The Playground Experiment: Task-Independent Development of a Curious Robot

Pierre-Yves Oudeyer and **Frédéric Kaplan** and **Verena V. Hafner** and **Andrew Whyte**
{oudeyer, kaplan, hafner,whyte}@csl.sony.fr
Sony Computer Science Lab, Paris

Abstract

This paper presents the mechanism of Intelligent Adaptive Curiosity. This is an intrinsic motivation system which pushes the robot towards situations in which it maximizes its learning progress. It makes the robot focus on situations which are neither too predictable nor too unpredictable. This mechanism is a source of autonomous mental development for the robot: the complexity of its activities autonomously increases and a developmental sequence appears without being manually constructed. We test this motivation system on a real robot which evolves on a baby playmat with objects that it can learn to manipulate. We show that it first spends time in situations which are easy to learn, then shifts progressively its attention to situations of increasing difficulty, avoiding situations in which nothing can be learnt.

The challenge of autonomous mental development

All humans develop in an autonomous open-ended manner through life-long learning. So far, no robot has this capacity. Yet, building such a robot is one of the greatest challenges to robotics today, and is the long-term goal of the growing field of developmental robotics ((Lungarella *et al.* 2003)).

There are two characteristic properties of human infant development that should inspire us. First of all, development involves the progressive increase of the complexity of the activities of children with an associated increase of their capabilities. Moreover, infants' activities have always a complexity which is well fitted to their current capabilities. Children undergo a developmental sequence during which each new skill is acquired only when associated cognitive and morphological structures are ready. For example, children learn first to roll over, then to crawl and sit, and only when these skills are operational, they begin to learn how to stand. Development is progressive and incremental. Inspired by this, some roboticists have realized that learning a given task could be made much easier for a robot if it followed a developmental sequence (e.g. "Learning form easy mission" ((Asada *et al.* 1996)). But very often, roboticists craft the developmental sequence by hand: they manually build simpler versions of a complex task and put the robot successively in versions of increasing complexity. This technique is very useful in many cases, but has shortcomings which limit severely our capacity to build robots that develop in an open-ended manner. Indeed, this is not practical: for each task that one wants the robot to learn, we have to design versions of this task of increasing complexity, and we also have to design manually a reward-function dedicated to this particular task. This might be all right if one is interested in only one or two tasks, but a robot capable of life-long learning should eventually be able to perform thousands of tasks.

This leads us to a second property of child development by which we should be inspired: it is autonomous and active. Of course, adults help by scaffolding their environment, but this is just a help: eventually, infants decide by themselves what they do, what they are interested in, and what their learning situations are. They are not forced to learn the tasks suggested by adults, they can invent their own. Thus, they construct by themselves their developmental sequence. Anyone who has ever played with an infant in its first year knows that for example it is extremely difficult to get the child to play with a toy that is chosen by the adult if other toys and objects are around. In fact, most often the toys that we think are adapted to them and will please them are not at all the one they prefer: they can have much funnier and instructive play experiences with adult objects, such as magazines, keys, or flowers. Also, most of the time infants engage in particular activities for their own sake, rather than as steps towards solving practical problems. This is indeed the essence of play. This suggests the existence of intrinsic motivations, which provide internal rewards during these play experiences. Such motivations are obviously useful, since they allow to learn many skills that will potentially be readily available later on for challenges and tasks which are not yet foreseeable.

In order to develop in an open-ended manner, robots should certainly be equipped with capacities for autonomous and active development, and in particular with intrinsic motivation systems, an engine for task-independant learning. This crucial topic is still largely an underinvestigated issue. Only a few researchers have suggested to implement drives for novelty or artificial curiosity. Schmidhuber, Thrun and Hermann provided initial implementations of artificial curiosity, but they did not integrate this concept within the problematics of developmental robotics ((Schmidhuber 1991), (Thrun 1995), and (Herrmann, Pawelzik, & Geisel 2000)). The first integrated view of developmental robotics that incorporated a proposal for a novelty drive was described by Weng and colleagues ((Weng 2002); (Huang & Weng 2002)). Then, Kaplan and Oudeyer proposed an implementation of artificial curiosity within a developmental framework ((Kaplan & Oudeyer 2003)), and Marshall, Blank and Meeden as well as Barto, Sing and Chentanez suggested variations on the novelty drive ((Marshall, Blank, & Meeden 2004), (Barto, Singh, & Chentanez 2004)). As we will explain later on in the paper, these pioneering systems have a number of limitations making them impossible to use on real robots in real uncontrolled environments. Furthermore, to our knowledge, it has not been shown yet how they could successfully lead to the autonomous formation of a developmental sequence comprising more than one stage. This means that typically they have allowed for the development and emergence of one level of behavioural patterns, but did not show how new levels of more complex behavioural patterns could emerge without the intervention of a human or a change in the environment provoked by a human.

In this paper, we will present a system called Intelligent Adaptive Curiosity, which is an intrinsic motivation system, coupled with the adequate action-selection mechanism, for autonomous and active development. It is based on two key concepts: 1) what the robot is ultimately interested in is the decrease of errors in predicting the consequences of its actions; 2) this decrease of errors in prediction is evaluated by comparing the current error rate to the one in similar sensorimotor contexts in the past rather than to the error rate in immediately preceding sensorimotor contexts. This system breaks the limitations of previous ones and is applied in a real world robotic set up, called the Playground Experiment.

Intrinsic motivation systems

The limits of existing systems

As stated in the last section, existing approaches to intrinsic motivations are all based on an architecture which comprises a machine which learns to anticipate the consequence of the robot's actions, and in which these actions are actively chosen according to some internal measures related to the novelty or predictability of the anticipated situation. Thus, the robots in these approaches can be described as having two modules: 1) one module implements a learning machine M which learns to predict the sensorimotor consequences when a given action is executed in a given sensorimotor context; 2) another module is a meta learning machine metaM which learns to predict the error that machine M makes in its prediction. The existing approaches can be divided into two groups, according to the way action-selection is made depending on the predictions of M and metaM.

In the first group (e.g. (Huang & Weng 2002); (Thrun 1995); (Marshall, Blank, & Meeden 2004); (Marshall, Blank, & Meeden 2004)) robots directly use the error predicted by metaM to choose which action to do[1]. The action that they choose at each step is the one for which metaM predicts the largest error in prediction of M. This has shown to be extremely efficient when the machine M has to learn a mapping which is learnable, deterministic and with homogeneous Gaussian noise ((Cohn, Atlas, & Ladner 1994); (Thrun 1995); (Weng 2002); (Barto, Singh, & Chentanez 2004)). But this method shows limitations when used in a real uncontrolled environment. Indeed, in such a case, the mapping that M has to learn is not anymore deterministic, and the noise is vastly inhomogeneous. Practically, this means that a robot using this method will for example be stuck by white noise or situations which are inherently too complex for its learning machinery. Suppose for example that there is a television in the room on a channel with no program and just "snow": obviously, nothing can be more unpredictable than this pattern and a robot motivated to find situations difficult to predict will stare fascinated in front of the television for hours.

A second group of models tried to avoid getting stuck in the presence of pure noise or unlearnable situations by using more indirectly the prediction of the error of M (e.g. ((Schmidhuber 1991); (Herrmann, Pawelzik, & Geisel 2000); (Kaplan & Oudeyer 2003))). In these models a third module that we call KGA for Knowledge Gain Assessor is added to the architecture. This new module enhances the capabilities of the meta-machine metaM: KGA allows to predict the mean error rate of M in the close future and in the next sensorimotor contexts. KGA also stores the recent mean error rate of M in the most recent sensorimotor contexts. The crucial point of these models is that candidate actions are evaluated using the expected difference between the expected mean error rate in the close future, and the mean error rate in the close past. The action which is chosen is that for which KGA predicts that it will lead to the greatest decrease of the mean error rate of M. Now the robot will not stay for a long time in front of white noise or in completely unlearnable situations because this does not lead to a decrease of its errors in prediction.

However, this method has always been tested in deterministic controlled environments with no noise, and thus never showed that it could actually overcome the limits of the previous method. And indeed, it is not difficult to realize that a new problem arises: the robot will now get stuck in the alternation of completely predictable and completely unpredictable situations, as explained by Oudeyer and Kaplan in (Oudeyer & Kaplan 2004). In front of a television with "snow", the robot will find that alternatively opening (noisy, very unpredictable visual image) and closing (completely black and predictable image) is very interesting.

Intelligent Adaptive Curiosity

We have developed an intrinsic motivation system which breaks the limitation that we just mentioned as we will show. We call this system **Intelligent Adaptive Curiosity** ((Oudeyer & Kaplan 2004)), which can be abbreviated as IAC. It starts from the same idea than the second group that we presented: the robot is motivated by performing actions

[1] Of course, we are only talking about the "novelty" drive here: their robots are sometimes equipped with other competing drives or can respond to external human based reward sources

which lead to a decrease in the mean error rate in the predictions of M. Another formulation is to say that the robot is motivated to maximize its learning progress. The crucial feature which we add is the way this learning progress (or decrease in mean error rate) is evaluated. We do not compare anymore the expected mean error rate in the expected next sensorimotor contexts to the mean error rate of the most recent sensorimotor context, but rather we compare the expected mean error rate in the expected next sensorimotor contexts to the mean error rate which occured in similar sensorimotor contexts in the past. Briefly, we do not compare the performance between two activities which happen sequentially in time, but between two activities which are similar. We use the term "local learning progress evaluation". This way of evaluating learning progress keeps the robot away both from situations which are too predictable and from situations which are too unpredictable. Indeed, the pathologic behaviors that we described in the last section are avoided. This is why we call this kind of curiosity "intelligent". For example, if we take the example of the robot in a room with a "snow" television, then the situation in which it successively opens and closes its eyes in front of it is not any more interesting for the robot. Indeed, if the robot closes its eyes after looking at the white noise, the system will compare its performance in prediction not to the previous seconds in which the robot had its eyes open, because the sensorimotor context is quite different, but to the last time the robot had its eyes closed. And it will detect that there is no decrease in the mean error rate: it is not an interesting situation. Such a system is not trapped by noise or unlearnable situations.

We will now describe how this system can be fully implemented. This implementation can be varied in many manners, for example by replacing the implementation of the learning machines M, metaM and KGA. The one we provide is basic and was developed for its practical efficiency. Also, it will be clear to the reader that in an efficient implementation, the machines M, metaM and KGA are not easily separable (keeping them separate entities in the previous paragraphs was for reasons of keeping the explanation easier to understand).

The robot has a number of real-valued sensors $s_i(t)$ which are here summarized by the vector $S(t)$. Its actions are controlled by the setting of the real number values of a set of action/motor parameters $m_i(t)$, which we summarize as the vector $M(t)$. These action parameters can potentially be very low level (for example the speed of motors) or of a higher-level (for example the control parameters of motor primitives such as the biting or bashing movement that we will describe in the next section). We denote the sensorimotor context $SM(t)$ as the vector which summarizes the values of all the sensors and the action parameters at time t. In all that follows, there is an internal clock in the robot which discretized the time, and new actions are chosen at every time step.

Robots are equipped with a learning machine M which learns to predict $S(t+1)$ given $SM(t)$. This machine is going to be organized into a set of experts Exp_i, each being a specialist of a part of the sensorimotor space and responsible for the predictions corresponding to this part. This partition of the sensorimotor space will also be the basis of our local learning progress evaluation. Here, all these experts are nearest neighbours algorithms[2]. These experts store in memory all the associations between $SM(t)$ and the associated $S(t+1)$ that the robot encountered and which fits into their area of competence (a particular sensorimotor region). To anticipate the consequences of a new $SM(t)$, the closest examplar presented in memory is picked and the associated sensory consequence is used as a prediction.

At the beginning, there is only one expert Exp_1, responsible for the predictions of the whole sensorimotor space. Then, when a criterion C_1 is met, this expert is split into two experts. The criterion C_1 is the following: when the number of exemplars is above a threshold $T = 50$, then split. The next step is to decide how to split the initial expert. As we used nearest neighbours algorithms, spliting one expert into two just means spliting its set of exemplars into two sets, each of which defining the new expert. As explained in more details in (Oudeyer & Kaplan 2004), we split the set of examplars into two sets so that the sum of the variances of the point in the output space of each set, weighted by the number of examplars of each set, is minimal.

Each expert stores all the cutting dimensions and the cutting values that were used in its generation as well as in the generation of its parent experts. As a consequence when a prediction has to be done of the consequences of $SM(t)$, it is easy to find out which is the expert specialist for this case: it is the one for which $SM(t)$ satisfies all the cutting tests (and there is always a single expert which corresponds to each $SM(t)$).

Moreover, a prediction is made about all the actions that are actually executed. After the execution of these actions, the robot can measure the discrepancy between the sensory state $S(t+1)$ that it predicted and the actual sensory state $S_{actual}(t+1)$ that it measures. This provides an error of the prediction at time t:

$$E(t) = (S(t+1) - S_{actual}(t+1))^2$$

This error is stored by the expert which made the prediction. As a consequence, each expert keeps a list of its past errors:

$$E(t), E(t-1), E(t-2), ..., E(0)$$

Note that here t denotes a time which is specific to the expert, and not to the robot: this means that $E(t-1)$ might correspond to the error made by the expert in an action performed at $t - 10$ for the robot, and that no actions corresponding to this expert were performed by the robot since that time. This memory is then used by the expert to evaluate the expected decrease in the mean error rate in prediction (i.e. the learning progress) that a candidate action may provide. The method we use here is straightforward but revealed to be very efficient: the expected learning progress of every action in a given context which corresponds to the expert Exp_i is equal to the learning progress that has been

[2] We could very well use neural-networks, support vector machine, bayesian machines, etc.

achieved by this expert with the acquisition of its recent exemplars. More practically, the computation involves two steps:

- the mean error rate in prediction is computed at t and $t-\tau$:

  $E_{mean}(t) = mean(E(t), E(t-1), \ldots E(t-windowSize)$

  and

  $E_{mean}(t-\tau) = mean(E(t-\tau), \ldots E(t-\tau-windowSize))$

  where

  $$mean(E(i), \ldots, E(i-n)) = \frac{\sum_{j=i-n}^{i} E(j)}{n}$$

- the expected decrease in the mean error rate in prediction corresponding to a $SM(t)$ which fits the expert is defined as:

  $$DE(SM(t)) = E_{mean}(t) - E_{mean}(t-\tau)$$

  We can define the learning progress as:

  $$LP(SM(t)) = -DE(SM(t))$$

Eventually, when an expert is split into two experts, both new experts inherit the list of past errors from their parent expert, which allows them to make evaluation of potential learning progress right from the time of their creation.

Having explained this prediction machinery, the action-selection mechanism is straightforward:

- in a given sensorimotor context, which specifies the current values of the sensors and of the actuators, the robot makes a list of the possible actions which it can do; If this list is infinite, which is often the case since we work in continuous action spaces, a sample of candidate actions is generated;
- each of these candidate actions associated with the context makes a $SM(t)$ vector for which the robot finds out the corresponding expert; then this expert is used to provide an evaluation of the expected learning progress that might be the result of executing the candidate action (in addition to the sensory consequence of course);
- the action for which an expert expects the maximal learning progress is chosen and executed except in some cases when a random action is selected (the choice of a random action is tuned by a probability parameter P_{random} which is typically 0.2).

The Playground Experiment: the discovery of sensorimotor affordances

In a previous paper ((Oudeyer & Kaplan 2004)), we presented an implementation of this system in a simulated robot. We showed how IAC could allow the robot to develop in a noisy inhomogeneous environment, without being trapped by noise or the alternation between very unpredictable and very predictable situations. However, this experiment was in a simulated environment, and its complexity was limited.

Figure 1: The Playground Experiment

In this paper, we present a new experimental setup, called The Playground Experiment. This involves a physical robot as well as a more complex sensorimotor system and environment. We use a Sony AIBO robot which is put on a baby playmat with one toy that it can bite, and one toy that it can bash. (see figure 1). The environment is very similar to the ones in which two or three month old children learn their first sensorimotor skills. We have developped a web site which presents pictures and videos of this setup: http://playground.csl.sony.fr.

The robot is equipped initially only with three motor primitives: turning the head, bashing and crouch biting. Each of them is controlled by a number of real number parameters, which are the action parameters that the robot controls. The "turning head" primitive is controlled by the pan and tilt parameters of the robot's head. The "bashing" primitive is controlled by the strength and the angle of the leg movement. The "crouch biting" primitive is controlled by the depth of crouching (and the robot crouches in the direction in which it is looking at, which is determined by the pan and tilt parameters). Finally, the robot can combine the "turning head" primitive with bashing and biting, but bashing and biting cannot be combined. To summarize, a robot choosing an action has to choose a set of values for these 5 parameters (pan, tilt, bash strength, bash angle, crouch depth). All values are normalized between 0 and 1. When the bashing or the biting is not used, the values are set to -1.

The robot is equipped with three high-level sensors. There is one object visual detection sensor (Ov): it takes the value 1 when the robot sees one object, and 0 in the other case. In the playground, we use simple visual tags that we stick on the toys and are easy to detect from the image processing point of view. There is also a biting sensor (Bi): it takes the value 1 when the robot has something in its mouth and 0 otherwise. We use the cheek sensor of the AIBO. Finally, there is the oscillation sensor (Os): it takes the value 1 when the robot detects that there is something oscillating in front of it, and 0 otherwise. We use the infra-red distance sensor of the AIBO to implement this high-level sensor. This sensor can detect for example when there is an object that has been bashed in the direction of the robot's gaze, but can also detect events due to human walking around the playground (we do not control the environment).

It is crucial to note that initially the robot knows nothing about sensorimotor affordances. For example, it does not know that the values of the object visual detection sensor are correlated with the values of its pan and tilt. It does not know that the values of the biting or object oscillation sensors can become 1 only when biting or bashing actions are performed towards an object. It does not know that some objects are more prone to provoke positive values of the Bi and Os sensors when only certain kinds of actions are performed in their direction. It does not know for example that to get a positive value of the oscillation sensor, bashing in the correct direction is not enough, because it also needs to look in the right direction (since its oscillation sensors are on the front of its head). These remarks allow to understand easily that a random strategy will not be efficient in this environment. If the robot would do random action selection, in a vast majority of cases nothing would happen (especially for the Bi and Os sensors).

The robot is equipped with the Intelligent Adaptive Curiosity system, and thus chooses its actions according to the potential learning progress that it can provide to one of its experts. In this experiment, the action perception loop takes about one second: when the robot chooses and executes an action, it waits that all its motor primitives have finished their execution, which lasts approximately one second, before choosing the next action. This allows the robot to make all the measures necessary for determining adequate values of $(Ov,\ Bi,\ Os)$.

Results

During an experiment we continuously measure a number of features which help us to characterize the dynamics of the robot's development. First, we measure the frequency of the different kinds of actions that the robot does in a given time window. More precisely, every 100 actions and in the last 100 actions we measure: 1) the percentage of actions which do not involve the biting and the bashing motor primitive (i.e. the robot's action boils down to just looking in a given direction); 2) the percentage of actions which involve the biting motor primitive; 3) the percentage of actions which involve the bashing motor primitive. Second, we measure the distribution of values in each of the three sensory channels Ov, Bi and OS, every 100 actions and during the last 100 actions and we normalize these values by the distribution of the corresponding values in the case of random action selection. We normalize with the corresponding values of the random action selection method in order to show more clearly that some interesting and complex behaviours which are extremely rare with random action selection may become quite frequent when using Intelligent Adaptive Curiosity.

We will now show details of an example for a typical run of the experiment. All the curves corresponding to the measures we described are in figure 2. From the careful study of these curves, augmented with the study of the 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

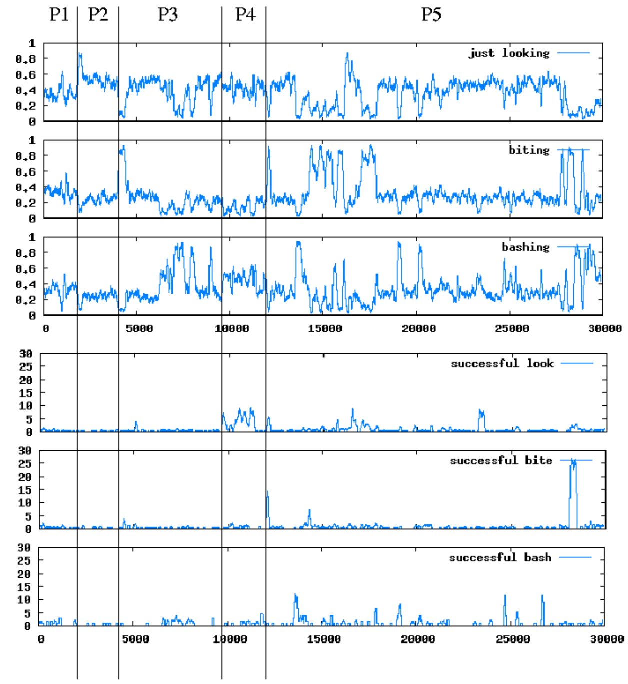

Figure 2: Top 3: Frequency for certain action types on windows 100 time steps wide. Top: bashing. Centre: Biting. Bottom: Just looking. Bottom 3: Distribution of values of the three sensors (Ov, Bi, Os) on windows 100 time steps wide, normalised regarding to the distribution of values in the case of random action selection.

than two phases of increasing behavioural complexity, i.e. we observe the emergence of several successive levels of behaviour patterns. Here are the different phases[3], which are visually denoted on the figure 2:

Phase 1: the robot has a short initial phase of random exploration and body babbling. This is very logical because during this period there are very few experts yet and so the sensorimotor space has not yet been partitioned in significantly different areas;

Phase 2: the robot stops using the biting and bashing primitives, and spends most of its time looking around. It has discovered that at this stage of its development, this kind of action is the greatest source of learning progress. The study of the curve measuring the frequency of 1 values in the Ov sensor shows that it does not see objects very often: it is in fact spending time learning that in many areas of the space there are no objects;

Phase 3: then there is a phase during which the robots begins to use a lot the biting and bashing primitives. It discovers that using these primitives sometimes produces something. Yet, the curve measuring the frequency of 1

[3] The organization of the developmental trajectory of the robot into 5 phases is of course created by us in a manner which simplifies the explanation of what is happening.

values in the Ov sensor as well as the close inspection of the traces of the experiment shows again that the robot is not oriented very often towards objects: this means that it has not discovered yet the fact that there is a relation both among the motor primitive (e.g. looking in the same direction as the movement of the bashing) and among action primitives and external objects (e.g. that biting or bashing can produce a result only if applied to an object);

Phase 4: then the robot discovers a new niche of learning progress at this stage of its development: it now starts to look often towards objects, as shown by the Ov curve. Yet, it is now half of the time stopping its action, and the rest of the time often bashing and sometimes biting, but with no specific association between the type of action (biting or bashing) and the objects towards which it is directed (the bitable or the bashable object). This means that the robot is here learning the precise location of objects as well as the fact that doing "something" towards an object can sometimes produce a reaction on the object and on its sensors.

Phase 5: Finally, the robot comes into a phase in which it discovers the precise affordances between action types and particular objects: it is now sometimes focusing either on trying to bite the bitable object, and on trying to bash the bashable object, as we can deduce from the curves showing the frequency of 1 values in the Bi and Os sensors. It is striking to note that during this phase, there are periods of time during which these coordinated motor primitives towards the right associated objects are 30 times more frequent as compared to the frequency of these situations in the random action selection case. Furthermore, it does actually manage to bite and bash successfully quite often, which of course is an emergent side effect of Intelligent Adaptive Curiosity and was not a preprogrammed task.

We made several experiments and each time we got a similar structure in which a self-organized developmental sequence pushed the robot towards activities of increasing complexity, in particular towards the progressive discovery of the sensorimotor affordances of various levels of detail. Nevertheless, we also observed that two developmental sequences are never exactly the same, and the number of phases sometimes changes a bit or intermediary phases are sometimes exchanged. It is interesting to note that this is also true for children: for example, some of them learn to crawl before they can sit, and vice versa. We are now trying to make statistical measures about the set of developmental sequences that are generated in our experiments in order to understand better how particular environment and embodiment conditions lead to the formation of recurrent developmental stages.

Acknowledgements

We would like to thank Jean-Christophe Baillie for its URBI systems which we used to program the robot and Luc Steels for precious comments on this work. This research has been partially supported by the ECAGENTS project founded by the Future and Emerging Technologies programme (IST-FET) of the European Community under EU R&D contract IST-2003-1940.

References

Asada, M.; Noda, S.; Tawaratsumida, S.; and Hosoda, K. 1996. Purposive behavior acquisition on a real robot by vision-based reinforcement learning. *Machine Learning* 23:279–303.

Barto, A.; Singh, S.; and Chentanez, N. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the International Conference on Developmental Learning*.

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

Herrmann, J.; Pawelzik, K.; and Geisel, T. 2000. Learning predicitve representations. *Neurocomputing* 32-33:785–791.

Huang, X., and Weng, J. 2002. Novelty and reinforcement learning in the value system of developmental robots. In *Proceedings of the 2nd international workshop on Epigenetic Robotics - Lund University Cognitive Studies 94*, 47–55.

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

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

Marshall, J.; Blank, D.; and Meeden, L. 2004. An emergent framework for self-motivation in developmental robotics. In *Internation Conference on Development and Learning*.

Oudeyer, P.-Y., and Kaplan, F. 2004. Intelligent adaptive curiosity: a source of self-development. In Berthouze, L.; Kozima, H.; Prince, C. G.; Sandini, G.; Stojanov, G.; Metta, G.; and Balkenius, C., eds., *Proceedings of the 4th International Workshop on Epigenetic Robotics*, volume 117, 127–130. Lund University Cognitive Studies.

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

Thrun, S. 1995. Exploration in active learning. In Arbib, M., ed., *Handbook of Brain Science and Neural Networks*. MIT Press.

Weng, J. 2002. A theory for mentally developing robots. In *Second International Conference on Development and Learning*. IEEE Computer Society Press.