Intrinsically Motivated Learning of Real World Sensorimotor Skills with Developmental Constraints

Pierre-Yves Oudeyer, Adrien Baranes, Frédéric Kaplan

Intrinsically Motivated Learning of Real World Sensorimotor Skills with Developmental Constraints

Pierre-Yves Oudeyer[1], Adrien Baranes[1], Frédéric Kaplan[2]

[1] INRIA, France
[2] EPFL-CRAFT, Switzerland

Abstract. Open-ended exploration and learning in the real world is a major challenge of developmental robotics. Three properties of real-world sensorimotor spaces provide important conceptual and technical challenges: unlearnability, high-dimensionality and unboundedness. In this chapter, we argue that exploration in such spaces needs to be constrained and guided by several combined developmental mechanisms. While intrinsic motivation, i.e. curiosity-driven learning, is a key mechanism to address this challenge, it has to be complemented and integrated with other developmental constraints, in particular: sensorimotor primitives and embodiment, task space representations, maturational processes (i.e. adaptive changes of the embodied sensorimotor apparatus), and social guidance. We illustrate and discuss the potential of such an integration of developmental mechanisms in several robot learning experiments.

A central aim of developmental robotics is to study the developmental mechanisms that allow life-long and open-ended learning of new skills and new knowledge in robots and animals (Asada et al., 2009; Lungarella et al., 2003; Weng et al., 2001). Strongly rooted in theories of human and animal development, embodied computational models are built both to explore how one could build more versatile and adaptive robots, as in the work presented in this chapter, and to explore new understandings of biological development (Oudeyer, 2010).

Building machines capable of open-ended learning in the real world poses many difficult challenges. One of them is exploration, which is the central topic of this chapter. In order to be able to learn cumulatively an open-ended repertoire of skills, developmental robots, like animal babies and human infants, shall be equipped with task-independent mechanisms which push them to explore new activities and new situations. However, a major problem is that the continuous sensorimotor space of a typical robot, including its own body as well as all the potential interactions with the open-ended surrounding physical and social environment, is extremely large and high-dimensional. The set of skills that can potentially be learnt is actually infinite. Yet, within a life-time, only a small subset of them can be practiced and learnt. Thus the central question: how to explore and what to learn? And with this question comes an equally important question: What *not* to explore and what *not* to learn? Clearly, exploring randomly and/or trying to learn all possible sensorimotor skills will fail.

Exploration strategies, mechanisms and constraints are needed and appear in two broad interacting families in animals and humans: internally guided exploration and socially guided exploration. Within the large diversity of associated mechanisms, as we will illustrate in this article, intrinsic motivation, a peculiar example of internal mechanism for guiding exploration, has drawn a lot of attention in the recent years, especially when related to the issue of open-ended cumulative learning of skills as shown by other chapters in this book (**??**).

Intrinsic motivation was identified in humans and animals as the set of processes which push organisms to spontaneously explore their environment even when their basic needs such as food or water are satisfied (Berlyne, 1960; Deci and Ryan, 1985; White, 1959). It is related to curiosity-driven learning and exploration, but is actually broader since it applies for example to the processes that push us to persist in trying to solve puzzles or improve our sport skills when not driven by extrinsic motivations such as the search for social status or money. A very large body of theories of intrinsic motivation, and its interaction with extrinsic motivation, has flourished in psychology and educational sciences at least since the middle of the 20th century (Ryan and Deci, 2000). Many of them have consisted in trying to understand which features of given activities could make them intrinsically motivating or "interesting" for a particular person at a particular moment of time. In this context, "interestingness" was proposed to be understood as related to concepts such as novelty (Hull, 1943; Montgomery, 1954), reduction of cognitive dissonances (Festinger, 1957; Kagan, 1972), optimal incongruity (Berlyne, 1960; Hunt, 1965), effectance and personal causation (De Charms, 1968; White, 1959), or optimal challenge (Csikszentmihalyi, 1996).

Following those ideas, either a priori or a posteriori, many computational systems were built to formalize, implement and evaluate intrinsically motivated exploration and learning, also referred as curiosity-driven machine learning or active learning (Lopes and Oudeyer, 2010). These models came from various fields such as statistics and "optimal experiment design" (e.g. Fedorov, 1972), active learning (e.g. Angluin, 1988; Castro and Novak, 2008; Chaloner and Verdinelli, 1995; Cohn et al., 1994; Thrun, 1992, reinforcement learning (e.g. Barto et al., 2004; Brafman and Tennenholtz, 2001; Schmidhuber, 1991; Sutton, 1990; Szita and Lorincz, 2008), computational neuroscience (e.g. Dayan and Belleine, 2002; Doya, 2002) and developmental robotics (e.g. Baranes and Oudeyer, 2009; Blank et al., 2002; Hart and Grupen, 2008; Huang and Weng, 2002; Oudeyer and Kaplan, 2007; Oudeyer et al., 2007; Schembri et al., 2007a; Schmidhuber, 2006, 2010 and **??**). Correspondingly, many formal measures of "interestingness", either heuristics or optimal regarding some criteria - and associated algorithms to compute them - were devised, including principles such as the maximization of prediction error (Meyer and Wilson, 1991; Thrun, 1992), the local density of already queried/sampled points (Whitehead, 1991), the maximization of the decrease of the global model variance (Cohn et al., 1996), or maximal uncertainty of the model (Thrun and Moller, 1992), among others. Those principles and algorithms were then integrated in various framings, a particularly interesting one being intrinsically motivated reinforcement learning, allowing to ap-

proach sequential decision problems in a unified approach (e.g. Barto et al., 2004; Schmidhuber, 1991 and **??**), and see (Kakade and Dayan, 2002; Sutton, 1990) and **??** for a related approach using exploration bonuses).

In spite of this diversity of techniques, many of these computational approaches where not designed initially for developmental learning and make assumptions that are incompatible with their use for learning in real developmental robots. Indeed, a combination of the following assumptions, which do not hold for a developmental robot, is often made for active exploration models:

- **Assumption 1**: It is possible to learn a model of the complete world/space within the life-time of the learning agent;
- **Assumption 2**: The world is learnable everywhere;
- **Assumption 3**: The noise is homogeneous;

These assumptions are very useful and relevant when the goal is to have a machine learn a predictive or control model of a whole bounded relatively small domain (e.g. sensorimotor space) and when it is yet very expensive to make one single measure/experiment. Examples include the control of automatic biological or chemical experiments (Faller et al., 2003; Kumar et al., 2010), or learning to visually recognize a finite set of visual categories (Tong and Chang, 2001). The associated techniques, such as for example those based on principles such as "search for maximal novelty or uncertainty", allow the learner to efficiently minimize the number of necessary experiments to perform in order to acquire a certain level of knowledge or a certain level of competence for controlling the given domain.

Furthermore, the models designed explicitly in a developmental learning framing were often elaborated and experimented in simple simulated worlds, even sometimes in discrete grid worlds, which allowed researchers to perform easily systematic experiments but introduced a bias on the properties of sensorimotor spaces. As a consequence, many of these models consisted in mechanisms that either also implicitly made the assumptions described in the previous paragraph, or could not (or were not shown in practice) to scale to real-world high-dimensional robot spaces.

Yet, the challenges of exploration and developmental learning become very different as soon as one uses real high-dimensional redundant bodies, with continuous sensorimotor channels, and an open-ended unbounded environment. Real sensorimotor spaces introduce three fundamental properties to which exploration and learning mechanisms should be robust:

- **Unlearnability**: There are very large regions of sensorimotor spaces for which predictive or control models cannot be learnt. Some of these regions of the sensorimotor space are definitively unlearnable, such as for example the relations between body movement and cloud movements (one cannot learn to control the displacement of clouds with ones own body actions) or the relation between the color of a cat and the color of the next car passing in the road (a developmental robot shall not be "spoon fed" with the adequate causal groupings of variables he may observe, but rather shall

discover by itself which are the sensible groupings). Some other regions of the sensorimotor space are unlearnable at a given moment of time/development, but may become learnable later on. For example, trying to play tennis is unlearnable for a baby who did not even learn to grasp objects yet, but it becomes learnable once he is a bit older and has acquired a variety of basic skills that he can re-use for learning tennis;

– **High-dimensionality**: A human child has hundreds of muscles and hundreds of thousands of sensors, as well as a brain able to generate new representational spaces based on those primitive sensorimotor channels, that are used by the organism to interact with novel objects, activities, situations or persons. Given this apparatus, even single specific skills such as hand-eye coordination or locomotion involve continuous sensorimotor spaces of very high low-level dimensions. Furthermore, action and perception consist in manipulating dynamic sequences within those high-dimensional spaces, generating a combinatorial explosion for exploration. As explained below, this raises the well known problem of the curse-of-dimensionality(Bishop, 1995), which needs to be addressed even for single specific skill learning involving the learning of forward and inverse models given a control space and a task space (Nguyen-Tuong and Peters, 2011; Sigaud et al., 2011);
– **Unboundedness**: Even if the learning organism would have a sort of "oracle" saying what is learnable and what is not at a given moment of time, real sensorimotor spaces would still have the property of unboundedness: the set of learnable predictive models and/or skills is infinite and thus much larger than what can be practiced and learnt within a life-time. Just imagine a one year-old baby who is trying to explore and learn how to crawl, touch, grasp and observe objects from various manners. First of all, with a given object in a given room, say for example a book, there is a very large amount of both knowledge and skills, of approximately equal interest for any measure of interestingness, to be learnt: e.g. learning to throw the book in various boxes in the room, at various lengths, with a various number of flips, with a final position on various sides, using various parts of the body (hands, shoulders, head, legs, ...), learning to predict the sequence of letters and drawings in it, learning to see what kind of noise it makes when torn up at various places with various strengths, hit on various objects, learning how it tastes, learning how it can fly out of the window, learning how individual pages fly when folded in various manners, ... Now, imagine what the same child may learn with all the other toys and objects in the room, then with all the toys in the house and in the one of neighbours. As it can walk, of course the child could learn to discover the map of his garden, and of all the places he could crawl to. Even with no increase of complexity, the child could basically always find something to learn. Actually, this would even apply if there would be no objects, no house, no gardens around the child: the set of skills he could learn to do with its sole own body, conceptualized as an "object/tool" to be discovered and learnt, is already unbounded in many respects. And even if obviously there are some cognitive and physical bounds on what he can learn (e.g. bounds on the possible speeds one can run at), those bounds are

initially unknown to the learner (e.g. the child initially does not know that it is impossible to run over a certain speed, thus he will need mechanisms to discover this and avoid spending its life trying to reach speeds that are not physically possible), and this is part of what is here called the challenge of unboundedness.

Considering some particular aspects of those challenging properties, in particular related to unlearnability, some specific computational models of "interestingness" and intrinsically motivated exploration were elaborated. In particular, measures of interestingness based on the *derivative* of the evolution of performances of acquired knowledge or skills, such as maximal increase in prediction errors, also called "learning progress" (Oudeyer et al., 2007; Schmidhuber, 1991), maximal compression progress (Schmidhuber, 2006), or competence progress (Bakker and Schmidhuber, 2004; Baranes and Oudeyer, 2010a; Modayil et al., 2010; Stout and Barto, 2010) were proposed. These measures resonate with some models in active learning, such as related to the principle of maximal decrease of model uncertainty (e.g. Cohn et al., 1996), but were sometimes transformed into heuristics which make them computationally reasonable in robotic applications (e.g. Oudeyer et al., 2007; Schmidhuber, 1991), which is not necessarily the case for various theoretically optimal measures. These measures also resonate with psychological theories of intrinsic motivation based on the concept of "optimal level" (e.g. Berlyne, 1960; Csikszentmihalyi, 1996; White, 1959; Wundt, 1874), which state that the most interesting activities are those that are neither too easy nor too difficult, i.e. are of the "right intermediate complexity". Yet, if modeled directly, "optimal level" approaches introduce the problem of what is "intermediate", i.e. what is/are the corresponding threshold(s). Introducing the *derivative* of knowledge or competences, such as in prediction progress or competence progress-based intrinsic motivations, allows us to transform the problem into a maximization problem where no thresholds have to be defined, and yet allowing learning agents to focus in practice on activities of intermediate complexity (e.g. Oudeyer et al., 2007).

A central property of the "interestingness" measures based on the increase of knowledge or competences is that they can allow a learning agent to discover which activities or predictive relations are unlearnable (and even rank the levels of learnability), and thus allow it to avoid spending too much time exploring these activities when coupled with an action-selection system such as in traditional reinforcement learning architectures, and where the reward is directly encoded as the derivative of the performances of the learnt predictive models of the agent (Schmidhuber, 1991). This has been demonstrated in various computational experiments (Baranes and Oudeyer, 2009; Oudeyer and Kaplan, 2006; Schmidhuber, 1991, 2006). An interesting side-effect of these measures, used in an intrinsic motivation system and in dynamical interaction with other brain modules as well as the body and the environment, is also the fact that it allows the self-organization of developmental stages of increasing complexity, sharing many similarities with both the structural and statistical properties of developmental trajectories in human infants (Oudeyer et al., 2007). Some models even

suggest that the formation of higher-level skills such as language and imitation bootstrapping could self-organize through intrinsically motivated exploration of the sensorimotor space and with no language specific biases (Oudeyer and Kaplan, 2006).

Yet, those approaches to intrinsically motivated exploration and learning address only partially the challenge of unlearnability, and leave largely unaddressed the challenges of high-dimensionality and unboundedness in real robots. First of all, while efforts have been made to make these approaches work robustly in continuous sensorimotor spaces, computing meaningful associated measures of interest still requires a level of sampling density which make those approaches become more and more inefficient as dimensionality grows. Even in bounded spaces, the processes for establishing measures of interestingness can be cast into a form of non-stationary regression problem, which as most regression problems in high-dimension faces the curse-of-dimensionality (Bishop, 2007). Thus, without additional mechanisms, like the ones we will describe in this chapter, the identification of unlearnable zones where no knowledge or competence progress happens is a process that becomes inefficient in high-dimensions. The second limit of those approaches if used alone relates to unboundedness. Actually, whatever the measure of "interestingness", if it is only based in a way or another on the evaluation of performances of predictive models or of skills, one is faced with the following circular problem:

- Those measures were initially designed to efficiently guide exploration;
- Those measures need to be "measured/evaluated";
- By definition, they cannot be known in advance, and the "measure of interestingness" of a given sensorimotor subspace can only be obtained if at least explored/sampled a little bit;
- In order to obtain meaningful measures, those sub-spaces cannot be too large, and are ideally quite local;
- In unbounded spaces, by definition all localities (even at the maximal granularity allowing to obtain meaningful measures, which is anyway initially unknown to the learner) cannot be explored/sampled within a life-time;
- Thus, one has to decide which sub-spaces to sample to evaluate their interestingness, i.e. one has to find an efficient meta-exploration strategy, and we are basically back to our initial problem with an equivalent meta-problem. This meta-problem for evaluating interestingness requires a less dense local sampling of subspaces than the problem of actually learning mappings and skills within those subspaces, but as the space is unbounded and thus infinite, this theoretical decrease in required sampling density does not make the meta-problem more tractable from a computational complexity point of view.

As a matter of fact, this argument can also be made directly starting from the framing of intrinsically motivated exploration and learning within the reinforcement learning framework, i.e. intrinsically motivated reinforcement learning. Indeed, in this framework one re-uses exactly the same machinery and architectures than in more traditional reinforcement learning, but instead of using a

reward function which is specific to a given practical problem, one uses a measure of interestingness such as the ones discussed above (e.g. a reward is provided to the system when high prediction errors, or high improvement of skills/options are observed). In such a way, the system can be made to learn how to achieve sequences of actions that will maximize the sum of future discounted rewards, e.g. the sum of future discounted prediction errors or competence progress. But essentially, this defines a reinforcement learning problem which has the same structure as traditional reinforcement learning problems, and especially similar to *difficult* traditional reinforcement learning problems given that the reward function will typically be highly non-stationary (indeed, prediction errors or competences and their evolution are both locally and globally non-stationary because of learning and of the external coupling of action selection and the reward function itself). Most importantly, as all reinforcement learning problems applied to unbounded/infinite state-spaces, exploration is a very hard problem (Sutton and Barto, 1998): even if the world would be discrete but with an unbounded/infinite number of states and associated number of options, how should exploration proceed? This problem is especially acute since when a "niche" of prediction or competence progress/errors has been well-explored and learnt, it provides no more rewards and new sources of intrinsic rewards must permanently be found. And as intrinsically motivated reinforcement learning was formulated as a way to explore efficiently the world to acquire potentially general skills that may be re-used later on for solving specific problems, we can indeed recast the meta-problem we just described as a meta-exploration problem.

Existing approaches of intrinsically motivated exploration can provide efficient mechanisms which allow a robot to decide whether it is interesting to *continue* or *stop* to explore a given sensorimotor subspace (or a local predictive model or a skill/option or simply a subset of states) which it has *already* began to explore a little bit. But due to unboundedness, strategies for exploration that may allow efficient organized acquisition of knowledge and skills need also mechanisms for answering to the question: What should *not* be explored *at all*? **The main argument that is put forward in this chapter is that complementary developmental mechanisms should be introduced in order to constrain the growth of the size, dimensionality and complexity of practically explorable spaces. Those mechanisms, that we call *developmental constraints* and are inspired by human development, should essentially allow the organism to automatically introduce self-boundings in the unbounded world (including their own body), such that intrinsically motivated exploration is allowed only within those bounds, and then progressively releasing constraints and boundings to increase the volume of explorable sensorimotor spaces, i.e. the diversity of explorable knowledge and skills**. Indeed, there is actually no mystery: efficient unconstrained exploration, even intrinsically motivated, of unbounded infinite complex spaces, especially high-dimensional spaces, is impossible within a life-time. Adaptive constraints and bounds have to be introduced, and ideally these constraints should be as little ad hoc, as little hand-tuned and as little

task-specific as possible while compatible with the real world (i.e. a real body within the real physical environment).

The study of developmental constraints complementing or interacting with or even integrated within intrinsically motivated exploration and learning has been the central topic of the research outlined in this chapter. This is achieved with the long-term goal of elaborating architectures allowing a robot to acquire developmentally a repertoire of skills of increasing complexity over a significant duration (at least on the order of one month) and in large high-dimensional sensorimotor spaces in an unbounded environment (which contrasts strongly with the existing experiments with real robots lasting most often a few minutes, at best a few hours, and allowing the acquisition of a limited repertoire of skills). Most of these developmental constraints that we are investigating are strongly inspired by constraints on human infant development, from which we take the fundamental insight that *complex acquisition of novel skills in the real world necessitates to leverage sophisticated innate capabilities/constraints as well as social constraints and constraints provided by self-organization* that may unfold with time in interaction with the environment during the course of epigenesis. In the following, we will describe some of them and explain how they may facilitate, sometimes considerably, the exploration and acquisition of complex skills in real-world sensorimotor spaces, more precisely:

- **Parameterized dynamic sensori and motor primitives, also referred as muscle synergies, and their use in adequate embodiments:** Human infants do not learn to control their whole body movements "pixel by pixel". Rather, they are born with muscle synergies, i.e. neurally embedded dynamical systems that generate parameterized coordinated movements, e.g. CPGs. These motor primitives, can considerably decrease the size of the explorable space and transform complex low-level action planning problems in higher-level low-dimensional dynamical system tuning problems. As we will show, their combination with intrinsic motivation is essential for the acquisition of dynamic motor skills in experiments like the Playground Experiment (Oudeyer and Kaplan, 2006; Oudeyer et al., 2007). We will also discuss the fact that adequate body morphologies can in addition facilitate the self-organization of movement structures and thus be potentially leveraged by intrinsically motivated exploration and learning;
- **Task-level intrinsically motivated exploration:** While biological bodies are very high-dimensional and redundant, motor tasks considered individually often consist in controlling effects in relatively low-dimensional task spaces. For example, while locomotion or reaching involve the coordination of a high number of muscular fibers, these activities aim at controlling only the three-dimensional trajectory of the body center of mass or of the hand. When human infants explore such sensorimotor spaces, they directly explore what they can do in the task space/space of effects (Bremner and Slater, 2003; Rochat, 1989), and rather spend their time exploring how to produce varieties of effects with sufficient means rather than exploring all means to achieve a single effect. Doing this, they exploit the low-dimensionality of task

spaces in combination with the high redundancy of their bodies. We will argue that similarly, intrinsic motivation in robots should operate directly in tasks spaces. We will illustrate the efficiency of this approach by presenting experiments using the SAGG-RIAC competence-based intrinsic motivation system, pushing the robot to actively explore and select goals in its task space;
- **Maturational constraints:** Human infants are not born with complete access to all their potential degrees of freedom. The neural system, partly through myelination, as well as the body, progressively grow, opening for control new muscle synergies and increasing the range and resolution of sensorimotor signals. We will illustrate how such maturational processes can be modeled and adaptively coupled with intrinsic motivation in the McSAGG-RIAC system (Baranes and Oudeyer, 2011), allowing a robot to learn skills like reaching or omnidirectional locomotion not only faster, but also with a higher asymptotic performance in generalization;
- **Social guidance:** Last but not least, social interaction should be a central companion to intrinsic motivation. The interaction between those two guiding mechanisms is at the centre of educational research (Ryan and Deci, 2000). We will argue that this shall probably also become the case in developmental robots, and discuss the various kinds of bi-directional interaction between social guidance and intrinsic motivation that shall be useful for open-ended learning in the real-world.

The choice of these families of developmental constraints on intrinsic motivation was here driven by our own investigations towards addressing the challenges of unlearnability, high-dimensionality and unboundedness, and is not intended to be a comprehensive list of potentially useful mechanisms (for example, developmental biases on representations, on mechanisms for creating abstractions, on operators for combining and re-using knowledge and skills, on statistical inference should be equally important, see **??**). Furthermore, as already explained earlier, we are still very far away from being able to address the challenge of open-ended cumulative learning in unbounded spaces in the real-world, and the approaches we present are still preliminary in this respect. Thus, our goal is mainly to draw attention to potential routes that may be pursued to address a fundamental problem of developmental robotics that has been so far largely overlooked.

1 Intrinsic motivation and embodied sensorimotor primitives

1.1 Bootstrapping learning in the "great blooming, buzzing confusion"

The problem of discovering structure and learning skills in the "great blooming, buzzing confusion" of a high-dimensional body equipped with a wide diversity of sensors like the eyes, ears, nose or skin, as stated by William James (James,

1890), might seem a daunting task. Hopefully, animal and human babies do not learn to see the world pixel-by-pixel, and likewise they do not learn to control their whole body movements "pixel by pixel". Rather, they are born with neurally embedded dynamical systems which on the sensori side allow them to be able to detect and track a number of higher-level structures right from the start, and on the motor side allow them to tune motor and muscle synergies which already generate parameterized coordinated movements (d'Avella et al., 2003; Lee, 1984; Ting and McKay, 2007). Examples of innate sensori primitives include visual movement detection and tracking systems (Bronson, 1974), basic human facial expression perception (Johnson, 2001; Meltzoff and Moore, 1977), or special auditory filters tuned for speech processing in humans (Sekuler and Blake, 1994). Examples of motor primitives include central pattern generators such as for leg oscillations (Cazalets et al., 1995), synergies for reaching with the hand (d'Avella et al., 2006), closing the fingers in a coordinated manner such as used in grasping (Weiss and Flanders, 2004), or of course skills such as breathing or swallowing (Dick et al., 1993). Of course, the existence of these primitives does not avoid the fundamental need for learning, even for the most basic skills: those primitives are typically parameterized, and thus can typically be seen as parameterized dynamical systems which semantics (affordances in particular), parameter values to be set and combination for achieving given tasks have to be learnt. For example, central pattern generators are typically neurally implemented as complex dynamical system generating oscillatory movements which can be tuned by controlling a number of high-level parameters (e.g. inputs to the neural dynamical system), and learning will consist for example in discovering that such a motor primitive can be used to "move" the whole body and in learning which tuning of the dynamical system produces which movement of the whole body. Yet, these sensorimotor primitives can considerably decrease the dimensionality, and thus the size of the explorable sensorimotor spaces and transform complex low-level action planning problems in simpler higher-level dynamical system tuning problems.

The use of a repertoires of innate parameterized sensorimotor primitives has been key in some of the most advanced real-world intrinsically motivated robot experiments so far, such as in the Playground Experiment (Oudeyer and Kaplan, 2006; Oudeyer et al., 2007) or in (Hart and Grupen, 2008) where primitives were based on sophisticated control-theoretic sensorimotor feedback loops. In parallel, several projects investigating the use of options in intrinsically motivated reinforcement learning can be related to this concept of motor primitives: for example, experiments such as in (Barto et al., 2004; Stout and Barto, 2010) assumed the existence of a number of innate temporally extended skill "templates", called "options", and corresponding to macro-actions which can be conceptualized as parameterized motor primitives. In those simulations, even if the world is discrete and finite, the system is nevertheless shown to be able to learn to achieve complex skills corresponding to long sequences of actions that are extremely difficult to learn with standard exploration procedures and only low-level actions. In other words, those simulations also provide examples of how innate motor

primitives can leverage the potentialities of intrinsically motivated exploration. To give a more precise illustration of such uses of sensorimotor primitives with intrinsic motivation, we will now outline the Playground Experiment (Oudeyer and Kaplan, 2006; Oudeyer et al., 2007). Other experiments such as those described in the references above would be equally relevant to illustrate this point and the reader is referred to them for more details.

1.2 Intrinsically motivated acquisition of affordances and skills in the Playground Experiment

The Playground Experiment was introduced as an experimental set-up allowing us to show how one particular kind of intrinsic reward system, called "Intelligent Adaptive Curiosity" (IAC), could allow a real-world robot with high-dimensional continuous sensorimotor channels to acquire continuously new skills of increasing complexity. As detailed in (Oudeyer et al., 2007), the central idea of IAC (which was later importantly refined in R-IAC (Baranes and Oudeyer, 2009)) was to push the robot to explore certain dynamic motor activities in certain sensorimotor contexts where its predictions of the consequences of its actions in given contexts were improving maximally fast, similarly to what was also proposed in (Schmidhuber, 1991). A specificity of IAC was the introduction of algorithmic heuristics allowing us to compute prediction progress efficiently and robustly in relatively large continuous sensorimotor spaces. Such an approach based on the optimization of prediction progress belongs to the family of "knowledge-based" intrinsic motivation systems (Oudeyer and Kaplan, 2008). Yet, even if driven by the acquisition of knowledge, the whole process is fundamentally active (active choice of actions or sequences of actions in given sensory contexts) and the forward models that are learnt can easily and efficiently be re-used for control as soon as one uses non-parametric statistical approaches such as in memory based approaches such as those presented in (Moore, 1992) and adapted in (Oudeyer et al., 2007), or multimap learning approaches such as in (Calinon et al., 2007; Ghahramani, 1993), or a mixture of non-parametric and multimap approaches (Cederborg et al., 2010). As a consequence, such active knowledge acquired through knowledge-based intrinsically motivated exploration can readily and directly be used for efficient control, as quantitatively shown in (Baranes and Oudeyer, 2009), and thus the IAC system allows the robot to learn a repertoire of *skills* of progressively increasing complexity in the Playground Experiment.

1.2.1 Parameterized Motor primitives in the Playground Experiment. As argued in the previous paragraph, previous articles presenting the Playground Experiment largely focused on the study of IAC and its role in the obtained results. Yet, a second essential ingredient was the use of parameterized dynamic motor primitives as well as sensory primitives, on top of which exploration and learning was actually happening. Here we try to emphasize the role of those innate (but still very plastic thanks to their parameters) structures to show two complementary points:

– These parameterized motor primitives consist in complex closed-loop dynamical policies which are actually temporally extended macro-actions (and thus could very well be described in terms of options), that include at the low-level long sequences of micro-actions, but controlled at the high-level only through the setting of a few parameters; Thus, behind the apparently "single-step look ahead" property of the system at the higher-level, the Playground Experiment shows the acquisition of skills consisting in complex *sequences* of actions;
– The use of those parameterized motor primitives allows the robot to encode those whole sequences of micro-actions into constrained compact low-dimensional static projections that permit an exploration with IAC that is made considerably easier than if all physically possible movements had been made possible and explorable "pixel-by-pixel";

1.2.2 Experimental setup and description of primitives The Playground Experiment setup involves a physical Sony AIBO robot which is put on a baby play mat with various toys, some of which affording learnable interactions, and an "adult robot" which is pre-programmed to imitate the vocalization of the learning robot when this later robot produces a vocalization while looking in the direction of the adult robot (see figure 1). The AIBO robot is equipped with four legs, each equipped with three degrees of freedom controlled by servomotors (the degrees of freedom are not controlled directly, but through the many dimensions of the control architecture of the motors), with one head with four degrees of freedom including a mouth, with a loudspeaker, as well as with a video camera, an infra-red distance sensor mounted on the chin and a microphone. Here, the back legs are blocked so that the robot is not able to locomote, similarly to young human infants in the first months of their life. Given such a rich sensorimotor apparatus in such an environment, it is clear that if action generation and exploration started at the level of millisecond-wide force commands in the motors and no further constraints on the movement profiles were added, and if perception started from the level of individual camera pixels or millisecond-wide spectrogram auditory features, the sensorimotor space would be so large that learning and development would be highly inefficient if not impossible.

In order to avoid this problem, the following parameterized motor and sensory primitives are made available to the robot, and can be used either alone or in combination (concurrently or in sequence in theory, but so far the Playground Experiment was only made with the concurrency combination operator):

– **Bashing motor primitive:** This motor primitive allows the robot to produce a bashing movement with either one of its two fore legs, and is parameterized by two real numbers indicating the strength and angle of the bashing movement. Based on these parameters, a lower-level control theoretic architecture first selects the appropriate group/synergy of motors to be used (depending on the angle, motors of the right or left leg shall be used), and then starting from a template movement of the tip of the leg in

Fig. 1. The Playground Experiment setup involves a physical Sony AIBO robot which is put on a baby play mat with various toys, some of which affording learnable interactions, and an "adult robot" which is pre-programmed to imitate the vocalization of the learning robot when this later robot produces a vocalization while looking in the direction of the adult robot. The learning robot is equipped with a repertoire of innate parameterized sensorimotor primitives, and learns driven by intrinsic motivation how to use and tune them to affect various aspects of its surrounding environment. Complex self-organized developmental trajectories emerge as a result of intrinsically motivated exploration, and the set of acquired skills and affordances increases along with time.

its operational/task space (Khatib, 1987), uses it to define a target trajectory to be followed by the tip of the leg with a certain acceleration profile (corresponding to the force parameter), which is then passed to a lower-level closed-loop action-selection mechanism which generates the appropriate motor currents/torques, in response to real-time position/speed/acceleration errors measured proprioceptively within the motor, at a frequency around 1 kHz and based on a standard PID algorithm (Chung et al., 2008). As a consequence, once the two high-level parameters of the primitive have been set, an automatic dynamical system/policy is generated and is launched to control leg movements, which thanks to the low-level PID servoing controller react and are robust to potential external perturbations. While the parameterization of these bashing movements compresses drastically the generation of movement, it still allows the robot to produce a constrained but very large

number of movements that are not unlike the reaching primitives of young human infants. Also to be noted is the fact that special values (-1,-1) are used for the parameters to inhibit the primitive (it is not launched), and this applies to all other motor primitives. Concurrency and combination of primitives are managed through this encoding mechanism;

- **Crouch biting motor primitive:** This motor primitive allows the robot to crouch down while opening the mouth and then finally closing the mouth, in which an object may potentially be bitten. It is parameterized by the amplitude of the crouch movement, and optionally by a parameter controlling the timing of the closing of the mouth. Furthermore, this primitive is such that it keeps the orientation of the head as it is before the primitive is launched, which basically allows to have the effect of this primitive partially controlled by the use of other concurrent motor primitives controlling the head, such as the "Turn head" primitive below. Once the parameters are set, a control-theoretic low-level system very similar to the one for the bashing motor primitive is launched: given the set of motors associated with this primitive, here those of the two legs and of the mouth, and a reference target trajectory of each of these motors (which shape is spline-like) and directly controlled by the parameters, a low-level PID based motion tracking system is launched to control the low-level sequences of motor torque/current commands;
- **Turning the head motor primitive:** This motor primitive allows the robot to direct its head in a direction determined by two parameters controlling its head pan and tilt. Again, those parameters trigger a lower-level control loop that gets the head from the current position to the desired orientation through low-level torque control. This motor primitive is essential for the robot since the head supports the camera and the infra-red sensor. Thus, this motor primitive allows the robot to direct its sensors in given directions of the environment;
- **Vocalization motor primitive:** This motor primitive allows the robot to produce vocalizations consisting of prototypical "baba" like sounds which are parameterized by their fundamental frequency, more precisely their mean pitch value. Of course the AIBO robot does not have a physical vocal tract, so a speech synthesizer is used instead (which may be seen himself as a model of a vocal tract), and the dynamic sounds to be produced are constrained to be basic syllables, corresponding to "canonical babbling" (innate stereotypical coordinated actions of many muscles in the human mouth (MacNeilage, 2008)), for which the robot is only authorized to modify the mean pitch;
- **Movement sensory primitive:** This sensory primitive allows the robot to assess whether something is moving, e.g. oscillating, right in the direction in front of its nose where the infrared sensor is positioned. It basically consists in a filter operating on short time windows of the past infrared sensor values, which is then saturated to provide a binary value (0 if no movement is detected, and 1 if a movement is detected). This sensory primitive is not unlike the innate movement detectors of the visual system of human infants (Bronson, 1974);

- **Visual object detection sensory primitive:** This sensory primitive allows the robot to assess whether an "object" is visually present in its narrow field of view, an object being defined as a group of pixels with certain saliency properties. In the Playground experiment, those measures of saliency were short cut by the use of visual markers directly put on objects to be perceived as "salient". This sensory primitive thus provides high-level filters upon the pixel matrix of the camera, which are functionally not unlike the facial presence and facial expression innate detectors in human infants ;
- **Mouth grabbing sensory primitive:** This sensory primitive allows the robot to assess whether he is holding something in the mouth or not, and relies on the use of a filter above the proprioceptive sensors in the mouth, saturated so that the end value is also binary;
- **Auditory pitch sensor:** This sensory primitive allows the robot to measure the mean pitch of the sounds perceived in the short past, typically being a continuous value when a vocalization has been produced by the other robot and either a value 0 or a random value for noises produces during motor interaction with objects. This sensor is automatically disabled while the learning robot is producing its own vocalizations (but it could very well not be the case, which would also produce interesting behaviours).

1.2.3 What the robot may explore and learn The sensorimotor primitives described in the previous paragraph constitute a significant amount of innate structures provided to the robot. Yet, those motor and sensory motor primitives are tools which semantics, parameterization, combination and affordances both among themselves and with the environment should be learnt. Indeed, from the point of view of the robot, each of these primitives are black boxes in which uninterpreted numbers can be sent, and from which uninterpreted numbers can be read. The relations among those black boxes, especially between the motor and sensory primitives, are also totally unknown to the robot. In practice, this means that the robot does not know initially things such as the fact that using the "bashing primitive" can produce controllable values in the "movement sensory primitive" (an object oscillating after being bashed) and when applied in certain regions of its sensorimotor space (with particular parameter values in relation to the position of an object), and in coordination with the "turning head primitive" which allows to direct the sensors in the direction of the physical effect of the bashing primitive. Another example is that the robot shall not know that predictable, and thus controllable, auditory sensations corresponding to the adult robot's vocalization shall be triggered by vocalizing while at the same time looking in the direction of the other robot, and the robot shall not know how particular parameters of the vocalization itself can affect the imitation of the adult robot (which is by the way perceived just as other standard salient "objects"). As a result, in the Playground Experiment, the robot has to explore and learn how the use of the motor primitives, with their continuous space of parameters, as well as their concurrent combination (e.g. bashing with given parameters achieved concurrently with turning the head with given pa-

rameters), allows (or does not allow) to predict and control the values of subsets of the sensory primitives. Details are described in (Oudeyer and Kaplan, 2006; Oudeyer et al., 2007).

In spite of the fact that those motor and sensory primitive considerably constrain and reduce the sensorimotor space to be explored for acquiring action knowledge and thus skills, it is still a rather large space in comparison with the physical time necessary to achieve one single sensorimotor experiment (i.e. experimenting how a vocalization with certain parameters might make a toy move - actually the robot shall discover that this is not possible - or make one of the surrounding objects - the adult robot - produce another vocalization). Indeed, in its most simple version, the above primitives still define a 6 dimensional continuous space of motor parameters and a 4 dimensional space of sensory parameters, constituting a 10 dimensional sensorimotor space. Such a dimensionality shall not be daunting for sampling and modeling an abstract space in a computer simulation where individual experiments last a few milliseconds and are thus cheap in terms of time. Yet, for robots as for living animals, actions take time and a reaching, bashing or vocalization attempt lasts at least two or three seconds. As a consequence, sampling randomly the space of sensorimotor experiments would lead to inefficient and slow acquisition of the learnable skills. This is the reason why the use of sensorimotor primitives in combination with intrinsically motivated exploration can really allow each mechanism to leverage the potentialities of each other.

1.2.4 Results of experiments We outline here the various results that came out of repeated experiments. For further details, the reader is referred to (Oudeyer and Kaplan, 2006; Oudeyer et al., 2007). During an experiment, which lasts approximately half-a-day, we store all the flow of values of the sensorimotor channels, as well as a number of features that help us to characterize the dynamics of the robot's development. Indeed, we measure the evolution of the relative frequency of the use of the different actuators and motor primitives (analogous measures were also used to study the behavioural structure of intrinsically motivated exploration in (Schmidhuber, 2002). In particular, we also constantly measure the direction in which the robot is turning its head. Figure 2 shows details of an example for a typical run of the experiment.

Table 1. Stages in the robot's developmental sequence

description	stage 0	stage 1	stage 2	stage 3
individuation of actions	-	+	+	+
biting and bashing with the right affordances	-	-	+	+
focused vocal interactions with the adult	-	-	-	+

Self-organization of developmental stages and affordance learning. From the careful study of the curves on figure 2, augmented with the study of the

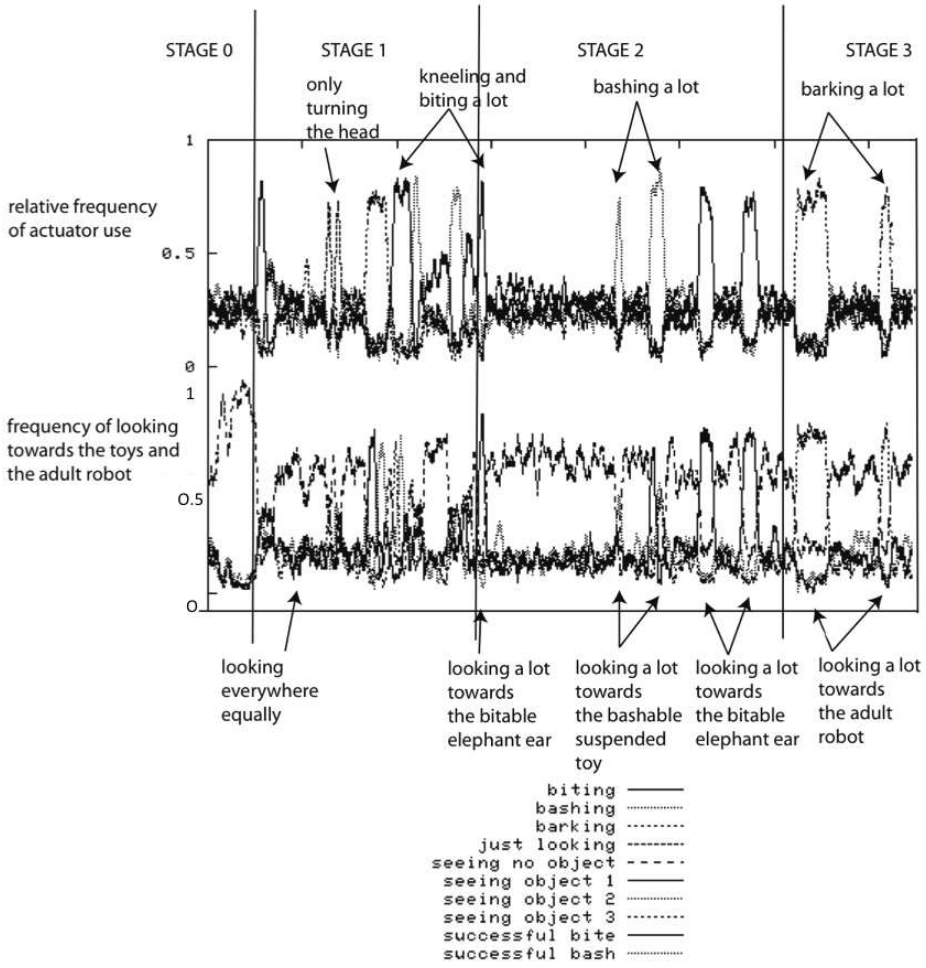

Fig. 2. Top curves: relative frequency of the use of different motor primitives in the Playground Experiment. Bottom curves: frequency of looking towards each object and in particular towards the "adult" pre-programmed robot. We can observe that the robot explores, and thus learns, progressively more complex and more affordant skills.

trace of all the situations that the robot encountered, we observe that (1) there is an evolution in the behavior of the robot; (2) this evolution is characterized by qualitative changes in this behavior; (3) these changes correspond to a sequence of more than two phases of increasing behavioral complexity, i.e. we observe the emergence of several successive levels of behavioral patterns. Moreover, it is possible to summarize the evolution of these behavioral patterns using the concept of stages, where a stage is here defined as a period of time during

which some particular behavioral patterns occur significantly more often than random and did not occur significantly more often than random in previous stages. This definition of a stage is inspired from that of Piaget (Piaget, 1952). These behavioral patterns correspond to combinations of clear deviations from the mean in the curves in figure 2. This means that a new stage does not imply that the organism is now only doing new things, but rather that among its activities, some are new. Here are the different stages that are visually denoted in figure 2 and table 1:

- **Stage 0:** The robot has a short initial phase of random exploration and body babbling. This is because during this period the sensorimotor space has not yet been partitioned in significantly different areas. During this stage, the robot's behavior is equivalent to the one we would obtain using random action selection: we clearly observe that in the vast majority of cases, the robot does not even look or act towards objects, and thus its action on the environment is quasi-absent. This is due to the fact that the sensorimotor space is vast and only in some small sub-parts some non-trivial learnable phenomena can happen given its environment.
- **Stage 1:** Then, there is a phase during which the robot begins to focus successively on playing with individual motor primitives, but without the adequate affordances: first there is a period where it focuses on trying to bite in all directions (and stops bashing or producing sounds), then it focuses on just looking around, then it focuses on trying to bark/vocalize towards all directions (and stops biting and bashing), then to bite, and finally to bash in all directions (and stops biting and vocalizing). Sometimes the robot not only focuses on a given actuator, but also looks in a focused manner towards a particular object at the same time: yet, there is no affordance between the actuator used and the object it is looking at. For example, the developing robot tries to bite the "adult" robot or to bark/vocalize towards the elephant ear. Basically, in this stage, the robot is learning to decompose its motor space into differentiable sub-parts which correspond to the use of different (combination of) motor primitives. This results from the fact that using one or two primitives at time (typically either bashing/biting/vocalizing together with turning the head in a particular direction) makes the $SM(t) \rightarrow S(t+1)$ easier to learn, and so at this stage in its development, this is what the robot judges as being the largest niche of learning progress.
- **Stage 2:** Then, the robot comes to a phase in which it discovers the precise affordances between certain motor primitives and certain particular "objects": it is now focusing either on trying to bite a biteable object (an elephant ear) and on trying to bash a bashable object (a suspended toy). Furthermore, the trace shows that it does actually manage to bite and bash successfully quite often, which shows how such capabilities can be learnt through general curiosity-driven learning since no reward specific to these specific tasks are pre-programmed. This focus on trying to do actions towards affordant objects is a result of the splitting mechanism of IAC (Oudeyer et al., 2007), which is a refinement of the categorization of the sensorimotor

space that allows the robot to see that, for example, there is more learning progress to be gained when trying to bite the biteable object that when trying to bite the suspended toy or the "adult" robot (indeed, in that case, nothing happens because they are too far, and so the situation is always very predictable and does not provide a decrease in the errors in prediction.).

– **Stage 3:** Finally, the robot comes to a phase in which it now focuses on vocalizing towards the "adult" robot and listens to the vocal imitations that it triggers. Again, this is a completely self-organized result of the intrinsic motivation system driving the behavior of the robot: this interest for vocal interactions was not pre-programmed and results from exactly the same mechanism which allowed the robot to discover the affordances between certain physical actions and certain objects. The fact that the interest in vocal interaction appears after the focus on biting and bashing comes from the fact that this is an activity which is a little bit more difficult to learn for the robot, given its sensorimotor space and the playground environment: indeed, this is due to the continuous sensory dimensions which are involved in vocalizing and listening, as opposed to the binary sensory dimensions which are involved in biting and bashing.

We made several experiments and each time we got a similar global structure in which a self-organized developmental sequence pushed the robot towards practicing and learning activities of increasingly organized complexity, particularly towards the progressive discovery of the sensorimotor affordances as well as the discovery for vocal interactions. In particular, in the majority of developmental sequences, there was a transition from a stage where the robot acted with the wrong affordances to a stage where it explored motor primitives with the right affordances and in particular finishing by a stage where it explored and focused on vocal interactions. Nevertheless, we also observed that two developmental sequences are never exactly the same, and the number of stages sometimes changes a bit or the order of intermediary stages is sometimes different. We then conducted systematic experiments to assess statistically those properties, as described in (Oudeyer et al., 2007), and we found that strong structural regularities were appearing in a statistically significant manner and at the same time that diversity of the details, and cases which varied importantly from the mean, appeared. This is particularly interesting since this duality between universal regularities and diversity in development pervades human infant development, as described in the developmental psychology literature (Berk, 2008; Fisher and Silvern, 1985), a property which has been so far only poorly understood and for which such a computational experiment suggest original hypothesis.

Formation of developmental cognitive categories. In addition to driving the exploration of the space of sensory and motor primitives and their combinations in given environmental contexts, the IAC architecture builds internal categorization structures, called "regions" (Oudeyer et al., 2007), and used to separate sensorimotor subspaces of various level of "interestingness", i.e. of various level of learnability/controllability. As argued in (Kaplan and Oudeyer, 2007), those categories initially made at the service of the intrinsic motivation

system, are formed gradually and their properties reflect important properties of fundamental general conceptual categories to be discovered by the human/robot child: in particular, it allows the learning agent to separated its own body - i.e. the self - (maximally controllable), from inanimate surrounding objects (moderately controllable) and from other living entities (less controllable but still important niches of learning progress), and finally from the unlearnable and uncontrollable. A similar approach to the bootstrapping of these fundamental cognitive categories was also presented in (Kemp and Edsinger, 2006).

Skill acquisition: From the knowledge of action consequences to direct control. In the Playground Experiment, exploration is driven by the search of maximal improvement of the predictions of the consequences of using motor primitives upon sensory primitives in given environments. Thus, it is actively driven by the acquisition of knowledge about the consequences of actions, i.e. by the acquisition of "forward models". If forward models would be encoded using parametric regression methods such as standard neural networks, then it would be complicated and highly inefficient to transform this knowledge into a competence, i.e. to reuse this knowledge to achieve practical goals, and thus one may say that the learning agent would not have learnt skills. Hopefully, research in robot learning and stochastic control theory based on statistical inference has shown that if forward models are encoded using certain forms of non-parametric models (Bishop, 2007), such as in memory-based approaches (Schaal and Atkeson, 1995), then there are simple and efficient methods to directly reuse the acquired knowledge to achieve efficient control, even in high-dimensional highly-redundant robot bodies (Baranes and Oudeyer, 2009; Moore, 1992). It has also been shown in the same literature that instead of either learning forward or inverse models with experimental data collected by a robot of the type in the Playground Experiment, one could learn both at the same time using multimap models, such as in Gaussian Mixture Regression (Calinon et al., 2007; Cederborg et al., 2010; Ghahramani, 1993). In the Playground Experiment, non-parametric models similar to (Moore, 1992) were used and thus allow the robot to acquire permanently new *skills* as it is exploring the world, even if driven by the acquisition of new *knowledge* about the consequences of its actions. A second collateral advantage of using non-parametric statistical approaches over parametric approaches such as standard neural networks is that it avoids catastrophic forgetting: new experiments by the robot allow it to acquire novel skills without forgetting any of the previously learnt knowledge and skills. In more recent experiments, such as experiments about a new version of IAC, called R-IAC, which is algorithmically more robust from several respects, a combination of non-parametric learning and multimap regression based on Gaussian mixtures was used and it was quantitatively shown how it could be reused efficiently for controlling a high-dimensional redundant body (Baranes and Oudeyer, 2009).

1.2.5 Embodiment and morphological computation We have just seen how sensorimotor primitives, viewed as dynamical systems controlling high-dimensional bodies but tuned by low-dimensional parameters, could be con-

siderably useful when combined with intrinsic motivation for learning complex sensorimotor skills in a real robot. Actually, the efficiency of those primitives is tightly related to the morphological properties of the body in which they are used. First, the inputs and structure of those primitives only make sense within a given body structure. Second, the outputs of those primitives do not entirely determine the movements/behaviour of the robot body: indeed, the physics of real-world robots is such that gravity, and its interaction with the inertia of the robot, in combination with the compliance and other dynamical properties of materials and actuators, also impact importantly the resulting movements/behaviour. Furthermore, the morphology of a robot might be more or less affordant with the environment (Gibson, 1986), and thus make the control of various aspects of the environment more or less easy to learn: for example, it will be much more difficult to learn how to grasp an egg for a robot with a gripper made of two stiff metal fingers than for a robot with a multi-finger soft compliant hand. Equally, a robot with an anthropomorphic head with a wide-angle camera will more easily trigger and perceive human social cues than a robot with no head and a narrow-angle camera directed to its foot.

Thus, the impact of morphology on control and behaviour is paramount. An adequately designed morphology can allow to significantly reduce the complexity of its traditional control code/system for a given set of tasks, and can even be conceptualized as replacing traditional digital control computations by "physical" or "morphological computation" (Paul, 2004; Pfeifer and Bongard, 2006; Pfeifer et al., 2007; Pfeifer and Scheier, 1999). A number of experiments exploring this principle have been presented in the literature, concerning skills such as grasping (Yokoi et al., 2004), quadruped locomotion (Iida and Pfeifer, 2004), robot fish swimming (Ziegler et al., 2006), insect navigation (Franceschini et al., 1992), as well as biped humanoid locomotion or emergent physical human-robot interfaces (Ly et al., 2011; Ly and Oudeyer, 2010; Oudeyer et al., 2011). The body itself, as a physical dynamical system subject to the laws of physics, should actually be considered as any other complex dynamical system, which can potentially generate spontaneously organized structures through self-organization (Ball, 1999).

As a consequence, the spontaneous structures potentially generated by the body complement and interact with the structures provided by sensorimotor primitives, and shall equally be leveraged for intrinsically motivated learning of sophisticated sensorimotor skills in the real world. Like when one uses a given set of innate sensorimotor primitives, a given body with given morphological properties is by definition particular. These innate constraints of course introduce biases: they will help the robot to acquire certain families of skills rather than other families of skills. But this is in no way incompatible with the goal of building machines capable of open-ended development and learning. Indeed, "open-ended learning" does not imply that the robot shall be able to learn universally *anything*, but rather simply that he shall be able to learn continuously novel skills. Again, due to unboundedness in particular, efficient universal skill learning in the real world is probably impossible, and constraints at all levels

need to be employed to make learning of particular *families* of skills in particular *families* of environment. This actually applies to intrinsic motivation systems themselves, for which no measure of "interestingness" might be universally useful, as argued in the evolutionary perspective presented in (Singh et al., 2010). Furthermore, using particular constraints, in particular morphological constraints, for a particular family of skills and environments does not mean either that they are necessarily ad hoc. For example the human body has very particular properties that considerably help the acquisition of a versatile and diverse repertoire of motor skills.

1.2.6 Limits and perspectives The Playground Experiment has shown how a high-dimensional robot could learn incrementally a repertoire of diverse and relatively complex skills and affordances through curiosity-driven exploration. We have argued above that these results could be obtained thanks to the use of innate parameterized motor and sensory primitives as much as to the use of an intrinsic motivation system. Yet, next to these promising results, many limits and avenues for improvement may be found in both the experimental setup and the algorithmic approach.

Firstly, while we have explained that knowledge based intrinsically motivated exploration allowed to acquire skills as a side effect when using non-parametric and/or multimap models, one could wonder what could be gained by using competence based intrinsically motivated exploration (Bakker and Schmidhuber, 2004; Baranes and Oudeyer, 2010a; Oudeyer and Kaplan, 2008; Schembri et al., 2007b; Stout and Barto, 2010), i.e. an architecture *directly* driven by the acquisition of skills (see also Rolf et al., 2010 for a related approach). In principle, there may be good reasons to use competence based approaches to mimic animal and human development given the central importance of skills for the survival of living beings, and for the usefulness of robots (see **??**). In the Playground Experiment, it may be noted that the motor parameter space is much larger and more complex than the space defined by the sensory primitives, i.e. the command/action space is much larger than the task/effect space. This is partly the result of having redundant motor primitives. If one is eventually interested in the skills that the robot may learn, and if skills are defined in terms of what changes in the external (or internal) environment the robot can produce by its actions (e.g. making an object move, being grasped, or produce sounds), then this means that one should prefer that the robot learns one strategy to achieve all possible effects rather than many strategies to achieve only a subset of potential effects. Thus, in such redundant spaces it would be interesting that exploration be driven directly by the evolution of performances for producing effects in task spaces, hence directly by evolution of competences. In section 2, we will present an example of such competence-based intrinsic motivation system where active exploration takes place directly in the task space (i.e. realizing what is sometimes called goal babbling Rolf et al., 2010), and show quantitatively how it can improve the speed of skill acquisition in a high-dimensional highly-redundant robot.

Secondly, while the use of such sensorimotor primitives in combination with intrinsic motivation is probably necessary for bootstrapping developmental learning in real sensorimotor spaces, it addresses only very partially and in a limited manner the fundamental problem of unboundedness. As shown above, the use of sensorimotor primitives can be used as a transform mapping a high-dimensional continuous problem into a much lower dimensional continuous problem. Harnessing dimensionality is fundamental, but it is nevertheless not sufficient to address unboundedness. Indeed, low-dimensional spaces could very well be themselves infinite/unbounded, e.g. one could typically have a motor or sensory primitive with parameters or values in an unbounded space, or alternatively one could very well have an infinite number of low-dimensional bounded sensorimotor primitives. In such contexts, intrinsic motivation systems face the meta-exploration problem: evaluating "interestingness" itself becomes very difficult. As argued above, unboundedness is probably an obstacle that shall not be attacked frontally. Rather, mechanisms for introducing "artificial" bounds are necessary (in addition to the bounds created by intrinsic motivation systems once interestingness has been efficiently evaluated). This is what was done in the Playground Experiment: all parameters of motor primitives, as well as all sensory values, were bounded in a compact hypercube within R^n, and there was a small number of motor primitives.

This self-bounding approach may be a bit too drastic and problematic for allowing open-ended acquisition of novel skills upon a longer life-time duration. A first aspect of the limits of such an artificial bounding is related to the very introduction of fixed relatively ad hoc bounds on the values of sensorimotor primitive parameters. It might be difficult to tune those bounds manually in order to allow the spaces to be explorable and learnable, and once the robot has reached these boundaries of what can be learnt and explore, an obstacle to further development appears. Introducing bounds is essential, but clearly autonomous open-ended development needs more flexibility. One possible way of addressing this challenge is to consider the possibility of using maturational mechanisms, inspired by the progressive growth and maturation of the body and brains of living beings, which permit to control the dynamic self-tuning and self-expansion of these bounds. This includes mechanisms controlling for example the progressive increase of the sensitivity of sensors, or of the number of degrees of freedoms and range of motor commands. Section 3 will present a system combining such maturational constraints with intrinsically motivated learning and draw some perspectives on the future challenges that this entails.

A second aspect of introducing such artificial boundings is related to the use of fixed and limited set of motor and sensory primitives. This equally limits the extent to which open-ended development may be achieved on a longer time scale. A first important direction of research in order to remove this barrier is to generalize the introduction of operators for combining primitives and make them recursive. Only a simple concurrency combination operator was available in the Playground Experiment, but many other kinds of operators could be imagined. The most obvious one is sequencing, allowing the robot to learn higher-level

skills involving plans based on the motor primitives (thus in addition to the low-level motor sequences inside the motor primitives), that may be coupled with operators allowing to encapsulate such plans/high-level skills into macros that can be re-used as atomic actions. Those objectives are at the centre of research combining intrinsically motivated reinforcement learning and option theory, and more generally approaches to cumulative learning, and the reader is referred to the following articles for an overview of those techniques (Bakker and Schmidhuber, 2004; Barto et al., 2004; Ring, 1994; Sutton et al., 1999; Wiering and Schmidhuber, 1997). A second equally important direction of research to go beyond a fixed set of sensorimotor primitives is social learning: mechanisms such as learning by imitation or demonstration may be very useful to help a robot acquire novel primitives and novel combinations of those primitives. More generally, while research on intrinsically motivated skill acquisition has largely focused on pure autonomous learning for methodological reasons, human infants learn and develop through a strong interaction of intrinsically driven learning and social guidance. Likewise, this interaction should probably be key in the strategies to be employed to face the challenge of open-ended development in an unbounded world, which we will discuss in section 4.

2 Intrinsically motivated exploration and learning directly in task spaces

As argued earlier in this article, robots are typically equipped with a very large sensorimotor space, in which motor policies are typically embedded in high-dimensional manifolds. Yet, many real world tasks consist in controlling/effecting only a limited number of sensory variables based on the use of high-dimensional motor commands. For example, a hand reaching task consists in positioning the hand in a three dimensional visual space, which contrasts with the many muscles that need to be activated to achieve this reaching. A biped locomotion task is defined in terms of the three-dimensional trajectory of the centre of mass, achieved with very high-dimensional control of all the degrees of freedom of the body. Related to this high dissimilarity between the dimensionality of many tasks and the dimensionality of their associated control/joint space, is the fact that human and robot motor systems are highly redundant. Goals in a task space (e.g. the position of the hand in three dimension) can typically be reached by many motor programs.

This property can importantly be exploited to design intrinsic motivation systems that drive exploration in such high-dimensional redundant spaces[3]. Knowledge-based intrinsic motivation systems and traditional active learning heuristics drive exploration by the active choice of motor commands and measure of their consequences, which allows to learn forward models that can be re-used as a side effect for achieving goals/tasks: this approach is sub-optimal in

[3] Part of the material presented in this section is adapted from (Baranes and Oudeyer, 2010a, tted)

many cases since it explores in the high-dimensional space of motor commands and considers the achievement of tasks only indirectly. A more efficient approach consists in directly actively exploring the space of goals within task spaces, and then learn associated local coupled forward/inverse models (possibly through local goal-driven active exploration) that are useful to achieve those goals. For example, if we consider the learning of a hand reaching task, the knowledge based approach would actively sample the set of joint motor commands and observe the resulting three dimensional hand position. This exploration process will not consider the distribution of explored hand position, and in addition to being embedded in a high-dimensional space if the arm has many degrees of freedom, may lead to learning many joint motor commands that produce the same hand position, while not necessarily learning how to reach many other hand positions. On the other hand, task-level exploration will directly and actively explore the space of goals, actively choosing three dimensional hand configurations to be reached and then launch a lower-level process for exploration of the joint space directed to the selected goal. Here, rather than learning many motor programs allowing to reach one goal, the robot will learn to reach many goals, maybe with few motor solutions for each goal. This allows to exploit redundancy and low-dimensionality of the task space. Such a task-level approach belongs the family of competence based intrinsic motivation systems (Oudeyer and Kaplan, 2008).

In the next section, we illustrate how this approach can be useful in the context of the SAGG-RIAC architecture. Actually, as it happens in SAGG-RIAC, task level/goal exploration and control level/joint exploration can be integrated in a single hierarchical active learning architecture. This architecture is organized in two levels: at a higher level, the robot chooses actively goals to explore (for example points in the visual space that may be reached by its hand), and at a lower level the robot actively performs local exploration to learn how to reach goals selected at the higher level. Hence, globally the exploration is guided by motor exploration in the task space, where goals are defined as particular configurations to reach (possibly under certain constraints, e.g. a goal may be to reach a given position with the tip of the arm through a straight line or while minimizing the spent energy). Yet, in spite of having a task space which dimensionality can be considerably smaller than the control space (e.g. often below five), sophisticated exploration heuristics have to be used due to a specific novel problem that appears in goal babbling/task level exploration. Indeed, a human or a robot does not know initially what parts of the task space are "reachable": the robot knows neither its learning limits nor its physical limits. If we take again the example of the reaching task, initially the robot will not know which part of the three dimensional visual space can or cannot be reached with its hand. Some goals may be impossible to reach because of physical limitation, some other goals may be too difficult to learn to reach given its inference capabilities, and some other goals may be too difficult to reach now but become reachable later on after learning basic motor programs that can be re-used for these more difficult goals. Thus, efficient exploration requires that the robot identifies quickly the parts of the task space where goals are not reachable at a given point of its

development, and focus exploration on trying to learn goals that are actually reachable, and thus learnable. This directly leads to the idea of transposing the concept of "prediction improvement' - characterizing the (non-)interestingness of motor commands in knowledge-based architectures - into a concept of "competence improvement" - characterizing the (non-)interestingness of goals in the task space.

2.1 SAGG-RIAC: multi-level active learning

In order to illustrate the interest of task level exploration, we outline here the SAGG-RIAC active learning architecture, introduced in (Baranes and Oudeyer, 2010a), and present it in the context of a reaching task example (but it can be trivially adapted to other sensorimotor spaces). We also present experiments evaluating the gain compared to knowledge-based architectures such as R-IAC. SAGG-RIAC transposes some of the basic ideas of R-IAC, combined with ideas from the SSA algorithm (Schaal and Atkeson, 1994), into a muti-level active learning algorithms, called **Self-Adaptive Goal Generation R-IAC algorithm (SAGG-RIAC)**. Unlike R-IAC that was made for active learning of forward models, we show that this new algorithm allows for efficient learning of inverse models in redundant robots by leveraging the lower-level dimension of the task space. The central idea of **SAGG-RIAC** consists in pushing the robot to perform babbling in the goal/operational space, as opposed to motor babbling in the actuator space, by self-generating goals actively and adaptively in regions of the goal space which provide a maximal competence improvement for reaching those goals. Then, a lower level active motor learning algorithm, inspired by the SSA algorithm (Schaal and Atkeson, 1994), is used to allow the robot to locally explore how to reach a given self-generated goal. Hence, it follows the inspiration of both the SSA algorithm, which constrains the exploration to a tube of data targeted to a specific goal, and of "learning progress" approaches to intrinsic motivation: it explores in an open-ended manner the space of goals, focusing on those where local improvement of the competence to reach them is maximal.

2.2 Global Architecture

Let us consider the definition of competence based models outlined in (Oudeyer and Kaplan, 2008), and extract from it two different levels for active learning defined at different time scales (Fig. 3):

1. The higher level of active learning (higher time scale) considers the *active self-generation and self-selection of goals*, depending on a measure of interestingness based on the level of achievement of previously generated goals.
2. The lower level of active learning (lower time scale) considers the *goal-directed active choice and active exploration* of lower-level actions to be taken to reach the goals selected at the higher level, and depending on another local measure of interestingness based on the evolution of the quality of learnt inverse and/or forward models.

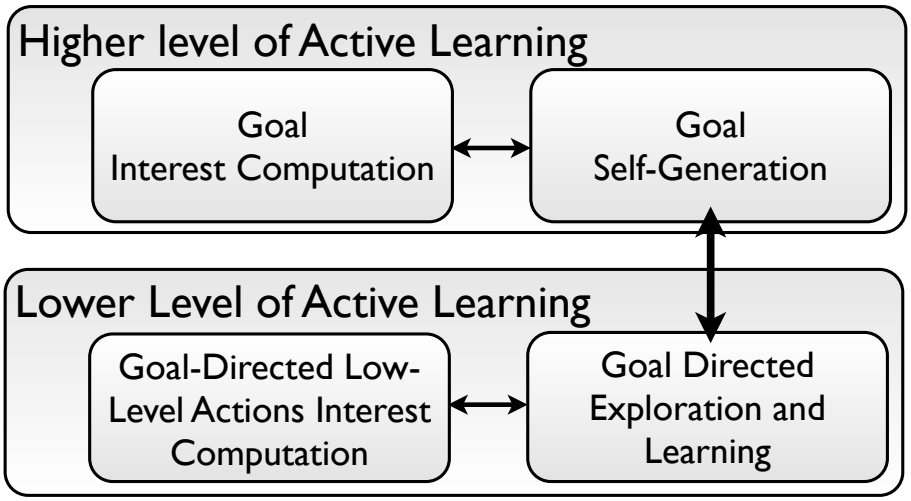

Fig. 3. Global Architecture of the SAGG-RIAC algorithm. The structure is composed of two parts defining two levels of active learning: a higher level which considers the active self-generation and self-selection of goals, and a lower level, which considers the goal-directed active choice and active exploration of lower-level actions, to reach the goals selected at the higher level.

2.2.1 Model Formalization Let us consider a robotic system whose configurations/states are described in both an actuator space S, and an operational/task space S'. For given configurations $(s_1, s'_1) \in S \times S'$, a sequence of actions $a = \{a_1, a_2, ..., a_n\}$ allows a transition toward the new states $(s_2, s'_2) \in S \times S'$ such that $(s_1, s'_1, a) \Rightarrow (s_2, s'_2)$. For instance, in the case of a robotic manipulator, S may represent its actuator/joint space, S' the operational space corresponding to the Cartesian position of its end-effector, and a may be velocity or torque commands in the joints.

In the frame of SAGG-RIAC, we are interested in the reaching of *goals*, from starting states. Also, we formalize starting states as configurations $(s_{start}, s'_{start}) \in S \times S'$ and goals, as a desired $s'_g \in S'$. All states are here considered as potential starting states, therefore, once a goal has been generated, the lower level of active learning always try to reach it by starting from the current state of the system.

When a given goal is set, the low-level process of goal-directed active exploration and learning to reach this goal from the starting state can be seen as exploration and learning of a motor primitive $\Pi_{(s_{start}, s'_{start}, s'_g, \rho, \mathbf{M})}$, parameterized by the initiation position (s_{start}, s'_{start}), the goal s'_g, constraints ρ (e.g. linked with the spent energy), and parameters of already learnt internal forward and inverse models $\mathbf{M}$.

Also, according to the self-generation and self-selection of goals at the higher level, we deduce that the whole process (higher and lower time scales) developed

in SAGG-RIAC can be defined as an autonomous system that explores and learns *fields of parameterized motor primitives*.

2.2.2 Lower Time Scale: Active Goal Directed Exploration and Learning The goal directed exploration and learning mechanism can be carried out in numerous ways. Its main idea is to guide the system toward the goal, by executing low-level actions, which allows it to progressively explore the world and create a model that may be reused afterwards. Its conception has to respect two imperatives :

1. A model (inverse and/or forward) has to be computed during the exploration, and has to be available for a later reuse, in particular when considering other goals.
2. A learning feedback has to be added, such that the exploration is active, and the selection of new actions depends on local measures about the evolution of the quality of the learnt model.

In the experiment introduced in the following, we will use a method inspired by the SSA algorithm introduced by Schaal & Atkeson (Schaal and Atkeson, 1994). Other kinds of techniques, for example based on natural actor-critic architectures in model based reinforcement learning (Peters and Schaal, 2008), could also be used.

2.2.3 Higher Time Scale: Goal Self-Generation and Self-Selection The goal self-generation and self-selection process relies on a feedback defined using a notion of competence, and more precisely on the competence improvement in given subregions of the space where goals are chosen. The following part details the technical formalization of this system.

2.2.4 Measure of Competence A reaching attempt in direction of a goal is defined as terminated according to two conditions:

1. A timeout related to a maximal number of micro-actions/time steps allowed has been exceeded.
2. The goal has effectively been reached.

We introduce a measure of competence $\gamma_{s'_g}$ for a given reaching attempt as depending on a measure of *similarity* C (i.e. typically a distance measure) between the state s'_f reached when the goal reaching attempt has terminated and the actual goal s'_g of this reaching attempt, and the respect of constraints ρ. The measure of similarity C, and thus the measure of competence, is as general as the measure of prediction error could be in RIAC. As seen below, we set equations so that $\gamma_{s'_g}$ is always a negative number, such that the lower the value is, the lower the competence (one can be unboundedly bad, i.e. the distance between the reached configuration and the goal can in general be growing towards

infinity), and the higher the value, the higher the competence (which becomes maximal when the goal is perfectly reached). Thus, we define $\gamma_{s'_g}$ as:

$$\gamma_{s'_g} = \begin{cases} C(s'_g, s'_f, \rho) \text{ if } C(s'_g, s'_f, \rho) \leq \varepsilon_C < 0 \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases}$$

with ε_C a tolerance factor where $C(s'_g, s'_f, \rho) > \varepsilon_C$ corresponds to a goal reached. Thus, a value $\gamma_{s'_g}$ close to 0 represents a system that is competent to reach the goal s'_g respecting constraints ρ. A typical instantiation of C, without constraints, is defined as $C(s'_g, s'_f, \emptyset) = \|s'_g - s'_f\|^2$, which is the direct transposition of prediction error in R-IAC (which here becomes goal reaching error). Yet, other equally general examples of similarity or distance measures could be used, possibly including normalizing terms such as in the experiments below.

Definition of Local Competence Progress The active goal self-generation and self-selection relies on a feedback linked with the notion of competence introduced above, and more precisely on the monitoring of the progress of local competences. We firstly define this notion of local competence: let us consider different measures of competence $\gamma_{s'_i}$ computed for reaching attempts to different goals $s'_i \in S', i > 1$. For a subspace called a region $R \subset S'$, we can compute a measure of competence γ_R that we call a *local measure* such that:

$$\gamma_R = \left(\frac{\sum_{s'_j \in R} (\gamma_{s'_j})}{|R|} \right) \tag{1}$$

with $|R|$, cardinal of R.

Let us now consider different regions R_i of S' such that $R_i \subset S', \bigcup_i R_i = S'$. (initially, there is only one region that is then progressively and recursively split, see below and figure 4). Each R_i contains attempted goals $\{s'_{t_1}, s'_{t_2}, ..., s'_{t_k}\}_{R_i}$, and corresponding competences obtained $\{\gamma_{s'_{t_1}}, \gamma_{s'_{t_2}}, ..., \gamma_{s'_{t_k}}\}_{R_i}$, indexed by their relative time order $t_1 < t_2 < ... < t_k | t_{n+1} = t_n + 1$ of experimentation inside this precise subspace R_i (t_i are not the absolute time, but integer indexes of relative order in the given subspace (region) being considered for goal selection). The interest value, described by equation 2, represents *the absolute value of the derivative of the local competence value inside R_i, hence the local competence progress, over a sliding time window of the ζ more recent goals attempted inside R_i* (see Baranes and Oudeyer, 2011 for a justification of the absolute value):

$$interest(R_i) = \frac{\left| \left(\sum_{j=|R_i|-\zeta}^{|R_i|-\frac{\zeta}{2}} \gamma_{s'_j} \right) - \left(\sum_{j=|R_i|-\frac{\zeta}{2}}^{|R_i|} \gamma_{s'_j} \right) \right|}{\zeta} \tag{2}$$

Goal Self-Generation Using the Measure of Interest Using the previous description of interest, the goal self-generation and self-selection mechanism has to carry out two different processes (see figure 4):

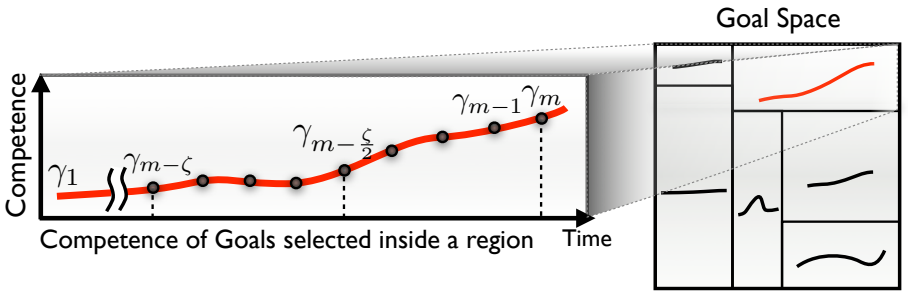

Fig. 4. Illustration of how a goal space can be split into subregions, in each of which competence progress is monitored. The action-selection system decides most of the time (typically 70 percent) to explore goals with regions of highest learning progress (the probability of choosing a region is proportional to competence progress), but still for meta-exploration dedicates a part of its time (typically 30 percent) to explore other randomly chosen regions.

1. Split of the space S' where goals are chosen, into subspaces, according to heuristics that allows to maximally distinguish areas according to their levels of interest;
2. Select the subspaces where future goals will be chosen;

Such a mechanism has been described in the Robust-Intelligent Adaptive Curiosity (R-IAC) algorithm introduced in (Baranes and Oudeyer, 2009), but was previously applied to the actuator space S rather than to the goal/task space S' as we do in SAGG-RIAC. Here, we use the same kind of methods like a recursive split of the space, each split being triggered once a maximal number of goals g_{max} has been attempted inside. Each split is performed such that is maximizes the difference of the interest measure described above, in the two resulting subspaces, this allows to easily separate areas of different interest, and thus, of different reaching difficulty.

Finally, goals are chosen according to the following heuristics that mixes three modes, and once at least two regions exist after an initial random exploration of the whole space:

1. *mode*(1): in $p_1\%$ percent (typically $p_1 = 70\%$) of goal selections, the algorithm chooses a random goal inside a region chosen with a probability proportional to its interest value:

$$P_n = \frac{|interest_n - \mathbf{min}(interest_i)|}{\sum_{i=1}^{|R_n|} |interest_i - \mathbf{min}(interest_i)|} \tag{3}$$

Where P_n is the probability of selection of the region R_n, and $interest_i$ corresponds to the current *interest* of regions R_i.

2. $mode(2)$: in $p_2\%$ of cases (typically $p_2 = 20\%$), the algorithm selects a random goal inside the whole space.
3. $mode(3)$: in $p_3\%$ (typically $p_3 = 10\%$), it performs a random experiment inside the region where the mean competence level is the lowest.

Developmental Constraints for the Reduction of the Initiation Set: to improve the quality of the learnt inverse model, we add a heuristic inspired by observations of Berthier et al. (Berthier et al., 1999) who noticed that infant's reaching attempts were often preceded by movements that either elevated their hand or moved their hand back to their side. By analogy, using such heuristics can directly allow a highly-redundant robotic system to reduce the space of initiation states used to learn to reach goals, and also typically prevent it from experimenting with too complex actuator configurations. Also, we add it in SAGG-RIAC, by specifying a rest position (s_{rest}, s'_{rest}) settable without any need of planning from the system, that is set for each $r \in n\mathbb{Z}$ subsequent reaching attempts.

2.3 Experimenting SAGG-RIAC on a reaching task

In the following, we consider a n-dimensions manipulator controlled in position and speed (as many of today's robots), updated at discrete time values, called *time steps*. The vector $\theta \in \mathbb{R}^n = S$ represents joint angles, and $x \in \mathbb{R}^m = S'$, the position of the manipulator's end-effector in m dimensions, in the Euclidian space S' (see Fig. 5 where $n = 15$ and $m = 2$). We evaluate how the SAGG-RIAC algorithm can be used by a robot to learn how to reach all reachable points in the environment S' with this arm's end-effector. Learning both the forward and inverse kinematics is here an online process that arises each time a micro-action is executed by the manipulator: by doing movements, the robot stores measures $(\theta, \Delta\theta, \Delta x)$ in its memory; these measures are then reused online to compute the Jacobian $J(\theta) = \Delta x/\Delta\theta$ locally, and the associated Moore-Penrose pseudo-inverse, to move the end-effector in a desired direction $\Delta x_{desired}$ fixed towards the self-generated goal. Also, in this experiment, where we suppose S' euclidian, and do not consider obstacles, the direction to a goal can be defined as following a straight line between the current end-effector's position and the goal (thus we avoid using complex planning, which is a separate problem and thus allows us to interpret more easily the results of the experiments).

2.3.1 Evaluation of Competence In this experiment, we do not consider constraints ρ and only focus on the reaching of goal positions x_g. We define the cost function C and thus the competence as linked with the Euclidian distance $D(x_g, x_f)$, between the goal position and the final reached position x_f, which is normalized by the starting distance $D(x_{start}, x_g)$, where x_{start} is the end-effector's starting position:

$$C(x_g, x_f, x_{start}) = -\frac{D(x_g, x_f)}{D(x_{start}, x_g)} \tag{4}$$

where $C(x_g, x_f, x_{start}) = min_C$ if $D(x_{start}, x_g) = 0$ and $D(x_g, x_f) \neq 0$.

2.3.2 Addition of subgoals Computing local competence progress in subspaces/regions typically requires the reaching of numerous goals. Because reaching a goal can necessitate several actions, and thus time, obtaining competence measures can be long. Also, without biasing the learning process, we improve this mechanism by taking advantage of the Euclidian aspect of S': we increase the number of goals artificially, by adding subgoals on the pathway between the starting position and the goal, where competences are computed. Therefore, considering a starting state x_{start} in S', and a self-generated goal x_g, we define the set of l subgoals $\{x_1, x_2, ..., x_l\}$ where $x_i = (i/l) \times (x_g - x_{start})$, that have to be reached before attempting to reach the terminal goal x_g.

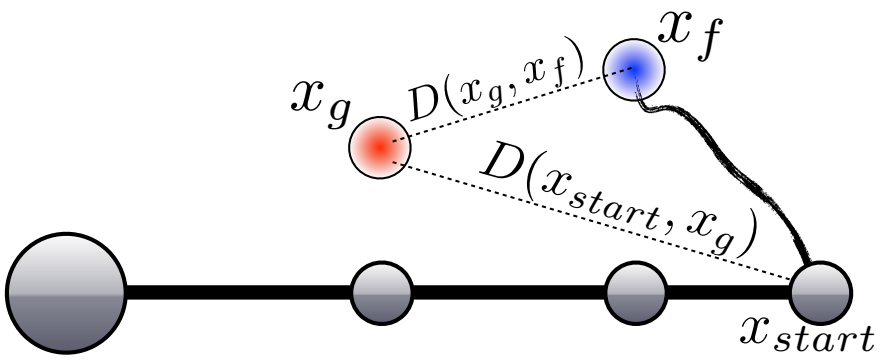

Fig. 5. Values used to compute the competence $\gamma_{s'_g}$, considering a manipulator of 3 degrees-of-freedom, in a 2 dimensions operational space. Here, the arm is set in a position called *rest position* $(\theta_{rest}, x_{rest})$.

2.3.3 Local Exploration and Reaching Here we propose a method, inspired by the SSA algorithm (Schaal and Atkeson, 1994), to guide the system to learn on the pathway toward the selected goal position x_g. The system is organized around two alternating phases: *reaching* phases, which involve a local controller to drive the system from the current position x_c towards the goal, and *local exploration* phases, which allows to learn the inverse model of the system in the close vicinity of the current state, and are triggered when the reliability of the local controller is too low. These mechanisms are stopped once the goal has been reached or a timeout exceeded. Let us here describe the precise functioning of those phases in our experiment:

Reaching Phase: the reaching phase deals with creating a pathway to the goal position x_g. This phase consists of determining, from the current position

x_c, an optimal movement to guide the end-effector toward x_g. For this purpose, the system computes the needed end-effector's displacement $\Delta x_{next} = v.\frac{x_c - x_g}{\|x_c - x_g\|}$ (where v is the velocity bounded by v_{max} and $\frac{x_c - x_g}{\|x_c - x_g\|}$ a normalized vector in direction of the goal), and performs the action $\Delta\theta_{next} = J^+.\Delta x_{next}$, with J^+, pseudo-inverse of the Jacobian estimated in the close vicinity of θ and given the data collected by the robot so far. After each action Δx_{next}, we compute the error $\varepsilon = \|\widetilde{\Delta x}_{next} - \Delta x_{next}\|$, and trigger the exploration phase in cases of a too high value $\varepsilon > \varepsilon_{max} > 0$.

Exploration Phase: this phase consists in performing $q \in \mathbb{N}$ small random explorative actions $\Delta\theta_i$, around the current position θ. This allows the learning system to learn the relationship $(\theta, \Delta\theta) \Rightarrow \Delta x$, in the close vicinity of θ, which is needed to compute the inverse kinematics model around θ.

2.4 Results

2.4.1 Goal Directed Exploration and Learning In the experiment introduced in this section, we consider the robotic system presented above with a n-DOF arm on a plane, with $n = 7$, 15, or 30 (thus the problem has respectively 16, 32 and 62 continuous dimensions, if one considers the fact that the motor space is spanned by the position and speed of each joint, and the task space has 2 dimensions) . We set the dimensions of S' as bounded in intervals $x_g \in [0; 150] \times [-150; 150]$, where 50 units is the total length of the arm, which means that the arm covers less than 1/18 of the space S' where goals can be chosen (i.e. the majority of areas in the operational/task space are not reachable, which has to be discovered by the robot). We fix the number of subgoal per goal to 5, and the maximal number of elements inside a region before a split to 50. We also set the desired velocity $v = 0.5$ units/movement, and the number of explorative actions $q = 20$. Moreover, we reset the arm to the rest position $(\theta_{rest}, x_{rest})$, where the arm is straight (position displayed in Fig. 5), every $r = 2$ reaching attempts. This allows the system to reduce the initiation set, and avoid experimenting with too complex joint positions where the arm is folded, and where the Jacobian is more difficult to compute.

2.4.2 Qualitative results Fig. 6 shows histograms of the self-generated goal positions (goals without subgoals) with $n = 15$, and created regions, after the execution of 200000 time steps (i.e. micro-actions). Each subfigure represents data obtained during a time window indexed on the number of generated goals: the first one (upper-left) shows that in the very beginning of learning (20 goals corresponds to 100 goals and subgoals), the system is already splitting the space and seems to discriminate the left third of the space, where the reachable area is (contoured by the black half-circle on each subfigure). Upper-right and lower-left subfigures show examples of areas where goals are generated over time, we can observe that the highest amount of goals that are chosen remains inside the reachable area: the system is indeed discovering that only a subpart is reachable, the interest value becoming null in totally unreachable areas where the

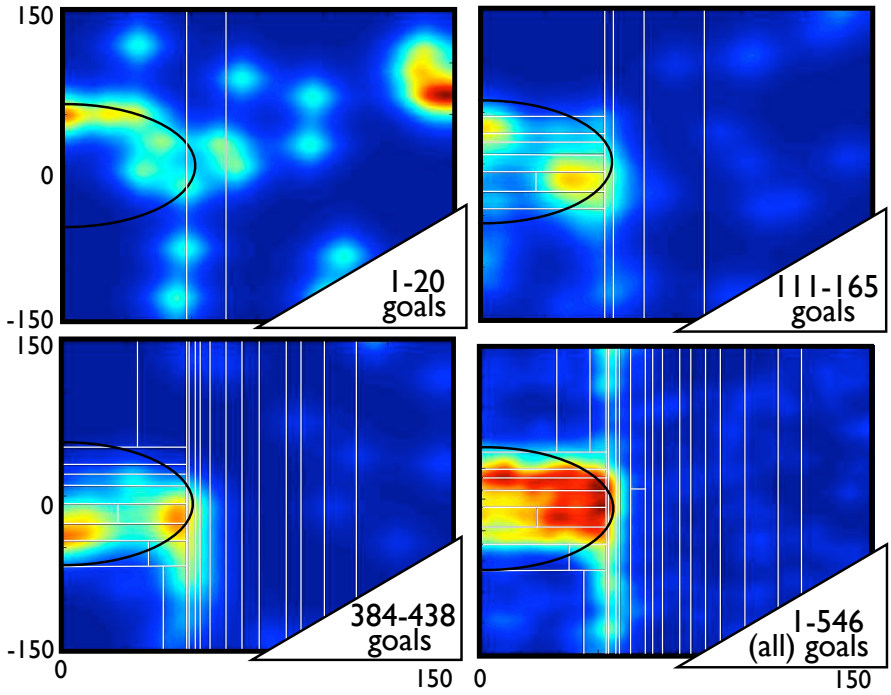

Fig. 6. Histograms of self-generated goals and regions with a 15 DOF robotic planar arm (split by white lines) displayed over time windows indexed by the number of performed goals, for an experiment of 200000 time steps (i.e. micro-actions). The black half-circle represents the contour of the area reachable by the arm according to its length of 50 units

competence typically takes small values, or even reach the threshold min_C. The last subfigure (lower-right) represents the position of all goals that have been self-generated and allows to observe that SAGG-RIAC is able to highly discriminate unreachable areas over time, and to focus its goal self-generation in the whole reachable subspace. Finally, observing regions, we can globally notice that the system splits the reachable space into regions in the first quarter of goal generations (upper-right subfigure), and then continues to split the space inside unreachable regions, in order to potentially release new areas of interest.

It is also important to notice that coupling the lower-level of active learning inspired by SSA with the heuristic of returning to x_{rest} every two subsequent goals creates an increasing radius of known data around x_{rest}, inside the reachable space. Indeed, the necessity to be confident in the local model of the arm to shift toward new positions makes the system progressively explore the space, and resetting it to its rest position makes it progressively explore the space by beginning close to x_{rest}. Finally, goal positions that are physically reachable but

far from this radius typically present a low competence to be reached initially, before the radius spreads enough to reach them, which creates new areas of interest, and explains the focalization on reachable areas far from x_{rest} (see Fig. 6). Therefore, the exploration also proceeds by going through reachable subspaces of growing complexity of reachability.

2.4.3 Quantitative Results In the following evaluation, we consider the same robotic system as previously and design two experiments. This experiment considers a large space $S' = [0; 150] \times [-150; 150]$, where one can evaluate the capability of SAGG-RIAC to discriminate and avoid to explore unreachable areas. Here, we repeated the experiment with $n = 7, 15, 30$ dofs as well as by taking two geometries for the arm: one where all segments have equal length, and one where the length is decreasing with the golden number ratio. All configurations of experiments were repeated 15 times in order to obtain proper statistics. We compare the following exploration techniques:

1. SAGG-RIAC
2. SAGG-Random, where goals are chosen randomly (higher-level of active learning (RIAC) disabled)
3. ACTUATOR-Random, where small random movements $\Delta\theta$ are executed.
4. ACTUATOR-RIAC, which corresponds to the original RIAC algorithm, which uses the decrease of the prediction error $(\theta, \Delta\theta) \rightarrow \Delta x$ to compute an interest value and split the space $(\theta, \Delta\theta)$.

Also, to be comparable to SAGG-RIAC, for each other techniques we reset the position of the arm to the *rest position* every *max* time steps, *max* being the number of time steps needed to consecutively reach the two more distant reachable positions. Fig. 7 shows the evolution of the capability of the system to reach 100 test goals (independently and uniformly distributed in the reachable area) using the inverse model learnt by each technique, starting from, half the time, the rest positions. The observations of the curves show several things. First, SAGG-RIAC allows both the robot to learn faster and to reach the highest absolute level of performance in *all* cases, and an ANOVA analysis shows a level of significance of this result $p = 0.002$ at the end of the experiment for 15 dofs. Thus, like in the previous experiment, we see that SAGG-RIAC allows both larger speed and higher generalization ability. The second observation that we can make is that random exploration in the joint space (ACTUATOR RANDOM) is the second best method for 7 dofs, then third for 15 dofs, then last for 30 dofs, illustrating the curse-of-dimensionality. Inversely, the evolution of performances of SAGG-Random and SAGG-RIAC degrades gracefully as dimension grow, showing how exploration in the operational space allows to harness the curse-of-dimensionality by exploiting the redundancy of the robotic system and the low-dimensionality of the operational space. Yet, one sees that SAGG-RIAC is consistently and significantly more efficient than SAGG-Random, which is explained by the fact that SAGG-Random pushes too often the robot to try to reach unreachable goals while SAGG-RIAC is capable of quickly focusing in

reachable goals. Finally, we again see that ACTUATOR-RIAC is not robust to the increase of dimensionality in such continuous spaces.

More systematic studies should of course be done, but these results already indicate the high potential of **competence based motor learning** in general, even using random Goal Self-Generation.

3 Maturationally constrained intrinsically motivated exploration

Task-level exploration by competence-based intrinsic motivation architectures can considerably decrease the dimensionality of the spaces to be explored and leverage the potentially high redundancy of sensorimotor spaces for more efficient exploration and acquisition of new skills. Yet, task spaces may often be unbounded or characterized by a volume of the "interesting" regions much smaller than the overall volume of the sensorimotor space, even when there are very few dimensions, and in those cases task-level exploration and competence-based architectures become as inefficient as knowledge-based architectures. For example, the three dimensional space around a human is unbounded, and both for reaching or locomotion tasks the space of potential goals to be reached is thus unbounded. As said above, the human or the robot does not know its limits initially, so why shall he not try to reach with its hand the mountains five kilometers ahead in its visual field, or why shall he not try to run at one hundred kilometers per hour? In the reaching experiment presented in previous section, the task space was in fact artificially and rather arbitrarily bounded: while including many parts that were not reachable, it was still severely bounded, which allowed goal exploration to succeed. As argued at the beginning of this article, mechanisms for self-bounding the explorable space are necessary, but they should be less ad hoc. To reach this objective, one may take inspiration from maturational mechanisms in biological organisms [4].

The progressive biological maturation of infant's brain, motor and sensor capabilities, including changes in morphological properties of the body, introduces numerous important constraints on the learning process (Schlesinger, 2008). Indeed, at birth, all the sensorimotor apparatus is neither precise enough, nor fast enough, to allow infants to perform complex tasks. The low visual acuity of infants (Turkewitz and Kenny, 1985), their incapacity to efficiently control distal muscles, and to detect high-frequency sounds, are examples of constraints reducing the complexity and limiting the access to the high-dimensional and unbounded space where they evolve (Bjorklund, 1997). Maturational constraints play an important role in learning, by partially determining a developmental pathway. Numerous biological reasons are part of this process, like the brain maturation, the weakness of infants' muscles, or the development of the physiological sensory system. In the following, we focus on constraints induced by

[4] Part of the material presented in this section is adapted from (Baranes and Oudeyer, 2011)

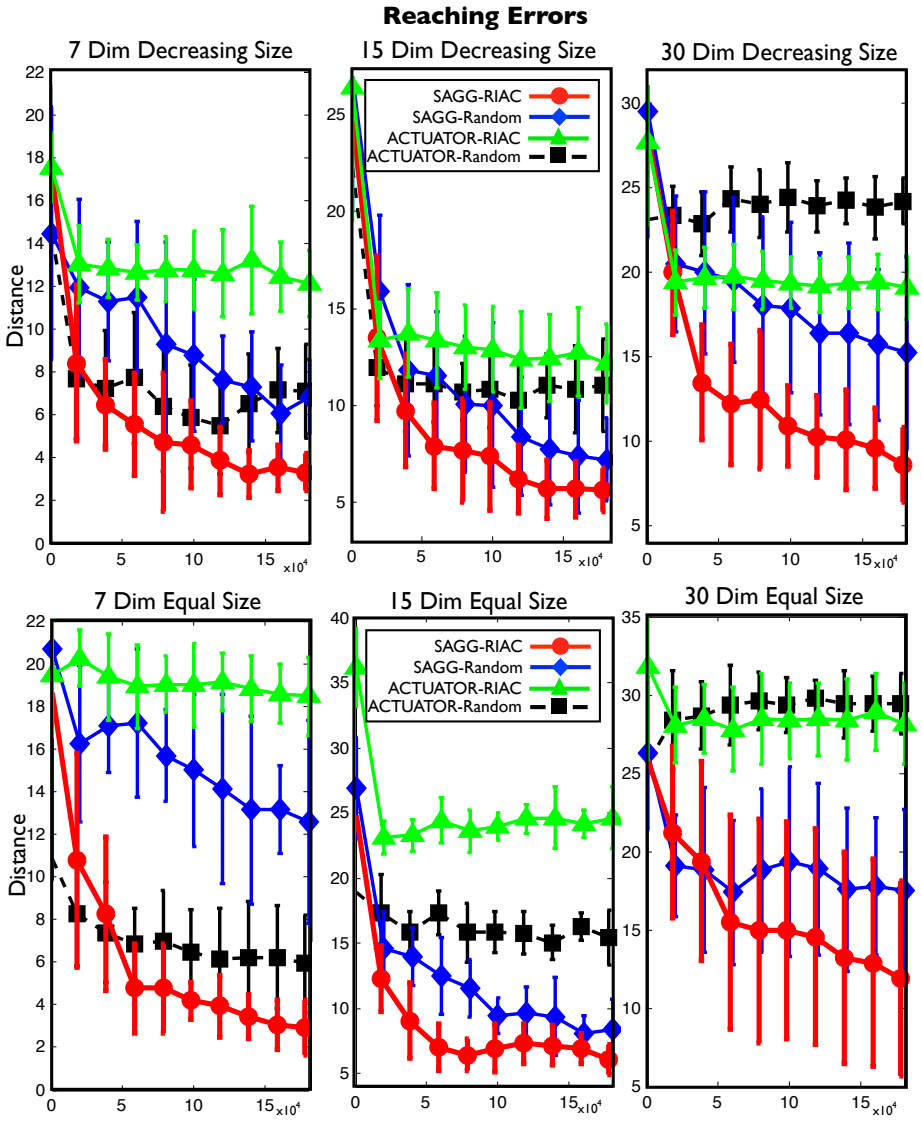

Fig. 7. Evolution of mean distances goal-end effector (reaching errors) after reaching attempts over an independently randomly generated set of test goals. Here SAGG-RIAC and SAGG-random are only allowed to choose goals within $S' = [0; 150] \times [-150; 150]$ (i.e. the set of reachable goals is only a small subset of eligible goals).

the brain **myelination** (Eyre, 2003). Related to the evolution of a substance called myelin, and usually qualified by the term white matter, the main impact

of myelination is to help the information transfer in the brain by increasing the speed at which impulses propagate along axons (connections between neurons). Here, we focus on the myelination process for several reasons, this phenomenon being responsible for numerous maturational constraints, effecting the motor development, but also the visual or auditory acuity, by making the number of degrees-of-freedom, and the resolution of sensory-motor channels increase progressively with time.

Actually, infants' brain does not come with an important quantity of white matter, myelination being predominantly a postnatal process, taking place in a large part during the first years of life. Konczak (Konczak et al., 1997) and Berthier (Berthier et al., 1999) studied mechanisms involved in reaching trials in human infants. In their researches, they expose that goal-directed reaching movements are ataxic in a first time, and become more precise with time and training. Also, they show that for all infants, the improving efficiency of control follows a proximo-distal way, which means that infants use in priority their torso and shoulder for reaching movements, and, progressively, their elbow (Berthier et al., 1999), see figure 8. This evolution of control capabilities comes from the increasing frequency of the muscular impulses, gradually, in shoulders, and elbows. This phenomenon, directly related to the myelination of the motor cortex, then allows muscles to become stronger at the same time, by training, which then increases their possibility to experiment wider sets of positions. Myelin is also responsible for brain responses to high visual and sound frequencies. Therefore, like introduced in (Turkewitz and Kenny, 1985), children are not able to detect details in images, which is also a reason, of imprecise reaching movements.

Coupling maturation and intrinsic motivation. Maturational mechanisms could easily be integrated with an intrinsic motivation system, where the space explorable with intrinsic motivation would grow as new degrees of freedom and higher resolutions or ranges are released and timed by a maturational clock. If the maturational clock is purely dependent on physical time, then maturation influences intrinsically motivated exploration. But what may be potentially even more useful is that maturation could in return be accelerated or slowed down based on how fast (or slow) new competences are acquired by the intrinsically motivated organism. This means that intrinsically motivated exploration would not only be a mechanism for deciding "what to learn", but also a self-regulating mechanism indirectly regulating the growth of the complexity of the very space in which exploration takes place. If one imagines that maturation not only releases new degrees of freedoms in sensorimotor channels, but also capabilities of statistical inference mechanisms (e.g. increasing the size of the hypothesis space to search), then one would have an intrinsic motivation system which actively explores its environment and at the same time actively regulates both the bounds of this explorable space and the capabilities of the learning system that it is driving. Such an integration of maturation and intrinsic motivation is what has been explored in the McSAGG architecture, coupling SAGG-RIAC with maturational constraints (Baranes and Oudeyer, 2010b). Such an approach is also supported by theories of brain and behaviour development that highlight

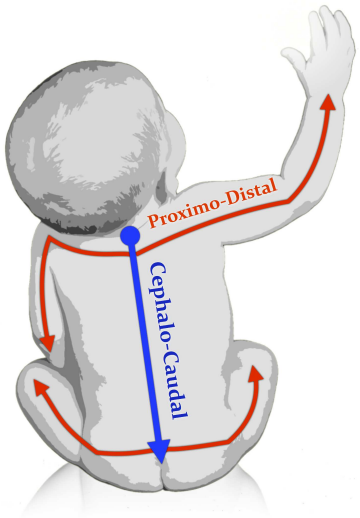

Fig. 8. The proximo-distal and cephalo-caudal law: infants explore and learn in priority their torso and shoulder for reaching movements, and progressively their elbow, and the same process happens in the gradual exploration and mastery of the neck-feet axis.

the strong interaction between maturational processes and learning processes (Johnson, 2001). The following section presents an outline of the system as well as results that show how it can improve the efficiency of intrinsically motivated learning on the same reaching task than in previous section.

3.1 McSAGG: Maturationally constrained competence based architecture

It is important to notice the multi-level aspect of maturational constraints: maturation can apply to motor actions/commands in the joint/control space as well as to goals in task spaces; also, maturational constraints can apply to sensors, such as the capacity to discriminate objects, and so here, to declare a goal as reached. The global idea is to control all of these constraints using an evolving term $\psi(t)$, called **adaptive maturational clock**, which increase, influencing the lifting of constraints, depends on the global learning evolution, and is typically non-linear. The main problem raised is to define a measure allowing the robot learner to control the evolution of this clock. For instance, in the Lift-Constraint, Act, Saturate (LCAS) algorithm, (Lee et al., 2007) use a simple discrete criteria based on a saturation threshold. (Lee et al., 2007) consider a robotic arm whose the end-effector's position is observed in a task space. This task space is segmented into spherical regions of specified radius used as output for learning the forward kinematics of the robot. Each time the end-effector

explores inside a region, this one is activated. Once every region is activated, saturation happens, and the radius of each region decreases so that the task space becomes segmented with a higher resolution, and allows a more precise learning of the kinematics. In the following section, we take inspiration from the LCAS algorithm and define a measure based on the competence progress allowing us to control a continuous and non linear evolution of the maturational clock.

3.1.1 Stage Transition: Maturational Evolution and Intrinsic Motivations Often considered as a process strictly happening in the first years of life, myelin continues to be produced even in adults while learning new complex activities (Scholz et al., 2009). Also, in a developmental robotics frame, we set the maturational clock $\psi(t)$ which controls the evolution of each release of constraint, as depending on the learning activity, and especially on the progress in learning by itself. Here, the main idea is to increase $\psi(t)$ (lifting constraints), *when the system is in a phase of stabilization of its global competence level, after a phase of progression* (see Fig. 9). This stabilization is shown by a low derivative of the averaged competence level computed in the whole goal space S' in a recent time window $[t_{n-\frac{\zeta}{2}}, t_n]$, and the progression, by an increase of these levels in a preceding time window $[t_{n-\zeta}, t_{n-\frac{\zeta}{2}}]$. We thus use a description analogous to the notion of competence progress used to define our measure of interest. Therefore, considering competence values estimated for the ζ last reaching attempts $\{\gamma_{s'_{n-\zeta}}, ..., \gamma_{s'_n}\}_{S'}$, $\psi(t)$ evolves until reaching a threshold ψ_{max} such that:

$$\psi(t+1) = \psi(t) + min\left(max_{evol}; \frac{\lambda}{CP(\{\gamma_{s'_{n-\zeta/2}}, ..., \gamma_{s'_n}\})}\right)$$

$$\text{if} \begin{cases} 0 < CP(\{\gamma_{s'_{n-\zeta/2}}, ..., \gamma_{s'_n}\}) < max_{CP} and \\ CP(\{\gamma_{s'_{n-\zeta}}, ..., \gamma_{s'_{n-\zeta/2}}\}) > 0 and \psi(t) < \psi_{max} \end{cases}$$

and $\psi(t+1) = \psi(t)$ otherwise, where max_{evol} is a threshold limiting a too rapid evolution of ψ, max_{CP} a threshold defining a stable competence progress, λ a positive factor, and $CP(\{\gamma_{s'_{n-\zeta/2}}, ..., \gamma_{s'_n}\})$ a measure of competence progress (in the experiments presented in this section, no absolute value is used, i.e. intrinsic rewards are only provided for increases in competence). As the global interest of the whole space is typically non-stationary, the maturational clock becomes typically non-linear, and stops its progression when the global average of competence decreases, due to the lifting of previous constraints. In Fig. 9, the increase of $\psi(t)$ is denoted as evolution periods.

3.1.2 Constraints Modeling

Constraints over the control space In this model, we concentrate on three kinds of maturational constraints over the control and perceptual space, i.e. constraints on motor programs explorable to reach goals and over perception used to achieve

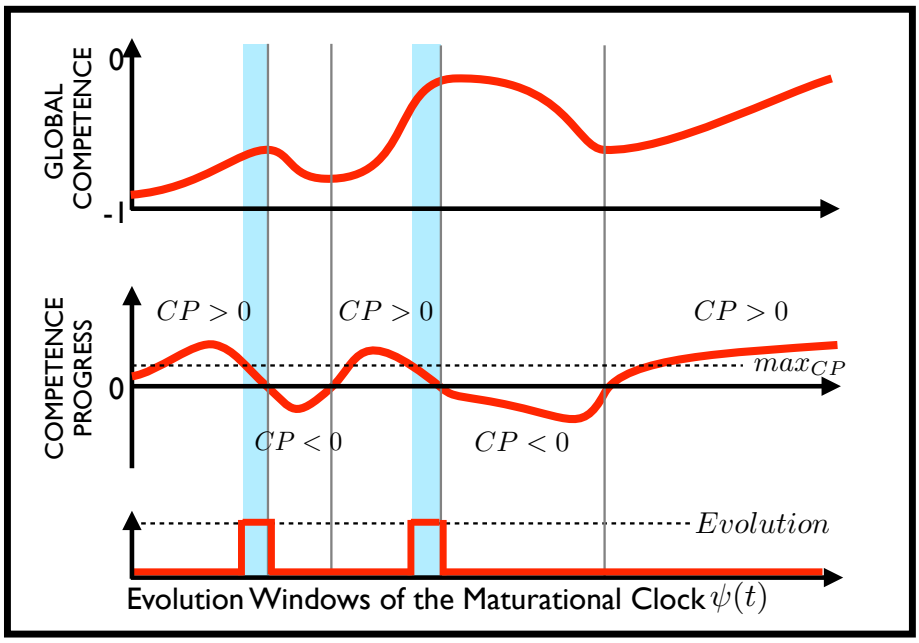

Fig. 9. Values used to compute the competence $\gamma_{s'_g}$, considering a manipulator of 3 degrees-of-freedom, in a 2 dimensions operational space.

those programs or evaluate whether goals are reached or not, directly inspired by consequences of the myelination process, and which are controlled by $\psi(t)$. These constraints are general and can be integrated in numerous kinds of robots.

The first constraint describes the *limitation of frequency of muscular impulses*, applied to the control of the limbs, which is responsible of the precision and complexity of control (Konczak et al., 1997). Also corresponding to the frequency of feedback updating movements to achieve a trajectory, we define the constraint $f(t)$ as increasing with the evolution of the maturational clock:

$$f(t) = \left(-\frac{(p_{max} - p_{min})}{\psi_{max}} . \psi(t) + p_{max} \right)^{-1} \tag{5}$$

Where p_{max} and p_{min} represents maximal and minimal possible time periods between control impulses.

The second studied constraint relies on the sensor abilities. Here, we consider the *capacity to discriminate objects* as evolving over time, which here corresponds to an evolving value of ε_D, the tolerance factor allowing to decide of a goal as reached. We thus set ε_D as evolving, and more precisely, decreasing over the maturational clock, from $\varepsilon_{D_{max}}$ to $\varepsilon_{D_{min}}$:

$$\varepsilon_D(t) = -\frac{(\varepsilon_{D_{max}} - \varepsilon_{D_{min}})}{\psi_{max}}.\psi(t) + \varepsilon_{D_{max}} \quad (6)$$

Finally, we set another constraint, implementing a mechanism analogous to the proximo-distal law described above. Here, we consider the ranges r_i within which motor commands in the control space can be chosen, as increasing over maturational time following a proximo-distal way over the structure of the studied embodied system. This typically allows larger movements and further goals to become explorable, and thus learnable:

$$r_i(t) = \ min(\psi(t).k_i, rmax_i) \quad (7)$$

Where k_i represents an intrinsic value determining the difference of evolution velocities between each joint. In the case of a robotic arm such as in the reaching task, and if one denotes $i = 1, ..., l$ the joints in the control space, then the proximo-distal law can be implemented by choosing $k_1 \geq k_2 \geq ... \geq k_n$, where k_1. In the quantitative experiments below, we only use this later constraint.

Constraints over the goal space As explained above, evolving maturational constraints can also be set on the space of explorable goals, in which active learning algorithm such as SAGG-RIAC can learn to discover reachable and unreachable areas. A simple but powerful (as shown below) manner to model those constraints is to let the robot start from a relatively small volume goal space around one or several seeds, and then have the goal space grow as a sphere which radius R_{goal} increases with the maturational clock:

$$R_{goal} = \ \psi(t).G_{const} \quad (8)$$

3.1.3 Experiments on maturationally constrained learning of arm kinematics Here, we consider a simulated robotic arm with the same reaching task than in the previous section on task-level exploration.

In a first qualitative experiment, we consider the case of a n=3 DOF arm, put in a two dimensional environment. We set the arm with a global length of 50 units, and fix the proportion of each limb as 3/5, 2/5, and 1/5 of this length and fix $\psi_{max} = 10$. Fig. 10 (a) shows the different constraints $r_i(t)$, ε_D and $f^{-1}(t)$ over values that take the maturational clock $\psi(t)$. We can firstly observe increasing ranges $r_i(t)$, defined such that $r_3(t) < r_2(t) < r_1(t)$, which respects the proximo-distal constraint meaning that joints closer to the basis of the arm have a controllable range which increase faster than further joints. Fig. 10 (a) also shows the evolutions of $\varepsilon_D(t)$, from 5 to 1 units over $\psi(t)$, and $f^{-1}(t)$, representative of the time period between the manipulator's update control signals, from 3 to 1 time steps. The evolution of the frequency has been decided as being not continuous, to let us observe the behavior of the algorithm when a sudden change of complexity arises for a constraint. We run an experiment over 15000

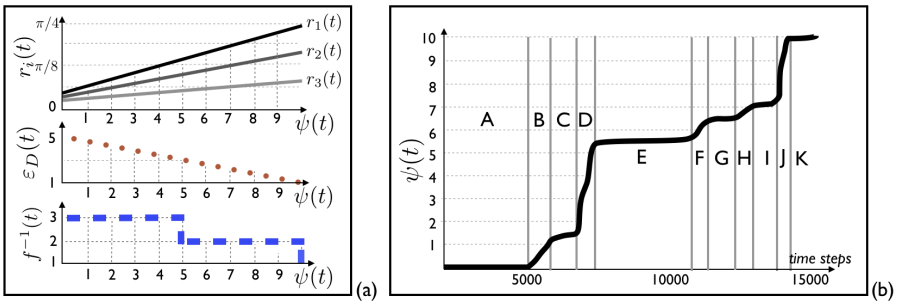

Fig. 10. (a) Exploration of maturational constraints over values taken by the maturational clock $\psi(t)$, for a manipulator of 3-dof. (b) evolution of the maturational clock over time, for a given experiment. Vertical splits are added manually, to let appear what we call *maturational stages*, which are described as periods between important changes of the evolution of $\psi(t)$ (change of the second derivative of $\psi(t)$).

time steps, which corresponds to the selection of about 7500 goals. During the exploration, we observe the evolution of the maturational clock $\psi(t)$ over time (black curve in Fig. 10 (b)) which evolves non-linearly, depending on the global progress of competence. Letters from A to K are added from an external point of view, they are described as periods between important changes of the evolution of $\psi(t)$ (evolution of the second derivative of $\psi(t)$) and represent what we call *maturational stages*. We describe two types of stages, *stationary stages* like A, C, E, G, I, K, where the maturational clock evolves slowly, which corresponds to time period (over time steps) where the global competence progress is either stable or negative, and *evolution stages*, like B, D, F, H, J, where the maturational clock is evolving with a high velocity.

We can emphasize two important maturational stages : the first one, A, which corresponds to a non-evolution of $\psi(t)$; this is due to the need of the mechanism to obtain a minimal number of competence measures, before computing the global progress to decide of a release of constraints. Also, the stable stage E, which appears after that $\psi(t)$ reaches the value 5 can be explained by the sudden change of frequency $f(t)$ from $1/3$ to $1/2$ update per time step, that is produced precisely at $\psi(t) = 5$. This is an effective example that clearly shows the capability of the McSAGG algorithm to slow down the evolution of the maturational clock in cases of an important change of complexity of the accessible body and world, according to constraints.

Another experiment can be made to assess the quantitative gain that can be obtained by using maturational constraints in terms of acquired competence to reach goals spread in the reachable space. In this experiment, we use $n = 15$ degrees of freedom in the robotic arm. Figure 11 shows the evolution of competences to reach a set of goals uniformly spread over the reachable space and chosen independently from the exploration process. We observe that using

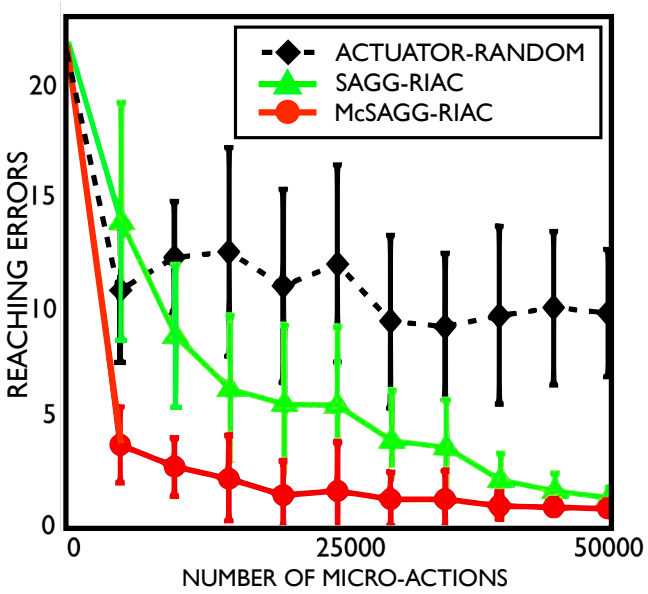

Fig. 11. Comparison of the evolution of acquired reaching competences among various exploration architectures: (1) random exploration in the joint space (Actuator-random), (2) SAGG-RIAC and (3) McSAGG-RIAC (i.e. SAGG-RIAC with maturational constraints). The y axis represents the mean reaching errors over 100 goals generated uniformly and independently from the exploration processes.

maturational constraints still improves importantly SAGG-RIAC, which was itself already shown to improve other active learning strategies as explained in the previous section.

3.1.4 Maturationally constrained learning of quadruped locomotion: coupling maturational constraints, motor synergies and intrinsic motivation The following experiment gives an example of how the combination of three families of constraints presented so far, i.e. intrinsically motivated exploration in the task space, motor synergies and maturational constraints, can leverage each other in order to allow a robot to learn a field of motor skills in a high-dimensional complex motor space: learning omnidirectional quadruped locomotion (see figure 12).

Robotic Setup In the following experiment, we consider a quadruped robot. Each of its leg is composed of two joints, the first one (the closest to the robot's body) is controlled by two rotational DOF, and the second by one rotation (one DOF). Each leg therefore consists of 3 DOF, the robot having in its totality 12 DOF (see figure 12).

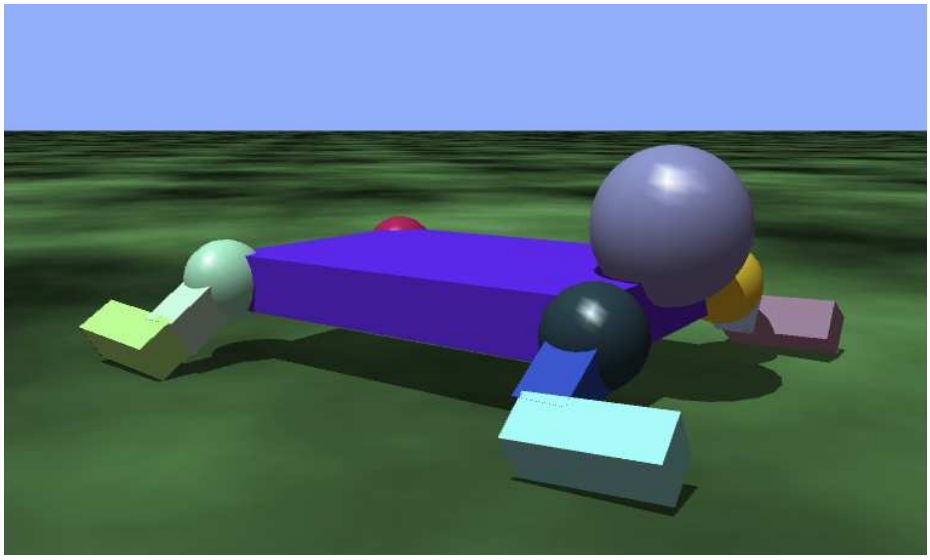

Fig. 12. The simulated quadruped. Physics is simulated using ODE and the Breve simulator (http://www.spiderland.org/)

This robot is controlled using motor synergies Υ which directly specify the phase and amplitude of sinusoids which control the precise rotational value of each DOF trajectory over time (the synergies define target trajectories which are then dynamically tracked by a closed loop low-level PID controller). These synergies are parameterized using a set of 24 continuous values, 12 representing the phase ph of each joint, and the 12 others, the amplitude am: $\Upsilon = \{ph_{1,2,..,12}; am_{1,2,..,12}\}$. Each experimentation consists of launching a motor synergy Υ for a fixed amount of time, starting from a fixed position. After this time period, the resulting position x_f of the robot is extracted into 3 dimensions: its position (u, v), and its rotation θ. The correspondence $\Upsilon \rightarrow (u, v, \theta)$ is then kept in memory as a learning exemplar.

The three dimensions u, v, θ are used to define the goal space of the robot. Also, it is important to notice that precise areas reachable by the quadruped cannot be estimated beforehand. In the following, we set the original dimensions of the goal space to $[-45; 45] \times [-45; 45] \times [-2\pi; 2\pi]$ on axis (u, v, θ), which was a priori larger than the reachable space. Then, after having carried out numerous experimentations, it appeared that this goal space was actually more than 25 times the size of the area accessible by the robot (see red contours in Fig. 14).

The implementation of our algorithm in such a robotic setup aims to test if the SAGG-RIAC and McSAGG-RIAC driving methods allows the robot to learn to attain a maximal amount of reachable positions, avoiding the selection

of many goals inside regions which are unreachable, or alternatively that have previously been visited.

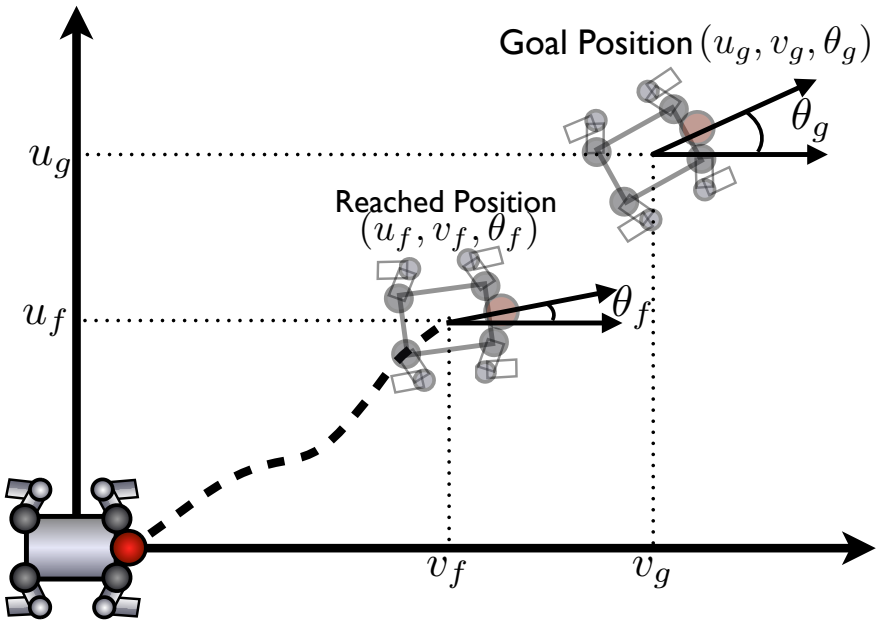

Fig. 13. Example of experimentation of the quadruped and illustration of beginning position, goal position (u_g, v_g, θ_g), and a corresponding reached position (u_f, v_f, θ_f) whose value are used to compute the measure of competence.

Measure of competence In this experiment, we do not consider requirements ρ and only focus on reaching goal positions $x_g = (u_g, v_g, \theta_g)$. In each iteration the robot is reset to a rest configuration and goal positions are defined in the robot's own body referential (see Fig. 13). We define the cost function C and thus the competence as linked with the Euclidian distance $D(x_g, x_f)$ after a reaching attempt, which is normalized by the original distance between the rest position, and the goal $D(x_{origin}, x_g)$ (See Fig. 13). This allows, for instance, assigning a same competence level when considering a goal at 1km from the origin position, which the robot approaches at 0.1km, and a goal at 100m, which the robot approaches at 10m.

In this measure of competence, we consider the rotation factor θ, and compute the Euclidian distance using (u, v, θ). Also, dimensions of the goal space are rescaled in [0;1]. Each dimension therefore has the same weight in the estimation

of competence (an angle error of $\theta = \frac{1}{2\pi}$ is as important as an error $u = \frac{1}{90}$ or $v = \frac{1}{90}$).

$$C(x_g, x_f, x_{start}) = -\frac{D(x_g, x_f)}{||x_g||} \tag{9}$$

where $C(x_g, x_f, x_{start}) = 0$ if $||x_g|| = 0$.

Local Exploration and Reaching Reaching a goal x_g necessitates the estimation of a motor synergy Υ_i leading to this chosen state x_g. Considering a single starting configuration (the rest configuration) for each experimentation, and motor synergies Υ, the forward model which defines this system can be written as the following:

$$\Upsilon \rightarrow (u, v, \theta) \tag{10}$$

Here, we have a direct relationship which only considers the 24 parameters $\{ph_{1,2,..,12}; am_{1,2,..,12}\}$ as inputs of the system, and a position in (u, v, θ) as output. We thus have a direct relationship and no context, or possible set-point, as used in the arm experiment. Also, when considering the inverse model $(u, v, \theta) \rightarrow \Upsilon$ that has to be estimated, the low-level of active learning that we use cannot be directly derived from SSA. Instead, we use the following optimization mechanism that can be divided into two different phases: a reaching phase, and an exploration phase.

3.1.5 Reaching Phase The reaching phase deals with reusing the data already learned to compute an inverse model $((u, v, \theta) \rightarrow \Upsilon)_L$ in the locality L of the intended goal $x_g = (u_g, v_g, \theta_g)$. In order to create such an inverse model (numerous can exist), we extract the potentially more reliable data using the following method:

we first compute the set L of the l nearest neighbors of (u_g, v_g, θ_g) and their corresponding motor synergies using an ANN method (Muja and Lowe, 2009):

$$L = \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_l\} \tag{11}$$

Then, we consider the set M which contains l sets of m elements:

$$M = \left\{ \begin{array}{l} \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_1 \\ \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_2 \\ ... \\ \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_l \end{array} \right\}$$

where each set $\{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_i$ corresponds to the m nearest neighbors of each Υ_i, $i \in L$, and their corresponding resulting position (u, v, θ).

For each set $\{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}_i$, we estimate the standard deviation σ of their motor synergies Υ :

$$N = \cup_{i \in M} \{\sigma\left(\Upsilon_j \in \{\{u, v, \theta, \Upsilon\}_{1,...,m}\}_i\right)\} \tag{12}$$

Finally, we select the set $O = \{\{u, v, \theta, \Upsilon\}_1, \{u, v, \theta, \Upsilon\}_2, ..., \{u, v, \theta, \Upsilon\}_m\}$ inside M such that it minimizes the standard deviation of its synergies:

$$O = \arg\min_N M \tag{13}$$

From O, we estimate a linear inverse model ($(u, v, \theta) \rightarrow \Upsilon$) by using a pseudo-inverse as introduced in the reaching experiment, and obtain the synergy Υ_g which corresponds to the desired goal (u_g, v_g, θ_g).

Exploration Phase The system here continuously estimates the distance between the goal x_g and the closest already reached position x_c. If the reaching phase does not manage to make the system come closer to x_g, i.e. $D(x_g, x_t) > D(x_g, x_c)$, with x_t as last experimented synergy in an attempt towards x_g, the exploration phase is triggered.

In this phase the system first considers the nearest neighbor $x_c = (u_c, v_c, \theta_c)$ of the goal (u_g, v_g, θ_g) and get the corresponding known synergy Υ_c. Then, it adds a random noise $rand(24)$ to the 24 parameters $\{ph_{1,2,..,12}, am_{1,2,..,12}\}_c$ of this synergy Υ_c which is proportional to the Euclidian distance $D(x_g, x_c)$. The next synergy $\Upsilon_{t+1} = \{ph_{1,2,..,12}, am_{1,2,..,12}\}_{t+1}$ to experiment can thus be described using the following equation:

$$\Upsilon_{t+1} = \begin{pmatrix} \{ph_{1,2,..,12}, am_{1,2,..,12}\}_c \\ + \lambda.rand(24).D(x_g, x_c) \end{pmatrix} \tag{14}$$

where $rand(i)$ returns a vector of i random values in $[-1; 1]$, $\lambda > 0$ and $\{ph_{1,2,..,12}, am_{1,2,..,12}\}_c$ the motor synergy which corresponds to x_c.

Constraining the Goal Space In the following, we constrain the goal space using the same maturational mechanism used in the maturationnally constrained reaching task presented above. The goal space starts as a small sphere centered around the position $(u, v, \theta) = (0, 0, 0)$, which corresponds to the rest position where the quadruped starts every displacement. Then, according to the evolution of the maturational clock, the radius of this sphere increases, until covering the entire goal space.

Constraining the Control Space Due to the high number of parameters controlling each motor synergy, the learning mechanism faces a highly redundant system. Also, because our framework considers important the fact of performing a maximal amount of tasks (i.e. goals in the task space), instead of different ways to perform a same task, constraints on the control space can be considered.

Let us consider the 24 dimensional space controlling phases and amplitudes as defined as $S = [-2\pi; 2\pi]^{12} \times [0; 1]^{12}$. We set the constrained subspace where possible values can be taken as $[\mu_i - 4\pi\sigma; \mu_i + 4\pi\sigma]^{12} \times [\mu_j - \sigma; \mu_j + \sigma]^{12} \in S$,

where μ defines a seed, different for each dimension, around which values can be taken according to a window of size 2σ, σ being function of the maturational clock $\psi(t)$. The value of those seeds is typically an innate constraint over the maturational constraints that should be the result of a biological evolutionary process. As we did not use evolutionary optimization here, we took a short cut by handcrafting the value of each seed according to the following mechanism: first, we run an experiment only using constraints in the goal space. Once this experiment terminated, we compute histograms of phases and amplitude experimented with during the exploration process. Then, the seed selected for each dimension corresponds to the maximum of the histogram, which represents the most used value during this experiment. Whereas different seeds could be imagined, we found that these handcrafted seeds were adequate for the learning efficiency of the robot.

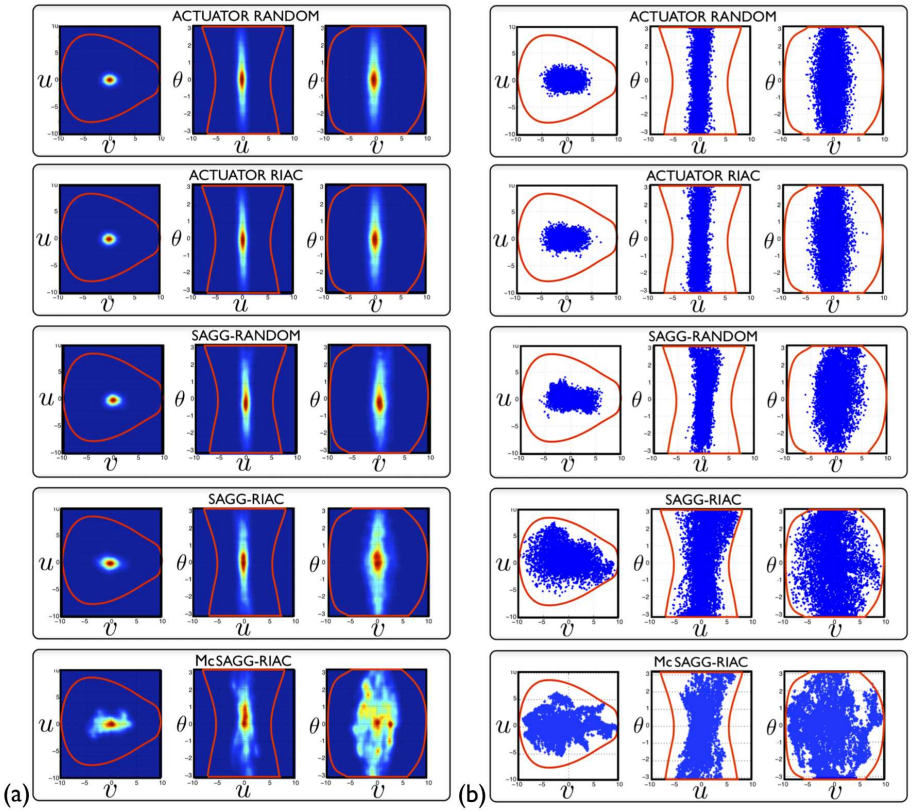

Fig. 14. Histograms of positions explored by the quadruped inside the goal space u, v, θ after 10000 experimentations (running a motor synergy during a fixed amount of time), using different exploration mechanisms.

Qualitative Results Figure 14 (a), presents representative examples of histograms of positions explored by the quadruped inside the goal space u, v, θ after 10000 experimentations (running of motor synergies during the same fixed amount of time), and (b) shows examples of the repartitions of positions inside the goal space after 10000 experimentations, when using the following exploration mechanisms:

ACTUATOR-RANDOM corresponds to a uniform selection of parameters controlling motor synergies (values inside the 24 dimensional space of phases and amplitude). ACTUATOR-RIAC corresponds to the original version of R-IAC (Baranes and Oudeyer, 2009) that actively generates actions inside the same space of synergies as ACTUATOR-RANDOM. SAGG-RANDOM is a method where the learning is situated at the level of goals which are generated uniformly in the goal space u, v, θ. Here the low-level of active learning used is the same as in SAGG-RIAC. Then, the SAGG-RIAC method corresponds to the self-generation of goals actively inside the whole goal space while McSAGG-RIAC also considers maturational constraints in both control and goal spaces.

Displaying both histograms and reached positions (i.e. displacement in the robot's own referential) allows observing different important qualitative aspects of the learning process: whereas histograms efficiently show the relative quantities of positions which have been experimented in the goal space u, v, θ, they prevent the precise observation of the volume where positions have been reached. This information is then displayed by observing the repartition of visited positions, (see Figure 14 (b)).

Comparing the two first exploration mechanisms (ACTUATOR-RANDOM and ACTUATOR-RIAC) we cannot distinguish any notable difference, the space explored appears similar and the extent of explored space on the (u, v) axis is comprised in the interval $[-5; 5]$ for u and $[-2.5; 2.5]$ for v on both graphs. Nevertheless, these results are important when comparing histograms of exploration (Fig. 14 (a)) and visited positions (Fig. 14 (b)) to the size of the reachable area (red lines on Fig. 14). It indeed shows that, in the 24 dimensional space controlling motor synergies, an extremely large part of values leads to positions close to $(0, 0, 0)$, and thus do not allow the robot to perform a large displacement. It allows us to see that reaching the entire goal space is a difficult task, which could be discovered using exploration in the space of motor synergies, only after extremely long time periods. Moreover, we notice that the difference between u and v scales is due to the inherent structure of the robot, which simplifies the way to go forward and backward rather than shifting left or right.

Considering SAGG methods, it is important to note the difference between the reachable area and the goal space. In Figure 14, red lines correspond to the estimated reachable area that is comprised of $[-10; 10] \times [-10; 10] \times [-\pi; \pi]$, whereas the goal space is much larger: $[-45; 45] \times [-45; 45] \times [-2\pi; 2\pi]$. We are also able to notice the asymmetric aspect of its repartition according to the v axis, which is due to the decentered weight of the robot's head.

The SAGG-RANDOM method seems to slightly increase the space covered on the u and v axis compared to ACTUATOR methods, as shown by the higher

concentration of positions explored in the interval $[-5;-3] \cup [3;5]$ of u. However, this change does not seem very important when comparing SAGG-RANDOM to any previous algorithm.

SAGG-RIAC, contrary to SAGG-RANDOM, shows a large exploration range compared to other methods: the surface in u has almost twice as much coverage than using previous algorithms, and in v, up to three times; there is a maximum of 7.5 in v where the previous algorithms were at 2.5. These last results emphasize the capability of SAGG-RIAC to drive the learning process inside reachable areas that are not easily accessible (hardly discovered by chance). Nevertheless, when observing histograms of SAGG-RIAC, we can notice the high concentration of explored positions around $(0, 0, 0)$, the starting position where every experimentation is launched. This signifies that, even if SAGG-RIAC is able to explore a large volume of the reachable space, as shown in Fig. 14 (b), it still spends many iterations exploring the same areas.

According to the repartition of positions shown in Fig. 14 (b) for the McSAGG-RIAC exploration mechanism, we can notice a volume explored comparable to the one explored by SAGG-RIAC. Nevertheless, it seems that McSAGG-RIAC visits a slightly lower part of the space, avoiding some areas, while explored area seems to be visited with a higher concentration. This higher concentration is confirmed via observation of histograms of McSAGG-RIAC: indeed, whereas every other methods focused during a large part of their exploration time around the position $(0, 0, 0)$, McSAGG-RIAC also focuses in areas distant from this position. The higher consideration of different areas is due to constraints fixed in the goal space, which allows a fast discovery of reachable goals and ways to reach them, whereas without constraints, the system spends high amount of time attempting unreachable goals, and thus performs movements which have a high probability to lead to position close to $(0, 0, 0)$. Also, small areas not visited can be explained by the high focalization of McSAGG-RIAC in others, as well as the limitation of values taken in the control space.

Quantitative Results In this section, we aim to test the efficiency of the learned database to guide the quadruped robot to reach a set of goal positions from its rest configuration. Here we consider a test database of 100 goals and compute the distance between each goal attempted, and the reached position. Fig. 3.1.5 shows performances of methods introduced previously. Also, in addition to the evaluation of the efficiency of McSAGG-RIAC with constraints in both control and goal spaces (called McSAGG-RIAC In&Out in Fig. 3.1.5), we introduce the evaluation of McSAGG-RIAC when only using constraints on the goal space (McSAGG-RIAC Out).

First of all, we can observe the higher efficiency of SAGG-RIAC compared to methods ACTUATOR-RANDOM, ACTUATOR-RIAC and SAGG-Random which can be observed after only 1000 iterations. The high decreasing velocity of the reaching error (in the number of experimentations) is due to the consideration of regions limited to a small number of elements (30 in this experiment). It allows the system to create a very high number of regions within a small interval of

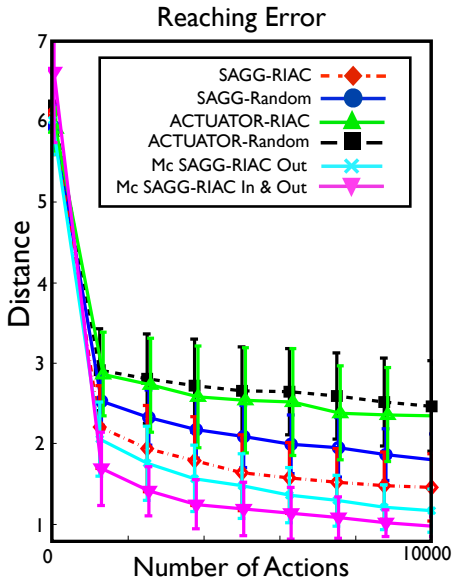

Fig. 15. Reaching Errors for different exploration methods.

time, which helps the system to discover and focus on reachable regions and its surrounding area.

ACTUATOR-RIAC shows slightly more efficient performances than ACTUATOR-RANDOM. Also, even if SAGG-RANDOM is less efficient than SAGG-RIAC, we can observe its highly decreasing reaching errors compared to ACTUATOR methods, which allows it to be significantly more efficient than these method when considered at 10000 iterations.

McSAGG-RIAC Out shows better results than SAGG-RIAC since the beginning of the evolution (1000 iterations), and decreases with a higher velocity until the end of the experiment. This illustrates the high potential of coupling constraints situated in the goal space and SAGG-RIAC in such a complex robotic setup.

Eventually, we can observe that using both constraints in both control and goal spaces as introduced by McSAGG-RANDOM In & Out allows to obtain significantly more efficient results than SAGG-RIAC without constraints ($p = 0.0055$ at the end of the exploration process), and better than when only using constraints in the goal space with a measure of significance $p = 0.05$.

In such a highly-redundant robot, coupling different types of maturational constraints with the SAGG-RIAC process thus allows to obtain significantly better performances than when using the SAGG-RIAC competence based intrinsic motivation algorithm without maturational constraints. It is important to note that "better performances" means here not only faster learning, but

better asymptotic generalization, as was already shown in earlier work on active learning (Cohn et al., 1994).

Summary These experiments exemplify the high efficiency of methods that drive the exploration at the level of goals, and then show that adding up maturational constraints improves significantly the efficiency of intrinsically motivated exploration and learning. As illustrated by qualitative results, SAGG methods, and especially SAGG-RIAC and McSAGG-RIAC, allow the robot to drive efficiently to explore large spaces containing areas hardly discovered by chance, when limits of reachability are impossible to predict. In spite of using motor primitives that already drastically reduce the huge space of physically possible quadruped movements, the dimensionality of the control space is still high, and in such spaces the experiment showed how McSAGG-RIAC could significantly improve the performances in generalization over SAGG-RIAC for learning precise omnidirectional locomotion.

Eventually, we conclude that the bidirectional coupling of maturational constraints and intrinsic motivation shall allow the self-focalization of goals inside maturationally restrained areas, which maximizes the information needed for constraints to evolve, increasing progressively the complexity of the accessible world, and thus of the acquired competences. Thus, it will be highly stimulating in the future to explore more systematically, and in various sensorimotor spaces, how the results outlined here could be reproduced and extended.

4 Integrating intrinsic motivation with socially guided learning

Because of its very nature, intrinsic motivation has often been studied separately (as in the developmental robotics literature), and opposed (as in the psychology and educational theory literature (Ryan and Deci, 2000)), to socially guided learning, many forms of which can be seen as extrinsically driven learning. Yet, in the daily life of humans, these two families of mechanisms strongly interact. Intrinsic motivation can motivate a child to follow the lessons of an interesting teacher. But reversely, and most importantly for the main question approached in this chapter, social guidance can drive a learner into new intrinsically motivating spaces/activities which it may continue to explore alone and for their own sake, but might not have discovered without the social guidance. For example, many people practice activities like Sudoku, tennis, painting, or sculpture driven by intrinsic motivation, but most of them discovered the very existence of those explorable activities by observing others practice it. Furthermore, while they practice activities like painting or Sudoku driven by intrinsic motivation, they may acquire new strategies for achieving those intrinsically motivated activities by observing others or by listening to their advices. Thus, social guidance is often a fundamental mechanism allowing an intrinsically motivated learner both to discover new potential explorable task spaces as well as new example of successful control strategies.

This role of social guidance shall quickly become essential and necessary for any robot built for open-ended development in high-dimensional unbounded spaces, even when equipped with advanced competence based intrinsic motivation operating on sophisticated sensorimotor primitives and with maturational constraints. Indeed, there are at least three important potential limits to the developmental constraints on intrinsic motivation presented so far in this chapter, that we describe in the following.

How to discover isolated islands of learnable goals? While maturational constraints in a task space can allow us to harness the problem of unboundedness by growing progressively the explorable region(s), it may as a side effect make it difficult for a sole intrinsic motivation system to discover disconnected islands of learnability in the whole task space. Let us take the example of a grasping task space defined in terms of the shape properties of objects. This is an unbounded space since the space of all objects around us is infinite (especially if we include things like walls, trees or clouds as potential objects to learn how to grasp). Using maturational constraints in a competence based approach would amount to start from a few basic object shapes for which to explore grasp motor commands and see how they work, and then progressively and continuously extend the space of explorable shapes to try to grasp. This would allow the robot to learn efficiently how to grasp a growing set of object shapes. Yet, the shape space is complicated and there are many object shapes which the robot could learn to grasp, but will never learn because either it would take too much time for the maturationally growing explorable space to reach them or because they are surrounded by unlearnable shapes that would prevent the explorable space to grow and reach them. Social guidance could be of essential help here: instead of waiting that the explorable space grows until reaching those objects, a simple social signal drawing the attention and motivation of the learner towards new specific objects in these islands could be used to start a new local and growing region of exploration around this object. The same applies to all kinds of task spaces. Initial work towards this objective has been presented in (Nguyen et al., 2011).

How to discover new explorable task spaces? All the experiments presented above, similarly to what exists in many other models of intrinsically motivated exploration in the literature, assumed a small finite set of pre-defined tasks spaces, i.e. a small set of groups of variables describing aspects of the world that the robot may learn to manipulate through exploration. The reaching task experiment involved only one such task space, while the Playground Experiment mixed several predefined task spaces. As a consequence, the robot could potentially achieve open-ended development of skills within those task spaces, but is was impossible for it to learn skills outside those task spaces, for the simple reason that they had no mechanism for discovering and representing novel task spaces, i.e. novel groups of variables potentially interesting to learn to manipulate. In principle, modifications of the associated intrinsic motivation architecture could be done to address this limitation. One could provide the robot with a very large set of potential task space variables, as well as with operators to build new

such variables, and then use a higher-level active exploration mechanisms which would generate, sample and select the new task spaces in which lower-level intrinsically motivated exploration could provide competence progress. Yet, even if one would constrain task spaces to be composed of only limited number of dimensions/variables (e.g. below six or seven), such an approach would have to face a very hard combinatorial problem. If one would like to learn the same kind of task variety as a human infant does, then the number of variables that may be considered by the learner shall be large. And as a mechanical consequence, the number of potential subsets of these variables, defining potentially new task spaces, would grow exponentially in such a way that even active exploration would be again inefficient. Thus here again social guidance may be essential: either through observation or direct gestural or linguistic guidance, a learner may infer the task dimensions/variables that characterize a new task space to explore later on through intrinsic motivation. The literature on robot learning by imitation/demonstration has already developed statistical inference mechanisms allowing to infer new task constraints/dimensions (Calinon et al., 2007; Cederborg et al., 2010; Lopes et al., 2009; Ng and Russell, 2000). These techniques could usefully be reused by intrinsically motivated learning architectures to expand efficiently the set of explorable task spaces. Initial work in this direction has for example been presented in (Thomaz and Breazeal, 2008).

How to discover new explorable sensorimotor primitives? Similarly to task spaces, all the experiments presented earlier, and many other approaches in the literature, assumed that the set of sensorimotor "tools", i.e. sensorimotor variables and the primitives operating on them, were predefined and finite. Of course, using techniques like reinforcement learning can allow a robot to learn how new sequences of motor commands can allow it to achieve a given task, but the dimensions in which these elementary commands are encoded are typically finite and predefined (but yet might be high-dimensional). Likewise, the set of sensorimotor primitives, providing the structure necessary to bootstrap the intrinsically motivated learning of many complicated tasks as argued earlier, is often predefined and thus limited. Just like for expanding the space of explorable task spaces, social guidance can also be an essential mechanism allowing a robot to discover new useful control dimensions, and new associated motor primitives, that it may reuse to explore a task space through intrinsic motivation. Let us take the example of the "playing tennis" space. In addition to being a space typically discovered by seeing others play tennis, observation of others is also crucial for the discovery of 1) which control variables are important, such as for example the relative position of the feet to the net at the moment of striking the ball, and 2) which motor primitives based on those control dimensions shall be explored, such as for example the fore hand, back hand or volley motor primitives. Once prototypical back hand or volley primitives have been observed, the learner can explore through intrinsic motivation the space of variations of these primitives and how to tune their parameters to the real time trajectory of the ball. But without the observation of such movements in others, learning to play tennis efficiently would be an extremely difficult challenge. Finally, at a finer

grain, when prototypes and dimensions for new sensorimotor primitives have been discovered, social guidance can continue to play in concert with intrinsic motivation by continuously providing new examples of variations of those primitives that may be repeated and explored by the learner (e.g. see Nguyen et al., 2011). Again, the literature on robot learning by imitation/demonstration has elaborated many algorithms allowing the acquisition of new motor primitives (Billard et al., 2008; Calinon et al., 2007; Cederborg et al., 2010; Grollman and Jenkins, 2010), and future research in developmental robotics might strongly benefit from integrating computational models of intrinsic motivation and socially guided learning.

5 Conclusion

Research on computational approaches to intrinsic motivation has allowed important conceptual advances in developmental robotics in the last decade, opening new horizons for the building of machines capable of open-ended learning and development. Many kinds of models and formalisms have been proposed and integrated in sophisticated architectures, such as in intrinsically motivated reinforcement learning, coming from a wide variety of research groups and backgrounds, ranging from machine learning, statistics, cognitive modeling to robotics. Diverse proof-of-concept experiments have also been achieved.

Yet, several important challenges are now in need to be addressed. Among them, a paramount objective is to study how it is possible to scale those conceptual and proof-of-concept initial results to the real world. How can a robot, equipped with a rich sensorimotor apparatus, develop new skills in an open high-dimensional unbounded uncontrolled environment, just as human children do? We are very far from having both conceptual and technical answer(s) to this question. Intrinsic motivation will certainly be a key element, but this chapter has shown that it can only become useful and used when other complementary mechanisms are harmoniously integrated in a single developmental architecture. Intrinsic motivation alone is indeed nearly helpless in unconstrained unbounded sensorimotor spaces. The growth of complexity should be controlled by the interaction of intrinsic motivation and other families of developmental constraints. In this chapter, we have argued that sensorimotor primitives, self-organization and embodiment, task space representations, maturational mechanisms, and social guidance should be considered as essential complements to intrinsic motivation for open-ended development in the real world. Other families of mechanisms, which we did not discuss, will be equally important, including developmental biases on representations, on mechanisms for creating abstractions, on operators for combining and re-using knowledge and skills and creating novel representations, or on statistical inference.

The challenges posed by the real world cannot be reduced to mere algorithmic complexity and/or efficiency problems. Unlearnability, high-dimensionality and unboundedness introduce new fundamental conceptual obstacles. Constraints and biases, either innate or self-organized, either static or evolving, are un-

avoidable for any real-world developmental learning system. And most probably, those constraints that include intrinsic motivation together with maturation or social guidance, will function efficiently only when integrated together properly. How shall social guidance, in its many different forms, be integrated with computational models of intrinsically motivated learning? How shall maturation and intrinsic motivation control each other? How novel sensorimotor primitives, and associated higher level representations and abstractions, can be constructed through the interaction of intrinsic motivation, maturation and social learning? Which constraints should be pre-wired explicitly, and which one should be self-organized? Can we understand how a given body/embodiment can help the development of certain families of skills rather than others? These fundamental questions shall become an essential target of research on intrinsic motivation and developmental robotics.

Furthermore, the unavoidability of constraints and biases in the real world indicates that no general purpose machine can be made capable of learning universally, and at the same time efficiently, anything in any environment. As a consequence, a set of fundamental conceptual questions raised by trying to scale to the real world concern the very concept of open-ended and task-independent learning: we should try to understand better how this property that we observe in human children is different from omni-capable any-task learning. Related to this, we should try to understand how certain families of constraints in certain families of environment allow or disallow the development of certain families of skills.

Finally, future research shall strongly rely on larger-scale, "more real world" experiments with high-dimensional robots in environments functionally and structurally as similar as those encountered by human infants. There were many good reasons for conducting toy-level experiments so far, but this had the consequence to short-cut much of the specific conceptual and technical difficulties posed by the real world. Confronting to the reality shall be an efficient constraint to guide research in a direction that may allow robots to acquire novel skills like human infants do. Furthermore, an associated methodological need for future research is to construct both explanations and understanding of our experiments such as to provide an appropriate emphasis on all components/constraints that allow these experiments to actually "work". Because one may be most interested by intrinsic motivation, it is sometimes tempting to emphasize the role of intrinsic motivation in the interpretation of experimental results. This is illustrated by the Playground Experiment: while often presented, including by us, under the light of the concepts of "learning progress" and intrinsic motivation, its success was actually due to the *combination* and *interaction* of "learning progress"-based intrinsic motivation with an innate parameterized repertoire of dynamic sensorimotor primitives.

As long held by the embodied and situated cognition literature, adaptive behavior as well as sensorimotor and cognitive development shall not be the result of isolated localized components, but rather the results of the dynamical interactions of all the components of a complete creature - mental and body compo-

nents - among themselves and with their physical and social environment. Thus, research on intrinsic motivation shall now focus on "boundaries": establishing coordinated links between intrinsic motivation and its functional boundaries, i.e. with the other constraints on exploration and learning of the complete creature, shall help robots to control better the progressive growth of the boundaries of their own knowledge and capabilities.

6 Acknowledgements

Many of the ideas presented in this paper benefitted from discussions and joint work with our colleagues, in particular Jérome Béchu, Fabien Benureau, Thomas Cederborg, Fabien Danieau, Haylee Fogg, David Filliat, Paul Fudal, Verena V. Hafner, Matthieu Lapeyre, Manuel Lopes, Olivier Ly, Olivier Mangin, Mai Nguyen, Luc Steels, Pierre Rouanet, Andrew Whyte. This research was partially funded by ERC Starting Grant EXPLORER 240007.

Bibliography

Angluin, D. (1988). Queries and concept learning. *Machine Learning*, 2:319–342.

Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M., and Yoshida, C. (2009). Cognitive developmental robotics: A survey. *IEEE Trans. Autonomous Mental Development*, 1(1).

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

Ball, P. (1999). *the Self-Made Tapestry-Pattern formation in nature*. Oxford University Press.

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

Baranes, A. and Oudeyer, P.-Y. (2010a). Intrinsically motivated goal exploration for active motor learning in robots: a case study. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*.

Baranes, A. and Oudeyer, P.-Y. (2010b). Maturationally constrained competence-based intrinsically motivated learning. In *Proceedings of IEEE International Conference on Development and Learning (ICDL 2010)*.

Baranes, A. and Oudeyer, P.-Y. (2011). The interaction of maturational constraints and intrinsic motivations in active motor development. In *Proceedings of IEEE ICDL-Epirob 2011*.

Baranes, A. and Oudeyer, P.-Y. (submitted). Competence-based intrinsic motivation for active motor learning in robots.

Barto, A., Singh, S., and Chenatez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *Proc. 3rd Int. Conf. Development Learn.*, pages 112–119, San Diego, CA.

Berk, L. (2008). *Child development*. Allyn and Bacon.

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

Berthier, N. E., Clifton, R., McCall, D., and Robin, D. (1999). Proximodistal structure of early reaching in human infants. *Exp Brain Res*.

Billard, A., Calinon, S., Dillmann, R., and Schaal, S. (2008). *Handbook of Robotics*, chapter Robot Programming by Demonstration. Springer.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press, Oxford, UK.

Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition.

Bjorklund, D. (1997). The role of immaturity in human development. *Psychological Bulletin*, 122(2):153–169.

Blank, D., Kumar, D., Meeden, L., and Marshall, J. (2002). Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture. *Cybernetics and Systems*, 36(2).

Brafman, R. and Tennenholtz, M. (2001). R-max: A general polynomial time algorithm for near-optimal reinforcement learning. In *Proceedings of IJCAI'01*.

Bremner, J. and Slater, A., editors (2003). *Theories of Infant Development*. Cambridge,MA:Blackwell.

Bronson, G. (1974). The postnatal growth of visual capacity. *Child. Dev.*, 45(4):873–890.

Calinon, S., Guenter, F., and Billard, A. (2007). On learning, representing and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man and Cybernetics, Part B.*

Castro, R. and Novak, R. (2008). Minimax bounds for active learning. *IEEE Trans. Information Theory*, 54.

Cazalets, J., Borde, M., and Clarac, F. (1995). Localization and organization of the central pattern generator for hindlimb locomotion in newborn rat. *Journal of Neuroscience*, 15:4943–4951.

Cederborg, T., Ming, L., Baranes, A., and Oudeyer, P.-Y. (2010). Incremental local online gaussian mixture regression for imitation learning of multiple tasks. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)*.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental desing: A review. *J. Statistical Science*, 10.

Chung, W., Fu, L.-C., and Hsu, S.-H. (2008). *Handbook of Robotics*, chapter Motion control, pages 133–159. Springer.

Cohn, D., Atlas, L., and Ladner, R. (1994). Improving generalization with active learning. *Mach. Learn.*, 15(2):201–221.

Cohn, D., Ghahramani, Z., and Jordan, M. (1996). Active learning with statistical models. *Journal of Artificial Intellifence Research*, 4:129–145.

Csikszentmihalyi, M. (1996). *Creativity-Flow and the Psychology of Discovery and Invention*. Harper Perennial, New York.

d'Avella, A., Portone, A., Fernandez, L., and Lacquaniti, F. (2006). Control of fast-reaching movement by muscle synergies combinations. *The Journal of Neuroscience*, 26(30):7791–7810.

d'Avella, A., Saltiel, P., and Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.*, 6:300–308.

Dayan, P. and Belleine, W. (2002). Reward, motivation and reinforcement learning. *Neuron*, 36:285–298.

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

Deci, E. and Ryan, M. (1985). *Intrinsic Motivation and self-determination in human behavior*. Plenum Press, New York.

Dick, T., Oku, Y., Romaniuk, J., and Cherniack, N. (1993). Interaction between cpgs for breathing and swallowing in the cat. *J. Physiol.*, 465:715–730.

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

Eyre, J. (2003). *Development and Plasticity of the Corticospinal System in Man*. Hindawi Publishing Corporation.

Faller, D., Klingmller, U., and Timmer, J. (2003). Simulation methods for optimal experimental design in systems biology. *Simulation*, 79:717–725.

Fedorov, V. (1972). *Theory of Optimal Experiment*. Academic Press, Inc., New York, NY.

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

Fisher, K. and Silvern, L. (1985). Stages and individual differences in cognitive development. *Ann. Rev. Psychol.*, 36:613–648.

Franceschini, N., Pichon, J., and Blanes, C. (1992). From insect vision to robot vision. *Phil. Trans. R. Soc. Lond. B*, 337:283–294.

Ghahramani, Z. (1993). Solving inverse problems using an em approach to density estimation. In Mozer, M., Smolensky, P., Toureztky, D., Elman, J., and Weigend, A., editors, *Proceedings of the 1993 Connectionist Models Summer School*.

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

Grollman, D. H. and Jenkins, O. C. (2010). Incremental learning of subtasks from unsegmented demonstration. In *International Conference on Intelligent Robots and Systems*, Taipei, Taiwan.

Hart, S. and Grupen, R. (2008). Intrinsically motivated hierarchical manipulation. In *Proceedings of the 2008 IEEE Conference on Robots and Automation (ICRA)*.

Huang, X. and Weng, J. (2002). Novelty and reinforcement learning in the value system of developmental robots. In Prince, C., Demiris, Y., Marom, Y., Kozima, H., and Balkenius, C., editors, *Proceedings of the 2nd international workshop on Epigenetic Robotics : Modeling cognitive development in robotic*

systems, pages 47–55. Lund University Cognitive Studies 94.

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

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

Iida, F. and Pfeifer, R. (2004). Cheap and rapid locomotion of a quadruped robot: Self-stabilization of bounding gait. In et al., F. G., editor, *Intelligent Autonomous Systems 8*.

James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Predd.

Johnson, M. (2001). Functional brain development in humans. *Nat. Rev. Neurosc.*, 2(7):475–483.

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

Kemp, C. and Edsinger, A. (2006). What can i control?: The development of visual categories for a robots body and the world that it influences. In *In 5th IEEE International Conference on Development and Learning (ICDL-06), Special Session on Autonomous Mental Development*.

Khatib, O. (1987). A unified approach for motion and force control of robot manipulators: The operational space formulation. *Robotics and Automation, IEEE Journal of*, 3(1):43–53.

Konczak, J., Borutta, M., and Dichgans, J. (1997). The development of goal-directed reaching in infants. learning to produce task-adequate patterns of joint torque. *Experimental Brain Research*.

Kumar, S., Narasimhan, K., Patwardhan, S., and Prasad, V. (2010). Experiment design, identification and control in large-scale chemical processes. In *Modelling, Identification and Control (ICMIC), The 2010 International Conference on*, pages 155 –160.

Lee, M., Meng, Q., and Chao, F. (2007). Staged competence learning in developmental robotics. *Adaptive Behavior*, 15(3):241–255.

Lee, W. (1984). Neuromotor synergies as a basis for coordinated intentional action. *J. Mot. Behav.*, 16:135–170.

Lopes, M., Melo, F., and Montesano, L. (2009). Active learning for reward estimation in inverse reinforcement learning. In *Proceedings of European Conference on Machine Learning (ECML/PKDD)*.

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

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

Ly, O., Lapeyre, M., and Oudeyer, P.-Y. (2011). Bio-inspired vertebral column, compliance and semi-passive dynamics in a lightweight robot. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*.

Ly, O. and Oudeyer, P.-Y. (2010). Acroban the humanoid: playful and compliant physical child-robot interaction. In *ACM Siggraph Emerging Technologies*, pages 1–1.

MacNeilage, P. (2008). *The Origin of Speech*. Oxford University Press.

Meltzoff, A. and Moore, M. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312):75–8.

Meyer, J. A. and Wilson, S. W., editors (1991). *A possibility for implementing curiosity and boredom in model-building neural controllers*. MIT Press/Bradford Books.

Modayil, J., Pilarski, P., White, A., Degris, T., and Sutton, R. (2010). Off-policy knowledge maintenance for robots. In *Proceedings of Robotics Science and Systems Workshop (Towards Closing the Loop: Active Learning for Robotics)*.

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

Moore, A. (1992). Fast, robust adaptive control by learning only forward models. In *Advances in Neural Information Processing Systems 4*.

Muja, M. and Lowe, D. (2009). Fast approximate nearest neighbors with automatic algorithm. In *International Conference on Computer Vision Theory and Applications (VISAPP'09)*.

Ng, A. Y. and Russell, S. (2000). Algorithms for inverse reinforcement learning. In *in Proc. 17th International Conf. on Machine Learning*, pages 663–670. Morgan Kaufmann.

Nguyen, M., Baranes, A., and Oudeyer, P.-Y. (2011). Bootstrapping intrinsically motivated learning with human demonstrations. In *Proceedings of IEEE ICDL-Epirob 2011*.

Nguyen-Tuong, D. and Peters, J. (2011). Model learning in robotics: a survey. *Cognitive Processing*.

Oudeyer, P.-Y. (2010). On the impact of robotics in behavioral and cognitive sciences: from insect navigation to human cognitive development. *IEEE Transactions on Autonomous Mental Development*, 2(1):2–16.

Oudeyer, P.-Y. and Kaplan, F. (2006). The discovery of 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*, 1:6.

Oudeyer, P.-Y. and Kaplan, F. (2008). How can we define intrinsic motivations ? In *Proc. Of the 8th Conf. On Epigenetic Robotics.*

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

Oudeyer, P.-Y., Ly, O., and Rouanet, P. (2011). Exploring robust, intuitive and emergent physical human-robot interaction with the humanoid acroban. In *Proceedings of IEEE-RAS International Conference on Humanoid Robots.*

Paul, C. (2004). Morphology and computation. In *Proceedings of the International Conference on the Simulation of Adaptive Behaviour.*

Peters, J. and Schaal, S. (2008). Natural actor critic. *Neurocomputing*, (7-9):1180–1190.

Pfeifer, R. and Bongard, J. C. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence (Bradford Books).* The MIT Press.

Pfeifer, R., Lungarella, M., and Iida, F. (2007). Self-organization, embodiment, and biologically inspired robotics. *Science*, 318:1088–1093.

Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence.* MIT Press, Boston, MA, USA.

Piaget, J. (1952). *The Origins of Intelligence in Childhood.* International University Press.

Ring, M. (1994). *Continual Learning in Reinforcement Environments.* PhD thesis, University of Texas at Austin.

Rochat, P. (1989). *Developmental Psychology*, 25:871–884.

Rolf, M., Steil, J., and Gienger, M. (2010). Goal babbling permits direct learning of inverse kinematics. *IEEE Trans. Autonomous Mental Development*, 2(3):216–229.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54 – 67.

Schaal, S. and Atkeson, C. G. (1994). Robot juggling: an implementation of memory-based learning. *Control systems magazine*, pages 57–71.

Schaal, S. and Atkeson, C. G. (1995). *Robot learning by nonparametric regression*, pages 137–153. elsevier.

Schembri, M., Mirolli, M., and Baldassare, G. (2007a). Evolving internal reinforcers for an intrinsically motivated reinforcement learning robot. In Demiris, Y., Scassellati, B., and Mareschal, D., editors, *Proceedings of the 6th IEEE International Conference on Development and Learning (ICDL2007)*.

Schembri, M., Mirolli, M., and G., B. (2007b). Evolution and learning in an intrinsically motivated reinforcement learning robot. In Springer, editor, *Advances in Artificial Life. Proceedings of the 9th European Conference on Artificial Life*, pages 294–333, Berlin.

Schlesinger, M. (2008). Heterochrony: It's (all) about time! In Studies, L. U. C., editor, *Proceedings of the Eighth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 111–117, Sweden.

Schmidhuber, J. (1991). Curious model-building control systems. In *Proc. Int. Joint Conf. Neural Netw.*, volume 2, pages 1458–1463.

Schmidhuber, J. (2002). Exploring the predictable. In Ghosh, S. and S., T., editors, *Advances in Evolutionary Computing: theory and applications*, pages 579–612. Springer-Verlag New York.

Schmidhuber, J. (2006). Optimal artificial curiosity, developmental robotics, creativity, music, and the fine arts. *Connection Science*, 18(2).

Schmidhuber, J. (2010). Formal theory of creativity. *IEEE Transation on Autonomous Mental Development*, 2(3):230–247.

Scholz, J., Klein, M., Behrens, T., and Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature neuroscience*, 12(11):1367–1368.

Sekuler, R. and Blake, R. (1994). *perception*. New-York:McGraw-Hill.

Sigaud, O., Salan, C., and Padois, V. (2011). On-line regression algorithms for learning mechanical models of robots: A survey. *Robotics and Autonomous Systems*, (0):–.

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

Stout, A. and Barto, A. (2010). Competence based intrinsic motivation. In *Proceedings of IEEE International Conference on Development and Learning (ICDL 2010)*.

Sutton, R. (1990). Integrated architectures for learning, planning, and reacting based on approximating integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the*

International Machine Learning Conference, pages 212–218.

Sutton, R. and Barto, A. (1998). *Reinforcement learning: an introduction*. MIT Press.

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

Szita, I. and Lorincz, A. (2008). The many faces of optimism: a unifying approach. In *Proceedings of ICML'08*.

Thomaz, A. and Breazeal, C. (2008). Experiments in socially guided exploration: Lessons learned in building robots that learn with and without human teachers. *Connection Science*, 20(2-3):91–110.

Thrun, S. (1992). The role of exploration in learning control. In White, D. and Sofge, D., editors, *Handbook for Intelligent Control: Neural, Fuzzy and Adaptive Approaches*. Van Nostrand Reinhold, Florence, KY, USA.

Thrun, S. and Moller, K. (1992). Active exploration in dynamic environments. In J. Moody, S. Hanson, R. L., editor, *Proc. of Advances of Neural Information Processing Systems 4*.

Ting, L. and McKay, J. (2007). Neuromechanics of muscle synergies for posture and movement. *Curr. Opin. Neurobiol.*, 17:622–628.

Tong, S. and Chang, E. (2001). Support vector machine active learning for image retrieval. In *Proceedings of the ninth ACM international conference on Multimedia*, MULTIMEDIA '01, pages 107–118. ACM.

Turkewitz, G. and Kenny, P. (1985). The role of developmental limitations of sensory input on sensory/perceptual organization. *J Dev Behav. Pediatr.*, 6(5):302–6.

Weiss, E. and Flanders, M. (2004). Muscular and postural synergies of the human hand. *J. Neurophysiol.*, 92:523–535.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., and Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, 291(599-600).

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

Whitehead, S. (1991). A study of cooperative mechanisms for faster reinforcement learning. Tr-365, University of Rochester.

Wiering, M. and Schmidhuber, J. (1997). Hq-learning. *Adaptive Behavior*, 6:219–246.

Wundt, W. (1874). *Grundzuge der physiologischen Psychologie*. Leipzig: Engelmann.

Yokoi, H., Hernandez, A., Katoh, R., Yu, W., Watanabe, I., and Maruishi, M. (2004). *Embodied artificial intelligence*, chapter Mutual adaptation in a prosthetics application. Springer LNAI 3139.

Ziegler, M., Iida, F., and Pfeifer, R. (2006). Cheap underwater locomotion: roles of morphological properties and behavioural diversity. In *Proc. Int. Conf. on Climbing and Walking Robots*.