitivists to view intelligence only as a form of sophisticated computation, researchers in embodied AI tried to define the kind of *morphological computation* realized by the body itself (Pfeifer and Bongard, 2007).To solve a problem like four-legged walking, it is easier and more efficient to build a body with the right intrinsic physical dynamics instead of building a more complex control system. One can replace the rigid members and powerful motors of the robot by a systems of elastic actuators inspired by the muscle-tendon dichotomy that is typical of the anatomy of quadruped animals. With such a body, one just needs a simple control system producing a periodic movement on each leg to obtain a nice elegant and adapted walking behavior. Once put on a given ground the robot stabilizes itself after a few steps and converges towards its "natural" gait. With such a system, the walking speed can not be arbitrary defined but corresponds instead to attractors of this dynamical system. Only an important perturbation can enable the robot to leave its natural walking gait and enter another attractor corresponding for instance to "trotting" (Pfeifer and Bongard, 2007).

Thus, in an attempt to suppress the gap inherited from the post-war

field division, embodied artificial intelligence emphasized the crucial importance of the body and illustrated its role for the elaboration of complex behavior: body morphological structure and animation processes must be thought as a coherent whole.

1.2.2 Stable kernels

In the beginning of the 1990s, robotic experiments from the new AI perspective focused essentially on reenacting insect adaptive behavior, examples strategically far from the classical AI programs playing chess. In the following years, some researchers tried to extend this embodied approach to build robots capable of learning like young children do. The idea what not to address one particular step in children development (like learning how to walk or how to talk), but to capture the open-ended, versatile, nature of children learning. In just a few months children incrementally learn to control their body, to manipulate objects, to interact with peers and caregivers. They acquire everyday novel complex skills that open them to new kinds of perception and actions. How could a machine ever do something similar? The objective of children-like general learning capabilities was not new as it was already clearly articulated in one of Turing's founding article of artificial intelligence (Turing, 1950). However, the sensorimotor perspective developed by the embodied approach gave to this challenge a novel dimension.

In asking how a machine could learn in an open-ended manner, researchers in epigenetic or developmental robotics (Lungarella et al., 2003; Kaplan and Oudeyer, 2006; Asada et al., 2009) partially challenged the basis of the embodied artificial intelligence approach and introduced a methodological shift. The importance of the body was still central as the focus was on developing sensorimotor skills intrinsically linked with a specific morphology and the structure of a given environment. However, while following an holistic approach, it seemed logical to identify inside a robotic system, a process independent of any particular body, ecological niche or task. Indeed, by definition, a mechanism that could drive the learning of an open-ended set of skills, cannot be specific to a particular behavior, environment or body. It must be general and disembodied.

Thus, the just reunited body must again be divided. But the division is not the one inherited from the punch-cards and the digital computer, the software/hardware gap. In this new methodological dualism, the objective is to separate (1) a potentially changing body envelope corre-

sponding to a sensorimotor space and (2) a kernel, defined as a set of general and stable processes capable of controlling any specific embodied interface. By differentiating a generic process of *incorporation* and fluid body envelopes, the most recent advances in epigenetic/developmental robotics permit to consider the body from a new point of view. Contrary to the traditional body schemata, grounded in anatomical reality, body envelopes are ephemeral spaces associated with a particular task or skill. Contrary to easily changeable animation programs used in robotics, we now consider a stable kernel, acting as an engine driving developmental processes. It is not the body that stays and the programs that change. It is precisely the contrary: the program stays, the embodiment changes.

Several kinds of kernels can be envisioned. Some of them lead to open developmental trajectories, others don't. Let's imagine a control room equipped with a set of measurement devices, a panel of control buttons, and most importantly, *no labels* on any of these devices. Imagine now an operator trying to guess how the whole system works despite the absence of labels. One possible strategy consists in randomly pushing buttons and observing the kind of signals displayed on the measurement devices. However, finding blindly correlation between these inputs and outputs could be very hard. For the operator a better strategy is to identify the contexts in which he progresses in his understanding of the effects of certain buttons and to explore further the corresponding actions.

It is possible to construct an algorithm that drives such kind of smart exploration. Given a set of input and output channels, the algorithm will try to construct a predictive model of the effect of the input on the output, given its history of past interactions with the system. Instead of trying random configuration, the algorithm detects situations in which its predictions progress maximally and chooses the input signal in order to optimize its own progress. Following this principle, the algorithm avoids the subspaces where the outputs are too unpredictable or on the contrary too predictable in order to focus on the actions that are most likely to make it progress (figure 1.1). We call these zones: "progress niches" [2]. The use of such an algorithm results in an organized exploration of an unknown space, starting with the simplest subspaces to progressively explore zones more difficult to model. The term "kernel"

[2] To discover these progress niches, the algorithm must explore regularly the entire space of possible actions. For such exploration the classifical trade-off between exploration and exploitation applies. The algorithm must be programmed to balance the exploitation of the best progress niches and the constant exploration to discover some new ones. Please refer to the appendix for a detailed implementation

is relevant for several reasons to describe the behavior of this algorithm. It is a *central* process, stable, unaffected by the peripheral embodied spaces. It is also the *origin* and the starting point of all the observed behavior.

Details of one version of this progress-driven kernel can be found in (Oudeyer et al., 2007) and also in the appendix of this chapter (see also earlier version in (Kaplan and Oudeyer, 2003; Oudeyer et al., 2005)). Many variants of such kind of intrinsic motivation systems have been or are currently being explored (see (Oudeyer and Kaplan, 2007) for a taxonomy). To our knowledge, the first computational system exploring progress-driven exploration was described by Schmidhuber in 1991 (Schmidhuber, 1991). He suggested to give intrinsic reward to a reinforcement learning controller in proportion to the predictor's error reductions, to motivate the controller to create actions that provoke more data that maximizes the predictor's future cumulative expected learning progress. In following papers, Schmidhuber described various techniques to obtain a similar behavior of the controller (Storck et al., 1995; Schmidhuber, 2006) [3]. Recently, different types of intrinsic motivation systems were explored, mostly in software simulations (Huang and Weng, 2002; Marshall et al., 2004; Steels, 2004). The term "intrinsically motivated reinforcement learning" has been used by Barto in this context (Barto et al., 2004). Interestingly, the mechanisms developed in these papers also show strong similarities with mechanisms developed in the field of statistics, where it is called "optimal experiment design" (Fedorov, 1972).

Coming back to our walking case study, let us now consider an experiment where a progress-driven kernel controls the movement of the different motors. For each motor, it chooses the period, the phase and the amplitude of a sinusoidal signal. The prediction system tries to predict the effect of the different set of parameters in the way the image captured by a camera placed on the robot's head is modified. This indirectly reflects the movement of its torso. At each iteration the kernel produces the values for the next parameter set in order to maximize the reduction of the prediction error (figure 1.2).

When one starts an experiment like this one, several sets of parameters are explored for a few minutes. The robot legs wobble in an apparently disorganized manner. Most of these attempts have very predictable effects: the robot just doesn't move. Errors in prediction stay at a minimal

[3] see his website for a complete list http://www.idsia.ch/ juergen/interest.html

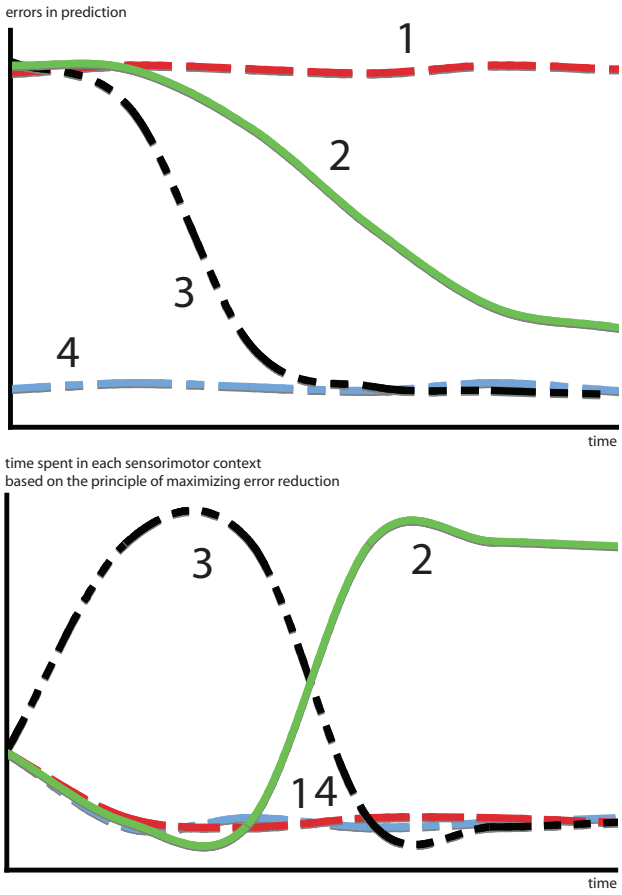

Figure 1.1 Confronted with four sensorimotor contexts characterized by different learning profiles, the exploration strategy of a progress-driven kernel consists in avoiding situations already predictable (context 4) or too difficult to predict (context 1), in order to focus first on the context with the fastest learning curve (context 3) and eventually, when the latter starts to reach a "plateau" to switch to the second most promising learning situation (context 2).

level: these situation are not interesting for the kernel. By chance, after thirty minutes or so, one movement leads the robots to make a slight move, in most cases a step backward. This new situation results first in an increase of the error in prediction but, as the robot experiences similar movements again, this error tends to decrease: the kernel has discovered a "progress niche".

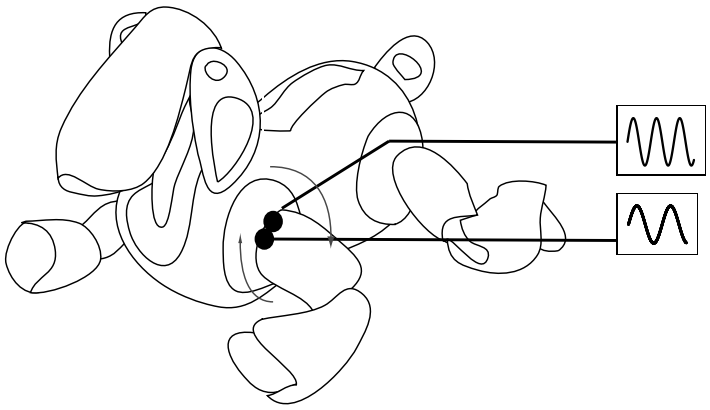

Figure 1.2 A robot can learn to walk just by exploring smartly a sensorimotor space. In the experiment. a progress-driven kernel controls the movement of the different motors of a four-legged robot. For each motor, it chooses the period, the phase and the amplitude of a sinusoidal signal. The prediction system tries to predict the effect of the different set of parameters in the way the image captured by a camera placed on the robot's head is modified. This indirectly reflects the movement of its torso. At each iteration the kernel produces the values for the next parameter set in order to maximize the reduction of the prediction error.

Then the robot will start exploring different ways to move backwards. During this exploration, it is likely that it discovers that certain modification of the parameters could lead to some sort of rotation movement, at least from an external observer's point of view. This is a new set of progress niches that the robot will learn to exploit when the skills for walking backwards will be essentially mastered.

In most experiments, it takes typically three hours for the kernel to find several subsets of parameters resulting in moving forward, backwards, sideways and to turn left and right. At no time in the process the robot was given the objective of learning to walk. Guided by the principle of maximizing the reduction of error in prediction, the robot ends up developing versatile locomotion skills. Actually, this versatility is the result of the unspecific nature of the kernel. A robot artificially motivated to go towards a specific object may not have learnt to walk backwards or to spin[4] .

The fact that walking backwards revealed itself to be a parameter subset easier to discover was not easy to foresee. Given the morphological physical structure of the robot and the kind of ground the robot was placed on during the experiments, the walking backward movement happened to be the first to be discovered. To know whether this progress niche is actually an attractor for most developmental trajectories, it is necessary to set up a bench of experimental trials, changing systematically the initial conditions, including the morphology of the robot itself. With such an experimental approach it becomes possible to study the developmental consequences of a physical modification of the body. A longer leg or a more flexible back can change importantly the structure of the progress niches and therefore the trajectory explored by the kernel. From a methodological point of view, the body becomes an *experimental variable*.

These robotic experiments naturally lead to novel questions addressed at other fields, including neurosciences (Can we identify the neural circuits that act as a kernel ? (Kaplan and Oudeyer, 2007a)), developmental psychology (Can we reinterpret the developmental sequences of young children as progress niches ? (Kaplan and Oudeyer, 2007c)) or in linguistics (Can we reconsider the debate on innateness in the language learning by reconsidering the role of the body in this process ?(Kaplan et al., 2007)).

1.2.3 Fluid body envelopes

A simple way to change the body envelope of a robot is to equipped it with a tool. Figure 1.3 a) shows how the body of four-legged robot can be simply extended by a helmet that plays the role of a prosthetic

[4] There exist many different gait patterns for four-legged robots. In the discussed experiment only a "walking" gait was discovered by the robot. We do not know whether other gaits, like trotting, could be discovered using the same approach.

finger.With this simple extension the robot can now push buttons, press on hard surfaces, even switch on or off other devices. This is a new space to explore.

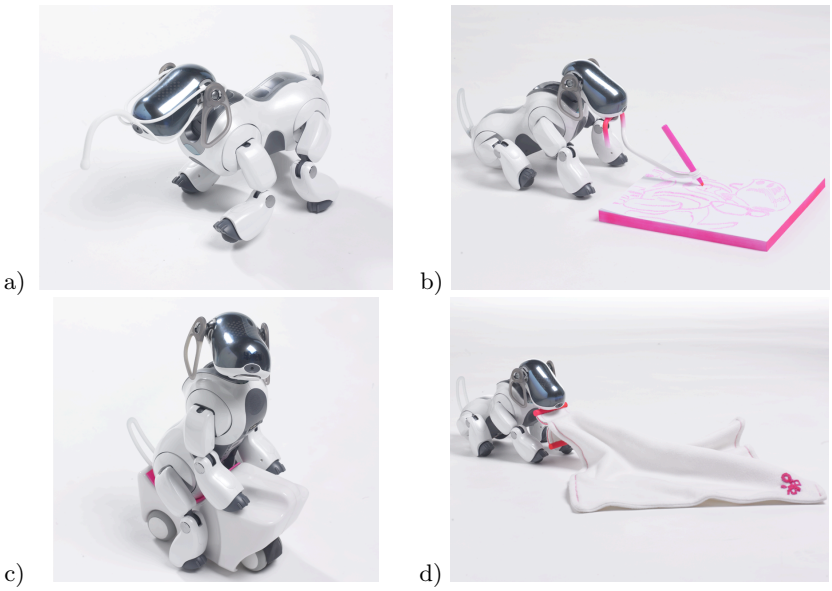

Figure 1.3 a) A helmet-finger extension. Design : ECAL / Stephane Barbier-Bouvet b) A pen holder extension. Design : ECAL / Meynet Bndicte, Burgisser Olivier, Xavier Rui,Wildi Sbastien et Reymond Simeon c) A scooter. Design : ECAL/ Clement Benoit and Moro Nicolas d) A blanket with a special handle. Design : ECAL / Meynet Benedicte, Burgisser Olivier, Xavier Rui,Wildi Sebastien et Reymond Simeon Photo ECAL / Milo Keller

Figure 1.3 b) shows the same idea with a pen holder. With this simple extension, the robot can now leave traces and use the environment as an external memory. A drawing is the temporal integration on a paper of a sequence of gestures. This simple pen holder opens a whole new space of exploration where the machine can learn to predict the relationship between a sequence of actions and particular kinds of representations. Such kind of anticipation is likely to be a fundamental milestone on the road towards higher-level cognition (Steels, 2003).

Figure 1.3 c) presents a small scooter adapted to the morphology of the robot. Learning how to move with this device is not very different form learning how to walk. Like in all the other cases, the progress-driven

kernel discussed in the previous section can be applied (Kaplan et al., 2006). The body changes but the program stays the same.

The progress-driven kernel does only give a partial understanding of the general process of incorporation. We illustrated how it could act on a single space, a single body envelope, like the parameter space resulting in versatile walking skills, but we have not shown how it could be used to shift between them. Incorporation as we described in our introduction involves complex sequences of body envelopes transformation. It involves recursive and hierarchical processes. Typically, once a robot would have learnt how to control its body to walk, it should be able to use these newly discovered walking primitives as basic elements for performing exploration of new spaces. A walking robot will certainly discover new objects, new environment for learning. Let's take for instance the case of the graspable blanket of figure 1.3 d). This blanket is equipped with a special handle adapted to the robot's "mouth". Learning to grasp the blanket is pretty similar than learning to grasp the pen holder we just mentioned. Once the robot would have learnt how to grasp this object, it could explore the specific space corresponding to walking with a blanket. This compositional process could continue endlessly.

Going from the exploration of a single envelope to a generic kernel capable of easily switching between hierarchical envelops is a difficult issue. In particular, it involves a mechanism permitting the formation of habits. The possibility of implementing these different features in a single generic kernel remains to be shown. However, several state of the art methods permit to move towards this goal and envision how such a kernel could work. Multilayer recurrent neural network architecture like the ones considered in (Schmidhuber, 1992; Tani and Nolfi, 1999; Tani, 2007) or the option framework (Sutton et al., 1999) permit hierarchical learning where chunks of behavior can be compiled and continuously adapted to be used later on (see also (Dayan et al., 1993; Ring, 1994; Wiering and Schmidhuber, 1997) for related methods). When a sensorimotor trajectory becomes easily predictable it becomes implicitly or explicitly associated with a dedicated expert predictor, responsible for both recognizing this specific sensorimotor situation and automatically choosing what do do. In other words, when a part of the sensorimotor space becomes predictable it is no longer necessary to explore it at a fine grained level, a higher level control is sufficient. In our walking example, routines for moving forward or backward, turning left or right could likewise become higher-level habits. When this is the case, the progress-

driven kernel could focus on other parts of the space, assuming these basic behavior routines to be in place.

Many challenges remains to be faced to explore the potential of the kernel/envelopes dichototomy. However, for the time being we would like to explore how this distinction could be relevantly used in developmental psychology and neuroscience.

1.3 A concept shift for developmental psychology

1.3.1 Incorporation: a misunderstood process

There is a long tradition of research that discusses the notion of body schema, body map, body image as if it was some stable notion that the child needs to discover or model. Such approach to the body does not give a good account of the flexibility of our embodiment. The relevance of considering the body not as a fixed, determined entity but as a fluid perceptually changing space has been argued by several philosophers (Merleau-Ponty, 1945), psychologists (Schilder, 1935), ethnographers (Warnier, 1999) and neuroscientists (Head and Holmes (1911) or for instance Iriki et al. (1996) for more recent studies). However, we are still far from having a precise model of this process and its relationship with attention, memory and learning.

By many respects, our skin is not the limit of our body. When we interact with tools and technical devices, our body extends its boundaries, changes shape. The stick, the hammer, the pen, the racket, the sword extend our hand and become, after some training, integral parts of our body envelope. Without thinking about it, we bend a bit more when we wear a hat and change the way we walk when we wear special shoes. This is also true for more complex devices. We are the car that we are driving. It took us many painful hours of training to handle it the right way. At the beginning it was an external body element, reacting in unpredictable ways. But once we got used to the dynamics of the machine, the car became like our second skin. We are used to its space occupation, the time necessary to slow down. Driving becomes as natural as walking, an unconscious experience.

Compared to a fixed body, the concept of envelopes that would be extensibles, stretchables, constantly changing, seems more relvant. If we want to fix a nail on wall, we will first pick a hammer. At this stage, the tool is *abstracted* from the environment. A few second later, when we pick

the hammer, we temporally extend our body envelope to include the tool in our hand. It disappears from our attention focus as a direct extension of our hand. It is *incorporated*. Once our goal has been reached, we put back the hammer and the tool becomes again an external object, ready to be used, but separated. This is the fundamental and misunderstood process of *incorporation*.

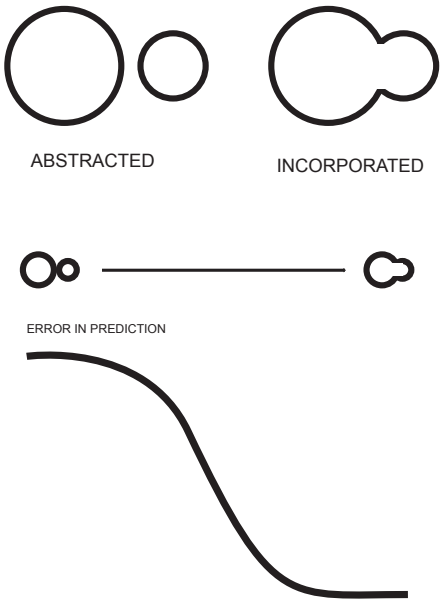

Figure 1.4 Illustration of the incorporation process. Objects can either be abstracted from the environment or incorporated as extension of our body. The process of incorporation takes time. Surprise or failure causes an incorporated objet to be abstracted again. When one learns to use an object, error in prediction corresponds to disincorporation of the object. The fewer the errors the more the object is incorporated.

The first time we use a hammer, we fail to fail to control it perfectly. Every time we fail to predict where the hammer will be, the tool becomes again abstracted, back in our attention focus. It takes time until we can successfully predict the consequences of our action with this "extended"

hand and it is only when prediction errors are very low that the object is fully incorporated (figure 1.4).

Before picking a hammer, we must first choose it among the other tools abstracted from our toolbox. Once picked, new objects, nails, become relevant for the pursuit of our goal. We don't think anymore of our extended hand, we focus on these new abstracted objects. In general, incorporation is a recursive process. At a given state of incorporation, certain objects are abstracted from the environment and become affordants. When one of these objects starts to be controlled and therefore incorporated, our attentional space changes and new objects get abstracted (figure 1.5).

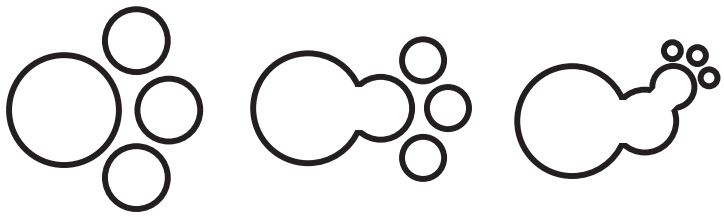

Figure 1.5 Incorporation is a recursive process. At a given state of incorporation, certain objects are abstracted from the environment and become affordants. When one of these objects start to be controlled and therefore incorporated, new objects get abstracted

These processes can be understood in a rather simple way if we consider the kernel/envelopes distinction. In the rest of this section we will discuss the opportunity of assuming the existence of a kernel, playing the role of an intrinsic motivation system, capable of driving the developmental and learning dynamics occuring for particular body envelopes. First, we will see that altough the term "kernel" was not used in this context, such a construct has been discussed in various forms in psychology literature. Then we will illustrate how some children developmental milestones can be interpreted in this framework.

1.3.2 A kernel for active exploration : history of a construct

In psychology, an activity is characterized as intrinsically motivated when there is no apparent reward except the activity itself (Ryan and Deci, 2000). People seek and engage in such activities for their own sake and not because they lead to extrinsic reward. In such cases, the person seems to derive enjoyment directly from the practice of the activity. Following this definition, most children playful or explorative activities can be characterized as being intrinsically motivated. Also, many kinds of adult behavior seem to belong to this category: free problem-solving (solving puzzles, crosswords), creative activities (painting, singing, writing during leisure time), gardening, hiking, etc. Such situations are characterized by a feeling of effortless control, concentration, enjoyment and a contraction of the sense of time (Csikszenthmihalyi, 1991).

A first bloom of investigations concerning intrinsic motivation happened in the 1950s. Researchers started by trying to give an account of exploratory activities on the basis of the theory of drives (Hull, 1943), which are non-nervous-system tissue deficits like hunger or pain and that the organisms try to reduce. For example, (Montgomery, 1954) proposed a drive for exploration and (Harlow, 1950) a drive to manipulate. This drive naming approach had many short-comings which were criticized in detail by White in 1959 (White, 1959): intrinsically motivated exploratory activities have a fundamentally different dynamics. Indeed, they are not homeostatic: the general tendency to explore is never satiated and is not a consummatory response to a stressful perturbation of the organism's body. Moreover, exploration does not seem to be related to any non-nervous-system tissue deficit.

Some researchers then proposed another conceptualization. Festinger's theory of cognitive dissonance (Festinger, 1957) asserted that organisms are motivated to reduce dissonance, that is the 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). However, these theories were criticized on the basis that much human behavior is also intended to *increase* uncertainty, and not only to reduce it. Human seem to look for some forms of optimality between completely uncertain and completely certain situations.

In 1965, Hunt developed the idea that children and adult 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). Whereas most of these researchers focused on the notion of optimal incongruity at the level of psychological processes, a parallel trend investigated the notion of optimal arousal at the physiological level (Hebb, 1955). As over-stimulation and under-stimulation situations induce fear (e.g. dark rooms, noisy rooms), people seem to be motivated to maintain an optimal level of arousal. A complete understanding of intrinsic motivation should certainly include both psychological and physiological levels.

Eventually, 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).

In the recent years, the concept of intrinsic motivation has been less present in mainstream psychology but flourished in social psychology and the study of practices in applied settings, in particular in professional and educational contexts. Based on studies suggesting that extrinsic rewards (money, high grades, prizes) actually destroy intrinsic motivation (an idea actually articulated by Bruner in the 1960s (Bruner, 1962)), some employers and teachers have started to design effective incentive systems based on intrinsic motivation. However, this view is currently at the heart of many controversies (Cameron and Pierce, 2002).

In summary, most psychological approaches of intrinsic motivation postulate that "stimuli worth investigating" are characterized by a particular relationship (incompatibility, discrepancy, uncertainly or on the contrary predictability) between an internal predictive model and the actual structure of the stimulus. This invites us to consider intrinsically motivating activities not only at the descriptive behavioral level (no apparent reward except the activity itself) but primarily in respect to particular internal models built by an agent during its own personal history

of interaction and to postulate the existence of an intrinsic motivation system, namely a kernel.

1.3.3 Reinterpretation of developmental patterns

How can we reinterpret particular developmental processes as being the result of a kernel playing the role of an intrinsic motivation system driving the infant into situations expected to result in maximal learning progress? Taking ground on preliminary experimental results, we discussed in (Kaplan and Oudeyer, 2007b) a scenario presenting the putative role of the progress drive for the development of early imitation. We argue in particular that the kernel/envelope distinction could help understanding why children focus on specific imitative activities at a certain age and how they progressively organize preferential interactions with particular entities present in their environment.

The kernel pushes the agent to discover and focus on situations which lead to maximal learning progress. As we already mentioned, we call these situations, neither too predictable nor too difficult to predict, "progress niches". Once discovered, progress niches progressively disappear as they become more predictable. The notion of progress niches is related to Vygotsky's *zone of proximal development*, where the adult deliberately challenges the child's level of understanding. Adults push children to engage in activities beyond their current mastery level, but not too far beyond so that they remain comprehensible (Vygotsky, 1978). We could interpret the zone of proximal development as a set of potential progress niches organized by the adult in order to help the child learn. But it should be clear that independently of the adults' efforts, what is and what is not a progress niche is ultimately defined from the child's point view. Progress niches share also similarities with Csikszentmihalyi's *flow experiences* (Csikszenthmihalyi, 1991). Csikszentmihalyi argues that some activities are *autotelic* when challenges are appropriately balanced with the skills required to cope with them (see also (Steels, 2004)). We prefer to use the term progress niche by analogy with ecological niches as we refer to a transient state in the evolution of a complex "ecological" system involving the embodied agent and its environment.

The experiments we described with robots illustrated how an agent can (1) separate its sensorimotor space into zones of different predictability levels and (2) choose to focus on the one which leads to maximal

learning progress, called a "progress niche". With this kind of operant models, it could be speculated that meaningful sensorimotor distinctions (self, others and objects in the environment) may be the result of discriminations constructed during a progress-driven process, where different envelopes are constructed and actively explored.

More specifically we can offer an interpretation of several fundamental stages characterizing infant's development during their first year.

Simple forms of imitative behaviour have been argued to be present just after birth. They could constitute a process of early identification. Some totally or partially nativist explanations could account for this early "like-me stance" (Meltzoff and Gopnick, 1993; Moore and Corkum, 1994). This would suggest the possibility of an early distinction between persons and things. If an intermodal mapping facilitating the match between what is seen and what is felt exists, the hypothesis of a kernel for active exploration would suggest that infants will indeed create a discrimination between such easily predictable couplings (interaction with peers) and unpredictable situations (all the other cases) and that they will focus on the first zone of their sensorimotor space that constitutes a "progress niche". Neonates imitation (when it occurs) would be the result of the exploitation of the most predictable coupling present just after birth[5].

During the first two months of their life, infants perform repeated body motion. They kick their legs repeatedly, they wave their arms. This process is sometimes referred as "body babbling". However, nothing indicates that this exploratory behaviour is randomly organised. Rochat argues that children are in fact performing self-imitation, trying to imitate themselves (Rochat, 2002). This would mean that children are structuring their own behaviour in order to make it more predictable and form this way "circular reactions" (Baldwin, 1925; Piaget, 1952). Such self-imitative behaviours can be well explained by the progress drive hypothesis. Sensorimotor trajectories directed towards the child's own body can be easily discriminated from trajectories directed towards other people by comparing their relative predictability difficulty. By many respects, making progress in understanding primary circular reactions is easier than in the cases involving other agents: Self-centered types of be-

[5] This particular interpretation of neotanal imitation shows an example on how can the kernel/envelope approach may lead to rethinking the innate/learnt distinction. In such a case, the "innateness" of the behavior is not "coded" in the genes, but is a direct result of the coupling between a socially structured environment and an innate bias towards behaviors leading to learning progress.

haviour are "progress niches". In such a scenario the "self" emerges as a meaningful discrimination for achieving better predictability. Once this distinction is made, progress for predicting the effects of self-centered actions can be rapidly made.

After two months, infants become more attentive to the external world and particularly to people. Parental scaffolding plays a critical role for making the interaction with the child more predictable (Schaffer, 1977). Parents adapt their own responses so that interactions with the child follow the normal social rules that characterize communicative exchanges (e.g. turn taking). Moreover, if an adult imitates an infant's own actions, it can trigger continued activity in the infant. This early imitative behaviour is referred as "pseudo-imitation" by Piaget (Piaget, 1962). Pseudo-imitation and focus on scaffolded adult behaviour could be seen as predictable effects of the progress drive. As the self-centered trajectories start to be well mastered (and do not constitute "progress niches" anymore), the child's focus shifts to another branch of the discrimination tree, the "self-other" zone.

After five months, attention shifts again from people to objects. Children gain increased control over the manipulation of some objects on which they discover "affordances" (Gibson, 1986). Parents recognize this shift and initiate interactions about those affordant objects. However, children do not alternate easily their attention between the object and their caregiver. A progress-driven process can account for this discrimination between affordant objects and unmastered aspects of the environment. Although this stage is typically not seen as imitative, it could be argued that the exploratory process involved in the discovery of the object affordances shares several common features with the one involved for self-centered activities: the child structures its world looking for "progress niches".

The concepts of kernel and envelope lead to robotic experiments that can be used as a "tool for thoughts" in developmental psychology. In that sense, it may help formulating new concepts useful for the interpretation of the developmental dynamics underlying children's development. For example, the existence of a kernel could explain why certain types of imitative behaviour are produced by children at a certain age and stop to be produced later on. It could also explain how discrimination between actions oriented towards the self, towards others and towards the environment may occur. However, we do not arge that a drive for maximizing learning progress could be the only motivational principle driving children's development. The complete picture is likely to include a complex

set of drives [6]. Developmental dynamics are certainly the result of the interplay between intrinsic and extrinsic forms of motivations, particular learning biases, as well as embodiment and environmental constraints. We believe that computational and robotic approaches can help specifying the contribution of these different components in the overall observed patterns and shed new light on the particular role played by intrinsic motivation in these complex processes.

1.4 A concept shift for neuroscience

1.4.1 Can we identify neural circuits coresponding to a kernel in the human brain ?

Can we identifiy neural circuits that could play the role of a kernel ? Or is it just a conceptual tool to understand how the brain learns ? In neuroscience, dominant views in behavioral neuropsychology have impeded for a long time discussions about putative intrinsic causes to behavior. Learning dynamics in brain systems are still commonly studied in the context of external reward seeking (food, sex, etc) and very rarely as resulting from endogenous and spontaneous processes. Actually, the term "reward" has been misleading as it is used in a different manner in neuropsychology and in machine learning (White, 1989; Wise, 1989; Oudeyer and Kaplan, 2007). In behavioral neuropsychology, rewards are primarily thought as objects or events that increased the probability and intensity of behavioral actions leading to such objects: "rewards make you come back for more" (Thorndike, 1911). This means the function of rewards is based primarily on behavioral effects interpreted in a specific theoretical paradigm. As Schultz puts it "the exploration of neural reward mechanisms should not be based primarily on the physics and the chemistry of reward objects but on specific behavioral theories that define reward function" ((Schultz, 2006) p. 91)

In computational reinforcement learning, a reward is only a numerical quantity used to drive an action-selection algorithm so that the expected cumulated value of this quantity is maximal in the future. In such context, rewards can be thought primarily as internal measures rather than external objects (as clearly argued by Sutton and Barto (Sutton and

[6] The model we discuss in this chapter and present in more details in the Appendix can be easily extended to account for this situation, just by transforming the intrinsic reward function into a linear combination of several reward sources

Barto, 1998)). This may explain why it is much easier from a machine learning perspective to considerer the intrinsic motivation construct as a natural extension of the reinforcement learning paradigm, whereas dominant behavioral theories and experimental methodology in neuroscience does not permit to consider such construct. This is certainly one reason why complex behaviors that do not involve any consummatory reward are rarely discussed.

In the absence of experimental studies concerning intrinsically motivated behaviors, we can consider what ressembles the most: exploratory behaviors. The extended lateral hypothalamic corridor, running from the ventral tegmental area to the nucleus accumbens, has been recognized as a critical piece of a system responsible for exploration. Pankseep calls it the SEEKING system (Panksepp, 1998) (different terms are also used as for instance behavioral activation system (Gray, 1990) or behavioral facilitation system (Depue and Iacono, 1989)). "This harmoniously operating neuroemotional system drives and energizes many mental complexities that humans experience as persistent feelings of interest, curiosity, sensation seeking and, in the presence of a sufficiently complex cortex, the search for higher meaning." ((Panksepp, 1998) p.145). This system, a tiny part compared to the total brain mass, is where one of the major dopamine pathway initiates (for a discussion of anatomical issue one can refer for instance to (Stellar, 1985; Rolls, 1999)).

The roles and functions of dopamine are known to be multiple and complex. Dopamine is thought to influence behavior and learning through two, somewhat decoupled, forms of signal: phasic (bursting and pausing) responses and tonic levels (Grace, 1991). A set of experimental evidence shows that dopamine activity can result from a large number of arousing events including novel stimuli and unexpected rewards (Hooks and Kalivas, 1994; Schultz, 1998; Fiorillo, 2004). On the other hand, dopamine activity is suppressed by events that are associated with reduced arousal or decreased anticipatory excitement, including the actual consumption of food reward and the omission of expected reward (Schultz, 1998). More generally, dopamine circuits appear to have a major effect on our feeling of engagement, excitement, creativity, our willingness to explore the world and to make sense of contingencies (Panksepp, 1998). More precisely, growing evidence currently supports the view of dopamine as a crucial element of incentive salience ("wanting processes") different from hedonic activation processes ("liking processes") (Berridge, 2007). Injections of GABA in the ventral tegmental area and of a dopamine receptor agonistic in the nucleus accumbens cause rats to stop searching for a su-

crose solution, but still drink the liquid when moved close to the bottle (Ikemoto and Panksepp, 1999). Parkinsonian patients who suffer from degeneration of dopaminergic neurons experience not only psychomotor problems (inability to start voluntary movement) but more generally an absence of appetite to engage in exploratory behavior and a lack of interest for pursuing cognitive tasks (Bernheimer et al., 1973). When the dopamine system is artificially activated via electrical or chemical means, humans and animals engage in eager exploration of their environment and display signs of interest and curiosity (Panksepp, 1998). Likewise, the addictive effects of cocaine, amphetamine, opioids, ethanol, nicotine and canabinoid are directly related to the way they activate dopamine systems (Carboni et al., 1989; Pettit and Justice Jr., 1989; Yoshimoto et al., 1991). Finally, too much dopamine activity are thought to be at the origins of uncontrolled speech and movement (Tourette's syndrome), obsessive-compulsive disorder, euphoria, overexcitement, mania and psychosis in the context of schizophrenic behavior (Bell, 1973; Weinberger, 1987; Grace, 1991; Weiner and Joel, 2002).

Things get even more complex and controversial when one tries to link these observation with precise computational models. Hypotheses concerning phasic dopamine's potential role in learning have flourished in the last ten years. Schultz and colleagues have conducted a series of recording of midbrain dopamine neurons firing patterns in awake monkeys under various behavioral conditions which suggested that dopamine neurons fire in response to unpredicted reward (see (Schultz, 1998) for a review). Based on these observations, they develop the hypothesis that phasic dopamine responses drive learning by signalling an error that labels some events as "better than expected". This type of signalling has been interpreted in the framework of computational reinforcement learning as analogous to the prediction error signal of the temporal difference (TD) learning algorithm (Sutton, 1988). In this scheme, a phasic dopamine signal interpreted as TD-error plays a double role (Houk et al., 1995; Barto, 1995; Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 2001; Doya, 2002; Baldassarre, 2002; Khamassi et al., 2005). First, this error is used as a classical training signal to improve future prediction. Second, it is used for finding the actions that maximize reward. This so-called actor-critic reinforcement learning architecture have been presented as a relevant model to account for both functional and anatomical subdivisions in the midbrain dopamine system. However, most of the simple mappings that were first suggested, in particular the association of the actor to matrisome and the critic to the striosome part

of the striatum are now seriously argued to be inconsistent with known anatomy of these nuclei (Joel et al., 2002).

Computational models of phasic dopamine activity based on the error signal hypothesis have also raised controvery for other reasons. One of them, central to our discussion, is that several stimuli that are *not* associated with reward prediction are known to activate the dopamine system in various manner. This is in particular the case for novel, unexpected 'never-rewarded' stimuli (Hooks and Kalivas, 1994; Ikemoto and Panksepp, 1999; Horvitz, 2000, 2002; Fiorillo, 2004). The classic TD-error model does account for novelty responses. As a consequence, Kakade and Dayan suggested to extend the framework including for instance "novelty bonuses" (Kakade and Dayan, 2002) that distort the structure of the reward to include novelty effects (in a similar manner that "exploration bonuses" permit to ensure continued exploration in theoretical machine learning models (Dayan and Sejnowski, 1996)). More recently, Smith and colleagues presented another TD-error model model in which phasic dopamine activation is modeled by the combination of "Surprise" and "Significance" measures (Smith et al., 2006). These attempts to reintegrate novelty and surprise components into a model elaborated in a framework based on extrinsic reward seeking may successfully account for a larger number of experimental observations. However, this is done in the expense of a complexification of a model that was not meant to deal with such type of behavior.

Some authors developed an alternative hypothesis to the reward prediction error interpretation, namely that dopamine promotes behavioural switching (Oades, 1985; Redgrave et al., 1999). In this interpretation, dopaminergic-neuron firing would be an essential component for directing attentional processes to unexpected, behaviorally important stimuli (related or unrelated to rewards). This hypothesis is supported by substantial evidence but stays at a very general explanation level. Actually, Kakade and Dayan argued that this interpretation is not incompatible with reward error-signaling hypothesis provided that the model is modified to account for novelty effect (Kakade and Dayan, 2002).

The incentive salience hypotheses, despite their psychological foundations, are not yet supported by many computational models. But they are some progress in this direction. In 2003, McClure and colleagues argued that incentive salience interpretation is not incompatible with the error signal hypothesis and presented a model where incentive salience is assimilated to expected future reward (McClure et al., 2003). Another recent interesting investigation can be found in (Niv et al., 2006) con-

cerning an interpretation of tonic responses. In this model, tonic levels of dopamine is modeled as encoding "average rate of reward" and used to drive response vigor (slower or faster responding) into a reinforcement learning framework. With this dual model, the authors claim that their theory "dovetails neatly with both computational theories which suggest that the phasic activity of dopamine neurons reports appetitive prediction errors and psychological theories about dopamine's role in energizing responses" (Niv et al., 2006).

In summary, despite many controversies, converging evidence seems to suggest that (1) dopamine plays a crucial role in exploratory and investigation behavior, (2) the meso-accumbens dopamine system is an important brain component to rapidly orient attentional resources to novel events. Moreover, current hypotheses may favor a dual interpretation of dopamine's functions where phasic dopamine is linked with prediction error and tonic dopamine involved in processes of energizing responses.

1.4.2 Tonic dopamine as a signal of expected prediction error decrease

We just reviewed several elements of the current complex debate on the role and function of dopamine in action selection and learning. Based on the investigation we conducted using the kernel/envelop dichotomy, we would like to introduce yet another interpretation of the potential role of dopamine by formulating the hypothesis that tonic dopamine acts as a signal of "progress niches", i.e. states where prediction error of some internal model is expected to decrease. As experimental researches in neuroscience have not really studied intrinsically motivated activities per se, it is not clear at this stage to decide whether this hypothesis is compatible or incompatible with the other interpretation of dopamine we have reviewed. Nevertheless, we can discuss how this interpretation fits with existing hypotheses and observations of the dopamine's functions.

We have just discussed the interpretation of tonic dopamine as a 'wanting' motivational signal (incentive salience hypothesis). In the context of intrinsically motivated behavior, we believe this view is compatible with the hypothesis of dopamine as signal of "progress niches". Dopamine acts as an invitation to investigate these "promising" states. This interpretation is also coherent with investigations that were conducted concerning human affective experience during stimulation of the dopamine circuits. When the lateral hypothalamus dopamine system is

stimulated (part of the SEEKING system previously discussed), people report a feeling that "something very interesting and exciting is going on" ((Panksepp, 1998), p.149 based on experiments reported in (Heath, 1963; Quaade et al., 1974)). This corresponds to subjective affective states linked with intrinsically motivating activities (Csikszenthmihalyi, 1991).

In addition, Berridge articulates the proposition that "dopamine neurons code an informational consequence of learning signals, reflecting learning and prediction that is generated elsewhere in the brain but do not cause any new learning themselves" ((Berridge, 2007), p.405). In this view, dopamine signals are a consequence and not a cause of learning phenomena happening elsewhere in the brain. This is consistent with the fact that dopamine neurons originating in the midbrain are recognized to have only sparse direct access to the signals information that needs to be integrated by an associative learning mechanism. All the signals that they receive are likely to be "highly processed already by forebrain structures before dopamine cells get much learning-relevant information" ((Berridge, 2007), p.406, see also (Dommett et al., 2005)).

In the model of a kernel presented in the appendix of this chapter, this progress signal is used as a reinforcement to drive action-selection and behavioral switching. This aspect of our architecture could lead to a similar interpretation of the role of dopamine in several previous (and now often criticized) actor-critic models of action-selection occurring in the basal ganglia (Houk et al., 1995; Barto, 1995; Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 2001; Doya, 2002; Baldassarre, 2002; Khamassi et al., 2005). Let's recall that the dorsal striatum receives glutamate inputs from almost all regions of the cerebral cortex. Striatal neurons fire in relation to movement of a particular body part but also to preparation of movement, desired outcome of a movement, to visual and auditory stimuli and to visual saccades toward a particular direction. In most actor-critic computational models of the basal ganglia, dopamine responses originating the substantia nigra is interpreted as increasing the synaptic strength, between currently active striatal input and output elements (thus shaping the policy of the actor in an actor-critic interpretation). With this mechanism, if the striatal outputs corresponds to motor responses and that dopamine cells become active in the presence of an unexpected reward, the same pattern of inputs should elicit the same pattern of motor outputs in the future. One of the criticism to this interpretation is that "if dopamine neurons respond to surprise/arousing events, regardless of appetitive or aversive values,

one would postulate that dopamine activation does not serve to increase the likelihood that a given behavioral response is repeated under similar input conditions" ((Horvitz, 2002) p. 70). Progress niches can be extrinsically rewarding (i.e progress in playing poker sometimes result in gaining some money) or aversive (i.e. risk-taking behavior in extreme sports). Therefore, we believe our hypothesis is compatible with interpretations of the basal-ganglia based action-selection circuits that control the choice of actions during cortico-striato-thalamo-cortical loops.

However, the precise architecture of this reinforcement learning architecture is at this stage very open. A seducing hypothesis would be that the much studied reinforcement learning architectures based on short prediction error phasic signals could be just reused with an internal self-generated reward, namely expected progress. This should lead to a complementary interpretation of the role of phasic and tonic dopamine in intrinsically motivated behavior in reinforcement. An alternative hypothesis is that tonic dopamine is directly used as a reinforcement signal. As previously discussed, Niv and colleagues assimilated the role of tonic dopamine to an average reward signal in a recent computational model (Niv et al., 2006), a view which seems to contradict the hypothesis articulated a few years ago that tonic dopamine signal reports a long-run average punishment rate (Daw et al., 2002). Our hypothesis is based on the difference of two long-run average prediction error rate (equation 1.3 of the model presented in the Appendix). We will now discuss how and where this progress signal could be measured.

1.4.3 Cortical microcircuits as both prediction and metaprediction systems

Following our hypothesis that tonic dopamine acts as signal of prediction progress, we must now guess where learning progress could be computed. For this part, our hypothesis will be that cortical microcircuits acts as both prediction and metaprediction systems and therefore have the possibilty of directly computing regional learning progress, through an unsupervised regional assignment as this is done in the computational model we have presented.

However, before considering this hypothesis let us briefly explore some alternative ones. The simpler one would be that progress is evaluated in some way or another in the limbic system itself. If indeed, as many authors suggests, phasic responses of dopamine neurons report prediction error in certain contexts, their integration over time could be easily

performed just through the slow accumulation of dopamine in certain part of neural circuitry (hypothesis discussed in (Niv et al., 2006)). By comparing two running average of the phasic signals one could get an approximation of equation 1.1 of the model presented in the appendix. However, to be appropriately measured, progress must be evaluated in regional manner, by local "expert" circuits. Although it is not impossible to imagine an architecture that would maintain such type of regional specialized circuitry in the basal ganglia (see for instance the multiple expert actor-critic architectures described (Khamassi et al., 2005)), we believe this is not the most likely hypothesis.

As we argued, scalability considerations in real-world structured inhomogeneous spaces favor architectures in which neural resources can be easily recruited or built for different kinds of initially unknown activities. This still leaves many possibilities. Kawato argues that, from a computational point of view, "it is conceivable that internal models are located in all brain regions having synaptic plasticity, provided that they receive and send out relevant information for their input and output" (Kawato, 1999). Doya suggested broad computational distinction between the cortex, the basal ganglia and the cerebellum, each of those associated with a particular type of learning problems, unsupervised learning, reinforcement learning and supervised learning, respectively (Doya, 1999). Another potential candidate location, the hippocampus has often been described as a comparator of predicted and actual events (Gray, 1982) and fMRI studies revealed that its activity was correlated with the occurrence of unexpected events (Ploghaus et al., 2000). Among all these possibilities, we believe the most promising direction of exploration is the cortical one, essentially because the cortex offers the type of open-ended unsupervised "expert circuits" recruitment that we believe are crucial for the computation of learning progress.

A single neural microcircuit forms an immensely complicated network with multiple recurrent loops and highly heterogeneous components (Mountcastle, 1978; Shepherd, 1988; Douglas and Martin, 1998). Finding what type of computation could be performed with such a high dimensional dynamical system is a major challenge for computational neuroscience. To explore our hypothesis, we must investigate whether the computational power and evolutionary advantage of columns can be unveiled if these complex networks are considered not only as predictors but performing both prediction and metaprediction functions (by not only anticipating future sensorimotor events but also its own errors in prediction and learning progress).

In recent years, several computational models explored how cortical circuits could be used as prediction devices. Maas and Markram suggested to view a column as a liquid state machine (LSM) (Maas et al., 2002) (which is somewhat similar to Echo State Networks described by Jaeger (Jaeger, 2001; Jaeger and Haas, 2004)). Like the Turing machine, the model of a LSM is based on a rigorous mathematical framework that guarantees, under idealized conditions, universal computing power for time series problems. More recently, Deneve, Duhamel and Pouget presented a model of a Kalman filter based on recurrent basis function networks, a kind of model that can be easily mapped onto cortical circuits (Deneve et al., 2007). Kalman filters share some similarity with the kind of metaprediction machinary we have discussed in this article, as they also deal with modeling errors made by prediction of internal models. However, we must admit that there is not currently any definitive experimental evidence or computational model that support precisely the idea that cortical circuit actually compute their own learning progress.

If indeed we could show that cortical microcircuits can signal this information to other parts of the brain, the mapping with our model based on a stable kernel for the active exploration of many different envelopes would be rather straighforward. Lateral inhibition mechanisms, specialization dynamics and other self-organizing processes that are typical of cortical plasticity should permit without problems to perform the type of regionalization of the sensorimotor space that an architecture like the one presented in the appendix features. Moreover, hierarchical organization that has been identified in the neocortical dynamics would naturally extend one of the main weakness of the present computational architecture: its difficulty to deal with hierarchical forms of learning. As previously argued, action-selection could then be realized by some form of subcortical actor-critic architecture, similar to the one involved in the optimization of extrinsic forms of rewards.

We believe this hypothesis is consistent from an evolutionary perspective, or at least that an "evolutionary story" can be articulated around it. The relatively "recent" invention of the cortical column circuits correlates roughly with the fact that only mammals seems to display intrinsically motivated behavior. Once discovered by evolution, cortical columns have multiplied themselves leading to the highly expanded human cortex (largest number of cortical neurons (10^{10}) among all animals, closely followed by large cetaceans and elephants (Roth and Dicke, 2005), over thousandfold expansion from mouse to man to provide 80% of the human brain). What can make them so advantageous from an evolutionary

point of view? It is reasonable to suppose that the kernel responsible of intrinsically motivated exploration appeared after (or on top of) an existing machinery dedicated to the optimization of extrinsic motivation. For an extrinsically-motivated animal, value is linked with specific stimuli, particular visual patterns, movement, loud sounds, or any bodily sensations that signal that basic homeostatic physiological needs like food or physical integrity are (not) being fulfilled. These animals can develop behavioral strategies to experience the corresponding situations as often as possible. However, when an efficient strategy is found, nothing pushes them further towards new activities. Their development stops there.

The apparition of a basic cortical circuit that could not only acts as predictor but also as metapredictor capable of evaluating its own learning progress can be seen as a major evolutionary transition. The brain manages now to produce its own reward, a progress signal, internal to the central nervous system with no significant biological effects on non-nervous-system tissues. This is the basis of an adaptive internal value system for which sensorimotor experiences that produce positive value evolve with time. This is what drives the acquisition of novel skills, with increasing structure and complexity. This is a revolution, yet it is essentially based on the old brain circuitry that evolved for the optimisation of specific extrinsic needs. If we follow our hypothesis, the unique human cortical expansion has to be understood as a coevolutionary dynamical process linking larger "space" for learning and more things to learn. In some way, it is human culture, as a huge reservoir of progress niches, that has put pressure in having more of these basic processing units.

The attentive reader should have noticed that there is something peculiar about the hypotheses we present here. We hypothesize that the cortical circuits offer the neural substrate for representing a very large number of sensorimotor spaces corresponding to our concept of body envelope. Additionaly we suppose that they can perform the local computation necessary to the evalutation of learning progress that is then relayed by subcortical structures. This means that in this view it would be wrong identify the subcortical structure to the stable kernel and the cortical ones to the fluid envelopes, as some of the crucial computational operations of the kernel are supposed to be performed locally by the cortical areas.

1.5 Concluding remarks

Recent research in robotics sets the stage, both theoretically and experimentally, for a new conception of the embodiment process that views the experience of the body as a fluid, constantly changing space. By extracting, on the one hand, the concept of generic and stable kernel, origin of the movement and action, and, on the other hand, the notion of changing body envelopes, robotics offers a novel framework for considering deep and complex issues linked with development and innateness. Indeed, what is development if not a succession of embodiment: not only a body that changes physically but the discovery of novel embodied spaces. Each new skill acquired changes the space to explore. Through incorporation, the body extends temporally including objects, tools, musical interfaces or vehicles as novel envelopes to explore with no fundamental differences with their biological counterpart (Warnier, 1999; Clark, 2004).

By pushing further this notion of fluid body envelopes, couldn't we consider symbolic reasoning and abstract thought as merely special forms of body extension? Lakoff and Nunez suggested very convincingly that there is a direct correspondence between sensorimotor manipulation and very abstract notion in mathematics (Lakoff and Nunez, 2001). Metaphorical transfer, one of most fundamental process to bootstrap higher-level of cognition, can be relevantly considered as a process of incorporation (Lakoff and Johnson, 1998). Eventually, couldn't we consider linguistic communication itself as just one particular case of embodied exploration (Oudeyer and Kaplan, 2006)? All these spaces could be explored relevantly by progress-driven kernel like the one we discussed in this chapter.

Robots have always introdued technological and philosophical questions (Kaplan, 2004, 2005; Asada et al., 2009). They help us think about ourselves by difference. Studying the development of robots with embodied spaces very different from our own, is probably the most promising way to study the role of our body in our own developmental processes. In that sense. robots are not models. They are physical *thought experiments*. That's why they can permit to consider apparently impossible splits, like the ones separating the body from the animation processes or, more recently, the distinction between a stable kernel and fluid body envelopes.

Appendix : a kernel for progress-driven exploration of sensorimotor envelopes

Building a kernel permitting a continuous search for learning progress implies complicated and deep issues. The idealized problem illustrated on figure 1.1 allowed us to make more concrete the intuition that focusing on activities where prediction errors decrease most can generate organized developmental sequences. Nevertheless, the reality is in fact not as simple. Indeed, in this idealized problem, four different sensorimotor situations/activities were predefined. Thus it was assumed that when the idealized machine would produce an action and make a prediction about it, it would be automatically associated with one of the predefined kinds of activities. Learning progress would then be simply computed by for example comparing the difference between the mean of errors in prediction at time t and at time $t-\theta$. On the contrary, infants do not come to the world with an organized predefined set of possible kinds of activities. It would in fact be contradictory, since they are capable of open-ended development, and most of what they will learn is impossible to know in advance. It also occurs for a developmental robot, for which the world is initially a fuzzy blooming flow of unorganized sensorimotor values. In this case, how can we define learning progress? What meaning can we attribute to "maximizing the decrease of prediction errors"?.

A first possibility would be just to compute learning progress at time t as the difference between the mean prediction errors at time t and at time $t - \theta$. But implementing this on a robot quickly shows that it is in fact nonsense. For example, the behavior of a robot motivated to maximize such a progress would be typically an alternation between jumping randomly against walls and periods of complete immobility. Indeed, passing from the first behavior (highly unpredictable) to the second (highly predictable) corresponds to a large decrease in prediction errors, and so to a large internal reward. So we see that there is a need to compute learning progress by comparing prediction errors in sensorimotor contexts that are similar, which leads us to a second possible approach.

In order to describe this second possibility, we need to introduce a few formal notations and precisions about the computational architecture that will embed intrinsic motivation. Let us denote a sensorimotor situation with the state vector $\mathbf{x}(\mathbf{t})$ (e.g. a given action performed in a given context), and its outcome with $\mathbf{y}(\mathbf{t})$ (e.g. the perceptual consequence of this action). Let's call M a prediction system trying to model this function, producing for any $\mathbf{x}(\mathbf{t})$ a prediction $\mathbf{y}'(\mathbf{t})$. Once the actual

evolution $\mathbf{y(t)}$ is known, the error $e_{\mathbf{x}}(t) = |\mathbf{y(t)} - \mathbf{y'(t)}|$ in prediction can be computed and used as a feedback to improve the performances of M. At this stage, no assumption is made regarding the kind of prediction system used in M. It could be for instance a linear predictor, a neural network or any other prediction method currently used in machine learning. Within this framework, it is possible to imagine a first manner to compute a meaningful measure of learning progress. Indeed, one could compute a measure of learning progress $p_{\mathbf{x}}(t)$ for every single sensorimotor situation $\mathbf{x}$ through the monitoring of its associated prediction errors in the past, for example with the formula:

$$p_{\mathbf{x}}(t) = < e_{\mathbf{x}}(t-\theta) > - < e_{\mathbf{x}}(t) > \tag{1.1}$$

where $< e_{\mathbf{x}}(t) >$ is the mean of $e_{\mathbf{x}}$ values in the last τ predictions. Thus, we here compare prediction errors in exactly the same situation $\mathbf{x}$, and so we compare only identical sensorimotor contexts. The problem is that, whereas this is an imaginable solution in small symbolic sensorimotor spaces, this is inapplicable to the real world for two reasons. The first reason is that, because the world is very large, continuous and noisy, it never happens to an organism to experience twice exactly the same sensorimotor state. There are always slight differences. A possible solution to this limit would be to introduce a distance function $d(\mathbf{x_m}, \mathbf{x_n})$ and to define learning progress locally in a point x as the decrease in prediction errors concerning sensorimotor contexts that are close under this distance function:

$$p_{\mathbf{x}}(t) = < e^{\delta}_{\mathbf{x}}(t-\theta) > - < e^{\delta}_{\mathbf{x}}(t) > \tag{1.2}$$

where $< e^{\delta}_{\mathbf{x}}(t) >$ denotes the mean of all $\{e_{\mathbf{x_l}} | d(\mathbf{x}, \mathbf{x_l}) < \delta\}$ values in the last τ predictions, and where δ is a small fixed threshold. Using this measure would typically allow the machine to manage to repeatedly try roughly the same action in roughly the same context and identify all the resulting prediction errors as characterizing the same sensorimotor situation (and thus overcoming the noise). Now, there is a second problem which this solution does not solve. Many learning machineries, and in particular the one used by infants, are fast and characterized by "one-shot learning". In practice, this means that typically, an infant who observes the consequence of a given action in a given context will readily be able to predict very well what happens if exactly the same action happens in the same context again. Learning machines such as

memory-based algorithms also show this feature. As a consequence, if learning progress is defined locally as explained above, a given sensorimotor situation will be typically interesting only for a very brief amount of time, and will hardly direct further exploration. For example, using this approach, a robot playing with a plastic toy might try to squash it on the ground to see the noise it produces, experiencing learning progress in the first few times it tries, but would quickly stop playing with it and typically would not try to squash it for example on the sofa or on a wall to hear the result. This is because its measure of potential learning progress is still too local.

Thus, we conclude that there really is a need to build broad categories of activities (e.g. squashing plastic toys on surfaces or shooting with the foot in small objects) as those pre-given in the initial idealized problem. The computation of learning progress will only become both meaningful and efficient if an automatic mechanism allows for the mental construction of these categories of activities, typically corresponding to not-so-small regions in the sensorimotor space. We have presented a possible solution, based on the iterative splitting of the sensorimotor space into regions $\mathcal{R}_n$. Initially, the sensorimotor space is considered as one big region, and progressively regions split into sub-regions containing more homogeneous kinds of actions and sensorimotor contexts (the mechanisms of splitting are detailed in (Oudeyer et al., 2007)). In each region $\mathcal{R}_n$, the history of prediction errors $\{e\}$ is memorized and used to compute a measure of learning progress that characterizes this region:

$$p_{\mathcal{R}_n}(t) = < e_{\mathcal{R}_n}(t-\theta) > - < e_{\mathcal{R}_n}(t) > \tag{1.3}$$

where $< e_{\mathcal{R}_n}(t) >$ is the mean of $\{e_{\mathbf{x}} | \mathbf{x} \in \mathcal{R}_{\mathbf{n}}\}$ values in the last τ predictions.

Given this iterative region-based operationalization of learning progress, there are two general ways of building a neural architecture that uses it to implement intrinsic motivation. A first kind of architecture, called monolithic, includes two loosely coupled main modules. The first module would be the neural circuitry implementing the prediction machine M presented earlier, and learning to predict the $\mathbf{x} \rightarrow \mathbf{y}$ mapping. The second module would be a neural circuitry *metaM* organizing the space into different regions $\mathcal{R}_n$ and modelling the learning progress of M in each of these regions, based on the inputs $(\mathbf{x}(\mathbf{t}), \mathbf{e_x}(\mathbf{t}))$ provided by M. This architecture makes no assumption at all on the mechanisms and representations used by the learning machine M. In particular, the split-

ting of the space into regions is not informed by the internal structure of M. This makes this version of the architecture general, but makes the scalability problematic in real-world structured inhomogeneous spaces where typically specific neural ressources will be recruited/built for different kinds of activities.

This is why we have developed a second architecture, in which the machines M and $metaM$ are tightly coupled. In this version, each region $\mathcal{R}_n$ is associated with a circuit $M_{\mathcal{R}_n}$, called an expert, as well as with a regional meta machine $metaM_{\mathcal{R}_n}$. A given expert $M_{\mathcal{R}_n}$ is responsible for the prediction of $\mathbf{y}$ given $\mathbf{x}$ when $\mathbf{x}$ is a situation which is covered by $\mathcal{R}_n$. Also, each expert $M_{\mathcal{R}_n}$ is only trained on inputs $(\mathbf{x}, \mathbf{y})$ where $\mathbf{x}$ belongs to its associated region $\mathcal{R}_n$. This leads to a structure in which a single expert circuit is assigned for each non-overlapping partition of the space. The meta-machine $metaM_{\mathcal{R}_n}$ associated to each expert circuit can then compute the local learning progress of this region of the sensorimotor space (See Figure 1.6 (b) for a symbolic illustration of this splitting/assignment process). The idea of using multiple experts has been already explored in several works including for instance (Jordan and Jacobs, 1994; Tani and Nolfi, 1999; Kawato, 1999; Doya et al., 2002; Baldassarre, 2002; Khamassi et al., 2005)

The basic circuits we just described permit to compute an internal reward $r(t) = p_{\mathcal{R}_n}(t)$, each time an action is performed in a given sensorimotor context, depending on how much learning progress has been achieved in a particular region $\mathcal{R}_n$. An intrinsic motivation to progress corresponds to the maximization of the amount of this internal reward. Mathematically, this can be formulated as the maximization of future expected rewards (i.e. maximization of the return), that is

$$E\{r(t+1)\} = E\{\sum_{t \geq t_n} \gamma^{t-t_n} r(t))\}$$

where γ $(0 \leq \gamma \leq 1)$ is the discount factor, which assigns less weight on the reward expected in the far future. We can note that at this stage, it is theoretically easy to combine this intrinsic reward for learning progress with the sum of other extrinsic rewards $r_e(t)$ coming from other sources, for instance in a linear manner with the formula $r(t) = \alpha.p_{\mathcal{R}_n}(t) + (1 - \alpha)r_e(t)$ (the parameter α measuring the relative weight between intrinsic and extrinsic rewards).

This formulation corresponds to a reinforcement learning problem (Sutton and Barto, 1998) and thus the techniques developed in this field can be used to implement an action selection mechanism which

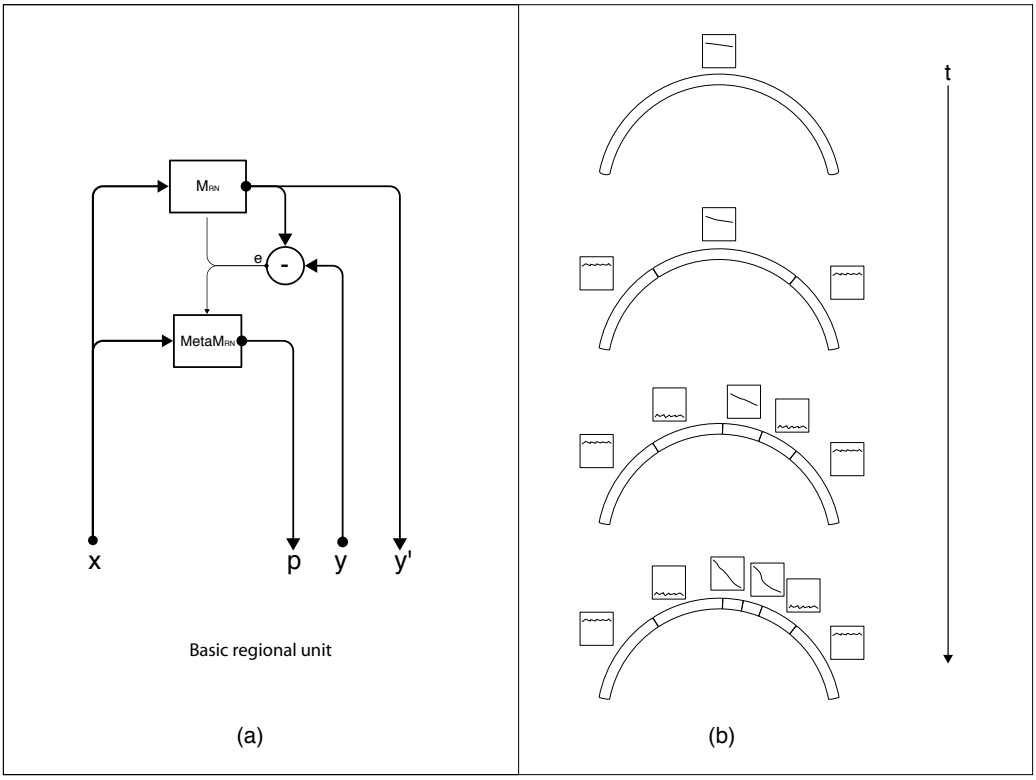

Figure 1.6 (a) An intrinsic motivation system is based on a population of regional units, each comprising an expert predictor $M_{\mathcal{R}_n}$ that learns to anticipate the consequence $\mathbf{y}$ of a given sensorimotor context $\mathbf{x}$ belonging to its associated region of expertise $\mathcal{R}_n$, and a metapredictor $metaM_{\mathcal{R}_n}$ modelling the learning progress of $M_{\mathcal{R}_n}$ in the close past. The learning progress defines the interestingness of situations belonging to a given context, and actions are chosen in order to reach maximally interesting situations. Once the actual consequence is known, $M_{\mathcal{R}_n}$ and $metaM_{\mathcal{R}_n}$ get updated. $metaM_{\mathcal{R}_n}$ re-evaluates the error curve linked with this context and computes an updated measure of the learning progress (local derivative of curve). (b) Illustration of the splitting/assignment process based on self-organized classification system capable of structuring an infinite continuous space of particular situations into higher-level categories (or kinds) of situations. An expert predictor/metapredictor circuit is assigned to each region.

will allow the system to maximize future expected rewards efficiently (e.g. Q-learning (Walkins and Dayan, 1992), TD-learning (Sutton, 1988), etc.). However, predicting prediction error reduction is, by definition, a highly non-stationary problem (progress niches appear and disappear in time). As a consequence, traditional "slow" reinforcement learning techniques are not well adapted in this context. In (Oudeyer et al., 2007), we describe a very simple action-selection circuit that avoids problems related to delayed rewards and makes it possible to use a simple prediction system which can predict $r(t+1)$, and so evaluate $E\{r(t+1)\}$. Let us consider the problem of evaluating $E\{r(t+1)\}$ given a sensory context $\mathbf{S(t)}$ and a candidate action $\mathbf{M(t)}$, constituting a candidate sensorimotor context $\mathbf{SM(t)} = \mathbf{x}(t)$ covered by region $\mathcal{R}_n$. In our architecture, we approximate $E\{r(t+1)\}$ with the learning progress that was achieved in $\mathcal{R}_n$ with the acquisition of its recent exemplars, i.e. $E\{r(t+1)\} \approx p_{\mathcal{R}_n}(t - \theta_{R_n})$ where $t - \theta_{R_n}$ is the time corresponding to the last time region $\mathcal{R}_n$ and the associated expert circuit processed a new exemplar. The action-selection loop goes as follows:

- in a given sensory $\mathbf{S(t)}$ context, the robot makes a list of the possible values of its motor channels $\mathbf{M(t)}$ which it can set; If this list is infinite, which is often the case since we work in continuous sensorimotor spaces, a sample of candidate values is generated;
- each of these candidate motor vectors $\mathbf{M(t)}$ associated with the sensory context $\mathbf{S(t)}$ makes a candidate $\mathbf{SM(t)}$ vector for which the robot finds out the corresponding region $\mathcal{R}_n$; then the formula we just described is used to evaluate the expected learning progress $E\{r(t+1)\}$ that might be the result of executing the candidate action $\mathbf{M(t)}$ in the current context;
- the action for which the system expects the maximal learning progress is chosen with a probability $1 - \epsilon$ and executed, but sometimes a random action is selected (with a probability ϵ), typically 0.35 in the following experiments).
- after the action has been executed and the consequences measured, the system is updated.

More sophisticated action-selection circuits could certainly be envisioned (see for example (Sutton and Barto, 1998)). However, this one revealed to be surprisingly efficient in the real-world experiments we conducted.

References

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Toshio, I., Yoshikawa, Y., Ogino, M., and Yoshida, C. 2009. Cogntive developmental robotics: A survey. *IEEE Transactions on Autonomous Mental Development*, **1**(1), 12–34.

Baldassarre, G. 2002. A modular neural-network model of the basal ganglia's role in learning and selection motor behaviors. *Journal of Cognitive Systems Research*, **3**(1), 5–13.

Baldwin, J.M. 1925. *Mental development in the child and the race*. New York: The Macmillan Company.

Barto, A.G. 1995. Adaptive critics and the basal ganglia. Pages 215–232 of: Houk, J.C., Davis, J.L., and Beiser, D.G. (eds), *Models of information processing in the basal ganglia*. Cambridge, MA, USA: MIT Press.

Barto, A.G., Singh, S., and Chentanez, N. 2004. Intrinsically Motivated Learning of Hierarchical Collections of Skills. In: *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*.

Bell, D.S. 1973. The experimental reproduction of amphetamine psychosis. *Arch Gen Psychiatry*, **29**, 35–40.

Berlyne, D.E. 1960. *Conflict, Arousal and Curiosity*. McGraw-Hill.

Bernheimer, H., Birkmayer, W., Hornykiewicz, Jellinger, K., and Seitelberger, F. 1973. Brain dopamine and the syndromes of Parkinson and Huntington: Clinical, morphological and neurochemical correlations. *Journal of the neurological Sciences*, **20**, 415–455.

Berridge, K. C. 2007. The debate over dopamine's role in reward: the case of incentive salience. *Psychopharmacology*, **191**, 391–431.

Brooks, R. 1999. *Cambrian intelligence: The early history of the new AI*. Cambridge, MA: The MIT Press.

Bruner, J. 1962. *On Knowing: Essays for the left hand*. Cambridge, MA: Harvard University Press.

Cameron, J., and Pierce, W.D. 2002. *Rewards and intrinsic motivation: Resolving the contoversy*. Bergin and Garvey Press.

Carboni, E., Imperato, A., Perezzani, L., and Di Chiara, G. 1989. Amphetamine, cocaine, phencyclidine and nomifensine increases extra-

cellullar dopamine concentrations preferentially in the nucleus accumbens of freely moving rats. *Neuroscience*, **28**, 653–661.

Clark, A. 2004. *Natural-born cyborgs: Minds, Technologies ad the Future of Human Intelligence*. Oxford, UK: Oxford University Press.

Csikszenthmihalyi, M. 1991. *Flow-the psychology of optimal experience*. Harper Perennial.

Daw, N., Kakade, S., and Dayan, P. 2002. Opponent interactions between serotonin and dopamine. *Neural Networks*, **15**, 603–616.

Dayan, P., and Sejnowski, T. J. 1996. Exploration bonuses and dual control. *Machine Learning*, **25**, 5–22.

Dayan, P., Hinton, G., Giles, C., Hanson, S., and Cowan, J. 1993. Feudal Reinforcement Learning. *Advances in Neural Information Processing Systems NIPS*.

De Charms, R. 1968. *Personal causation: the internal affective determinants of behavior*. New York: Academic Press.

Deci, E.L., and Ryan, R.M. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press.

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

Deneve, S., Duhamel, J-R., and Pouget, A. 2007. Optimal sensorimotor integration in recurrent cortical networks: A neural implementation of Kalman filters. *Journal of Neuroscience*, **27**(21), 5744–5756.

Depue, R.A., and Iacono, W.G. 1989. Neurobehavioral aspects of affective disorders. *Annual review of psychology*, **40**, 457–492.

Dommett, E., Coizet, V., Blatha, C.D., Martindale, J., Lefebre, V., Walton, N., Mayhew, J.E., Overton, P.G., and Redgrave, P. 2005. How visual stimuli activate dopaminergic neurons at short latency. *Science*, **307**, 1476–1479.

Douglas, R., and Martin, K. 1998. Neocortex. Pages 459–509 of: Shepherd, G. M. (ed), *The Synaptic Organization of the Brain*. Oxford University Press.

Doya, K. 1999. What are the computations of cerebellum, basal ganglia, and the cerebral cortex. *Neural Networks*, **12**, 961–974.

Doya, K. 2002. Metalearning and neuromodulation. *Neural Networks*, **15**(4–5).

Doya, K., Samejima, K., Katagiri, K., and Kawato, M. 2002. Multiple model-based reinforcement learning. *Neural computation*, **14**, 1347–1369.

Fedorov, V.V. 1972. *Theory of Optimal Experiment*. New York, NY: Academic Press.

Festinger, L. 1957. *A theory of cognitive dissonance*. Row, Peterson: Evanston.

Fiorillo, C. D. 2004. The uncertain nature of dopamine. *Molecular Psychiatry*, 122–123.

Gibson, J. 1986. *The ecological approach to visual perception*. Lawrence Erlbaum Associates.

Grace, A. A. 1991. Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: a hypothesis for the etiology of schizophrenia. *Neuroscience*, **41**(1–24).

Gray, J.A. 1982. *The neuropsychology of anxiety: an enquiry into the functions of the septo-hippocampal system*. Oxford: Clarendon Press.

Gray, J.A. 1990. Brain systems that mediate both emotion and cognition. *Cognition and Emotion*, **4**, 269–288.

Grey Walter, W. 1953. *The Living Brain*. 2nd 1967 edn. Penguin.

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

Head, H., and Holmes, G. 1911. Sensory disturbances from cerebral lesions. *Brain*, **32**(102).

Heath, R. G. 1963. Electrical self-stimulation of the brain in man. *American Journal of Psychiatry*, **120**, 571–577.

Hebb, D. O. 1955. Drives and the c.n.s (conceptual nervous system). *Psychological review*, **62**, 243–254.

Hooks, M.S, and Kalivas, P.W. 1994. Involvement of dopamine and excitatory amino acid transmission in novelty-induced motor activity. *Journal of Pharmacoloy, Experimental Therapeutics*, **269**, 976–988.

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

Horvitz, J-C. 2002. Dopamine gating of glutamatergic sensorimotor and incentive motivational input signals to the striatum. *Behavioral and Brain Research*, **137**, 65–74.

Houk, J.C., Adams, J.L., and Barto, A.G. 1995. A model of how the basal ganglia generate and use neural signals that predict reinforcement. Pages 249–270 of: Houk, J.C., Davis, J.L., and Beiser, D.G. (eds), *Models of information processing in the basal ganglia*. MIT press.

Huang, X., and Weng, J. 2002. Novelty and reinforcement learning in the value system of developmental robots. Pages 47–55 of: Prince, C., Demiris, Y., Marom, Y., Kozima, H., and Balkenius, C. (eds), *Proceedings of the 2nd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems*. Lund University Cognitive Studies 94.

Hull, C. L. 1943. *Principles of behavior: an introduction to behavior theory*. New-York: Appleton-Century-Croft.

Hunt, J. McV. 1965. Intrinsic motivation and its role in psychological development. *Nebraska symposium on motivation*, **13**, 189–282.

Ikemoto, S., and Panksepp, J. 1999. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Research Reviews*, **31**, 6–41.

Iriki, A., Tanaka, A., and Iwamura, Y. 1996. Coding of modified schema during tool use by macaque postcentral neurones. *Cognitive Neuroscience and Neuropsychology*, **7**(14), 2325–2330.

Jaeger, H. 2001. *The Echo State approach to analyzing and training recurrent neural networks*. Tech. rept. GMD Report 148, GMD - German National Research Institute for Computer Science.

Jaeger, H., and Haas, H. 2004. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science*, **304**(5667), 78–80.

Joel, D., Niv, Y., and Ruppin, E. 2002. Actor-critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, **15**, 535–547.

Jordan, M., and Jacobs, R. 1994. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, **6**(2), 181–214.

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

Kakade, S., and Dayan, P. 2002. Dopamine: Generalization and Bonuses. *Neural Networks*, **15**, 549–559.

Kaplan, F. 2004. Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. *International Journal of Humanoid Robotics*, **1**(3), 465–480.

Kaplan, F. 2005. *Les machines apprivoisees : comprendre les robots de loisir*. Coll. Automates Intelligents. Vuibert.

Kaplan, F., and Oudeyer, P-Y. 2003. Motivational principles for visual know-how development. Pages 73–80 of: Prince, C.G., Berthouze, L., Kozima, H., Bullock, D., Stojanov, G., and Balkenius, C. (eds), *Proceedings of the 3rd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic systems*. Lund University Cognitive Studies 101.

Kaplan, F., and Oudeyer, P-Y. 2006. Trends in Epigenetic Robotics: Atlas 2006. In: Kaplan, F., Oudeyer, P-Y., Revel, A., Gaussier, P., Nadel, J., Berthouze, L., Kozima, H., Prince, C., and Balkenius, C. (eds), *Proceedings of the Sixth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. LUCS 128.

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

Kaplan, F., and Oudeyer, P-Y. 2007b. The progress-drive hypothesis: an interpretation of early imitation. Pages 361–377 of: Nehaniv, C., and Dautenhahn, K. (eds), *Models and mechanisms of imitation and social learning: Behavioural, social and communication dimensions*. Cambridge University Press.

Kaplan, F., and Oudeyer, P-Y. 2007c. Un robot motivé pour apprendre : le role des motivations intrinseques dans le developpement sensorimoteur. *Enfance*, **59**(1), 46–58.

Kaplan, F., and Oudeyer, P-Y. 2008. Le corps comme variable experimentale. *La Revue Philosophique*, **2008**(3), 287–298.

Kaplan, F., and Oudeyer, P-Y. 2009. Stable kernels and fluid body envelopes. *SICE Journal of Control, Measurement, and System Integration*, **48**(1).

Kaplan, F., d'Esposito M., and Oudeyer, P-Y. 2006 (October). *AIBO's playroom*. http://aibo.playroom.fr.

Kaplan, F., Oudeyer, P-Y., and Bergen, B. 2007. Computational models in the debate over language learnability. *Infant and Child Development*, **17**(1), 55–80.

Kawato, M. 1999. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, **9**, 718–727.

Khamassi, M., Lachèze, L., Girard, B., Berthoz, A., and Guillot, A. 2005. Actor-Critic Models of Reinforcement Learning in the Basal Ganglia. *Adaptive Behavior*, **13**(2), 131–148.

Lakoff, G., and Nunez, R. 2001. *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York, NY.: Basic Books.

Lakoff, George, and Johnson, Mark. 1998. *Philosophy in the flesh: the embodied mind and its challenge to Western thought.* Basic Books.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. 2003. Developmental Robotics: A Survey. *Connection Science*, **15**(4), 151–190.

Maas, W., Natschlager, T., and Markram, H. 2002. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, **14**(11), 2531–2560.

Marshall, J., Blank, D., and Meeden, L. 2004. An Emergent Framework for Self-Motivation in Developmental Robotics. In: *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004).*

McClure, S., Daw, N.D., and Montague, P.R. 2003. A computational substate for incentive salience. *Trends in Neurosciences*, **26**(8).

Meltzoff, A., and Gopnick, A. 1993. The role of imitation in understanding persons and developing a theory of mind. Pages 335–366 of: S. Baron-Cohen, H. Tager-Flusberg, and D.Cohen (eds), *Understanding other minds.* Oxford, England: Oxford University Press.

Merleau-Ponty, M. 1942. *La structure du comportement.* Presses universitaires de France.

Merleau-Ponty, M. 1945. *Phénoménologie de la Perception.* Gallimard.

Montague, P.R., Dayan, P., and Sejnowski, T.J. 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, **16**, 1936–1947.

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

Moore, C., and Corkum, V. 1994. Social understanding at the end of the first year of life. *Developmental Review*, **14**, 349–372.

Mountcastle, V. 1978. An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System. In: Edelman, G., and Mountcastle, V. (eds), *The Mindful Brain.* MIT press.

Niv, Y., Daw, N.D., Joel, D., and Dayan, P. 2006. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, 507–520.

Oades, R.D. 1985. The role of noradrenaline in tuning and dopamine in switching between signals in the CNS. *Neuroscience Biobehavorial Review*, **9**, 261–282.

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

Oudeyer, P-Y., and Kaplan, F. 2007. What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, **1**(1).

Oudeyer, P-Y., Kaplan, F., Hafner, V. V., and Whyte, A. 2005. The Playground Experiment: Task-Independent Development of a Curious Robot. Pages 42–47 of: Bank, D., and Meeden, L. (eds), *Proceedings of the AAAI Spring Symposium on Developmental Robotics, 2005*.

Oudeyer, P-Y., Kaplan, F., and Hafner, V. 2007. Intrinsic Motivation Systems for Autonomous Mental Development. *IEEE Transactions on Evolutionary Computation*, **11**(1), 265–286.

Panksepp, J. 1998. *Affective neuroscience: the foundations of human and animal emotions*. Oxford University Press.

Pettit, H.O., and Justice Jr., J.B. 1989. Dopamine in the nucleus accumbens during cocaine self-administration as studied by in vivo microdialysis. *Pharmacol. Biochem. Behav.*, **34**, 899–904.

Pfeifer, R., and Bongard, J. 2007. *How the Body shapes the Way we think : How the body shapes the way we think: A new view of intelligence*. Cambridge, MA: MIT Press.

Pfeifer, Rolf, and Scheier, Christian. 1999. *Understanding intelligence*. Boston, MA, USA: MIT Press.

Piaget, J. 1952. *The origins of intelligence in children*. New York, NY: Norton.

Piaget, J. 1962. *Play, dreams and imitation in childhood*. New York: Norton Press.

Ploghaus, A., Tracey, I., Clare, S., Gati, J.S., Rawlins, J.N.P., and Matthews, P.M. 2000. Learning about pain: the neural substrate of the prediction error of aversive events. *PNAS*, **97**, 9281–9286.

Quaade, F., Vaernet, K., and Larsson, S. 1974. Stereotaxic stimulation and electrocoagulation of the lateral hypothalamus in obese humans. *Acta Neurochir.*, **30**, 111–117.

Redgrave, P., Prescott, T., and Gurney, K. 1999. Is the short latency dopamine response too short to signal reward error? *Trends in Neurosciences*, **22**, 146–151.

Ring, M. 1994. *Continual Learning in Reinforcement Environments*. Ph.D. thesis, University of Texas at Austin.

Rochat, P. 2002. Ego function of early imitation. In: Melzoff, A., and Prinz, W. (eds), *The Imitative Mind : Development, Evolution and Brain Bases*. Cambridge University Press.

Rolls, Edmund T. 1999. *The Brain and Emotion*. Oxford UP.

Roth, G., and Dicke, U. 2005. Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, **9**(5), 250–257.

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

Schaffer, H. 1977. Early interactive development in studies of Mother-infant interaction. Pages 3–18 of: *Proceedings of Loch Lomonds Symposium*. New York: Academic Press.

Schilder, P. 1935. *L'image du corps*. ed 1968 edn. Paris, France: Gallimard.

Schmidhuber, J. 1991. Curious model-building control systems. Pages 1458–1463 of: *Proceeding International Joint Conference on Neural Networks*, vol. 2. Singapore: IEEE.

Schmidhuber, J. 1992. Learning complex, extended sequences using the principle of history compression. *Neural Computation*, **4**(2), 234–242.

Schmidhuber, J. 2006. Optimal Artificial Curiosity, Developmental Robotics, Creativity, Music, and the Fine Arts. *Connection Science*, **18**(2), 173–187.

Schultz, W. 1998. Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, **80**, 1–27.

Schultz, W. 2006. Behavioral theories and the neurophysiology of reward. *Annual review of psychology*, **57**, 87–115.

Schultz, W., Dayan, P., and Montague, P.R. 1997. A neural substrate of prediction and reward. *Science*, **275**, 1593–1599.

Shepherd, G. M. 1988. A basic circuit for cortical organization. Pages 93–134 of: Gazzaniga, M. (ed), *Perspectives in memory research*. MIT Press.

Smith, A., Li, M., Becker, S., and Kapur, S. 2006. Dopamine, prediction error and associative learning: a model-based account. *Network: Computation in Neural System*, **17**, 61–84.

Steels, L. 1994. The Artificial Life Roots of Artificial Intelligence. *Artificial Life Journal*, **1**(1), 89–125.

Steels, L. 2003. Intelligence with representation. *Philosophical Transactions of the Royal Society A*, **361**(1811), 2381–2395.

Steels, L. 2004. The Autotelic Principle. Pages 231–242 of: Fumiya, I., Pfeifer, R., Steels, L., and Kunyoshi, K. (eds), *Embodied Artificial Intelligence*. Lecture Notes in AI, vol. 3139. Berlin: Springer Verlag.

Stellar, J.R. 1985. *The neurobiology of motivation and reward*. New York, NY.: Springer Verlag.

Storck, J., Hochreiter, S., and Schmidhuber, J. 1995. Reinforcement driven information acquisiation in non-deterministic environments. Pages 159–164 of: *Proceedings of ICANN 1995*, vol. 2. EC2 and CIE.

Suri, R.E., and Schultz, W. 2001. Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, **13**, 841–862.

Sutton, R.S. 1988. Learning to predict by the methods of temporal differences. *Machine Learning*, **3**(1), 9–44.

Sutton, R.S., and Barto, A.G. 1998. *Reinforcement learning: an introduction*. Cambridge, MA.: MIT Press.

Sutton, R.S., Precup, D., and Singh, S. 1999. Between MDPSs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, **112**, 181–211.

Tani, J. 2007. On the interactions between top-down anticipation and bottom-up regression. *Frontiers in Neurorobotics*, **1**.

Tani, J., and Nolfi, S. 1999. Learning to perceive the world as articulated : An approach for hierarchical learning in sensory-motor systems. *Neural Network*, **12**, 1131–1141.

Thorndike, E.L. 1911. *Animal intelligence: experimental studies*. New York, NY.: MacMillan.

Turing, A. 1950. Computing machinery and intelligence. *Mind*, **59**, 433–460.

Varela, F.J., Thompson, E., and Rosch, E. 1991. *The embodied mind : Cognitive science and human experience*. Cambridge, MA: MIT Press.

von Uexkull, J. 1909. *Umwelt und Innenwelt der Tiere*. Berlin: Springer.

Vygotsky, L. 1978. *Mind in society*. Harvard university press. the development of higher psychological processes.

Walkins, C.J.C.H., and Dayan, P. 1992. Q-learning. *Machine learning*, **8**, 279–292.

Warnier, J-P. 1999. *Construire la culture materielle. L'homme qui pensait avec les doigts*. PUF.

Weinberger, D. R. 1987. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch Gen Psychiatry*, **44**, 660–669.

Weiner, I., and Joel, D. 2002. Dopamine in schizophrenia: dysfunctional information processing in basal ganglia-thalamocortical split circuits. Pages 417–472 of: Chiara, G.D. (ed), *Handbook of experimental pharmacology, vol 154/II, Dopamine in the CNS II*. Springer.

White, N. M. 1989. Reward or reinforcement: what's the difference? *Neuroscience Biobehavorial Review*, **13**, 181–186.

White, R.N. 1959. Motivation reconsidered: The concept of competence. *Psychological review*, **66**, 297–333.

Wiering, M., and Schmidhuber, J. 1997. HQ-Learning. *Adaptive Behavior*, **6**(2), 219–246.

Wise, R.A. 1989. The brain and reward. Pages 377–424 of: Liebman, J.M., and Cooper, S.J. (eds), *The neuropharmacological basis of reward*. Clarendon Press.

Yoshimoto, K., McBride, W.J., Lumeng, L., and Li, T.-K. 1991. Alcohol stimulates the release of dopamine and serotonin in the nucleus accumbens. *Alcohol*, **9**, 17–22.