

Intrinsically motivated machines

Frederic Kaplan, Pierre-Yves Oudeyer

▶ **To cite this version:**

Frederic Kaplan, Pierre-Yves Oudeyer. Intrinsically motivated machines. Lungarella, M.; Iida, F.; Bongard, J.; Pfeifer, R. 50 Years of AI, Lungarella, M.; Iida, F.; Bongard, J.; Pfeifer, R., pp.n/a, 2008. inria-00420223

HAL Id: inria-00420223
https://inria.hal.science/inria-00420223

Submitted on 28 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Intrinsically motivated machines

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

(1) Ecole Polytechnique Federale de Lausanne - CRAFT
CE 1628 - Station 1 - CH 1015 Lausanne Switzerland - email: frederic.kaplan@epfl.ch
(2) Sony Computer Science Laboratory Paris
6 rue Amyot 75005 Paris France email: py@csl.sony.fr

Abstract. Children seem intrinsically motivated to manipulate, to explore, to test, to learn and they look for activities and situations that provide such learning opportunities. Inspired by research in developmental psychology and neuroscience, some researchers have started to address the problem of designing intrinsic motivation systems. A robot controlled by such systems is able to autonomously explore its environment not to fulfil predefined tasks but driven by an incentive to search for situations where learning happens efficiently. In this paper, we present the origins of these intrinsically motivated machines, our own research in this novel field and we argue that intrinsic motivation might be a crucial step towards machines capable of life-long learning and open-ended development.

Keywords: Intrinsic motivation, curiosity, exploration, meta-learning, development

1 Introduction

Have you ever noticed how much fun babies can have by simply touching objects, sticking them into their mouths, or rattling them and discovering new noises? Children seem to engage is such type of activities just for the sake of it. They seem intrinsically motivated to manipulate, to explore, to test - in one word - to learn and therefore they look for activities and situations that provide such learning opportunities. More than 50 years ago, Alan Turing prophetically announced that the childs mind would show us the way to artificial intelligence."Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?" [Turing, 1950]. We believe it is now time to take this advice seriously. Through hundreds of experiments and models - supervised, unsupervised, reinforced, active, passive, associative, symbolic, connectionist, hybrid, embodied, situated, distributed - we benefit now from a large collection of examples that show how a machine can learn. However, the issue of "why" would a machine learn (or how would it choose what to learn) has not been tackled with the same attention. This is what interests us here.

During the past five years, we have been working on algorithms that make robots eager to investigate their surroundings. These robots explore their environment in search of new things to learn: they get bored with situations that

are already familiar to them, and also avoid situations which are too difficult. In our experiments, we place the robots in a world that is rich in learning opportunities and then just watch how the robots develop by themselves. This research is based on a series of studies showing the importance of *intrinsic motivation* in human development and its neural correlates in the brain. The next sections give a general overview of these findings and discuss the origins of intrinsically motivated machines in artificial intelligence research. We then present our own research in this field though the discussion of a specific architecture and related robotic experiments.

2 What is intrinsic motivation?

In psychology, an activity is characterized as intrinsically motivated when there is no apparent reward except the activity itself. People seek and engage in such activities for their own sake and not because they lead to extrinsic reward. In such cases, the person seems to derive enjoyment directly from the practice of the activity. Following this definition, most children playful or explorative activities can be characterized as being intrinsically motivated. Also, much adult behaviour seem to belong to this category: free problem-solving (solving puzzles, crosswords), creative activities (painting, singing, writing during leisure time), gardening, hiking, etc. At the physiological level, it has been argued that intrinsically motivated activities are directly related to changes in the central nervous system and are quite independant from non-nervous tissues. On the contrary, extrinsic needs (e.g. hunger) are directly related to the state and management of non-nervous-systems tissues [Deci and Ryan, 1985]. Moreover, intrinsically motivated activities are generic in the sense that they can be produced by different kinds of sensory contexts. Finally, at a phenomenological level , a person engages in intrinsically motivating activities to experience particular feelings of competence and self-determination [Deci and Ryan, 1985]. Such situations are characterized by a feeling of effortless control, concentration, enjoyment and a contraction of the sense of time [Csikszenthmihalyi, 1991].

3 Intrinsic motivation in psychology and neuroscience

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

Moreover, exploration does not seem to be related to any non-nervous-system tissue deficit.

Some researchers then proposed another conceptualization. Festinger's theory of cognitive dissonance [Festinger, 1957] asserted that organisms are motivated to reduce dissonance, that is the incompatibility between internal cognitive structures and the situations currently perceived. Fifteen years later a related view was articulated by Kagan stating that a primary motivation for human is the reduction of uncertainty in the sense of the "incompatibility between (two or more) cognitive structures, between cognitive structure and experience, or between structures and behaviour" [Kagan, 1972]. However, these theories were criticized on the basis that much human behaviour is also intended to *increase* uncertainty, and not only to reduce it.

Human seem to look for some forms of optimality between completely uncertain and completely certain situations. In 1965, Hunt developed the idea that children and adult look for optimal incongruity [Hunt, 1965]. He regarded children as information-processing systems and stated that interesting stimuli were those where there was a discrepancy between the perceived and standard levels of the stimuli. For, Dember and Earl, the incongruity or discrepancy in intrinsically-motivated behaviours was between a person's expectations and the properties of the stimulus [Dember and Earl, 1957]. Berlyne developed similar notions as he observed that the most rewarding situations were those with an intermediate level of novelty, between already familiar and completely new situations [Berlyne, 1960].

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

Eventually, a last group of researchers preferred the concept of challenge to the notion of optimal incongruity. These researchers stated that what was driving human behaviour was a motivation for effectance [White, 1959], personal causation [De Charms, 1968], competence and self-determination [Deci and Ryan, 1985]. The difference with optimality theories is mainly a matter of point of view: in one case, human search for some form of optimality as defined by an abstract function, in the other case they look for a particular kind of feelings occurring during challenging situations.

Novel investigations in neuroscience concerning neuromodulation systems have complemented these findings. Although most experiments in this domain focus on the involvement of particular neuromodulators like dopamine for predicting extrinsic reward (e.g. food), some work lends credence to the idea that such neuromodulators might also be involved in the processing of types of intrinsic motivation associated with novelty and exploration (e.g. [Dayan and Belleine, 2002] and [Kakade and Dayan, 2002]). In particular, some studies suggest that dopamine

responses could be interpreted as reporting "prediction error" and not only "reward prediction error" [Horvitz, 2000]. At a more global level, Panksepp has compiled a set of evidence suggesting the existence of a SEEKING system responsible for exploratory behaviours. "This harmoniously operating neuroemotional system drives and energizes many mental complexities that humans experience as persistent feelings of interest, curiosity, sensation seeking and, in the presence of a sufficiently complex cortex, the search for higher meaning." [Panksepp, 1998]. However, the gap is still important between neuroscience accounts and research in psychology on intrinsic motivation.

4 The route to intrinsically motivated machines

During the last ten years, the machine learning and robotics community has begun to investigate architectures that permit incremental and active learning (see for instance [Thrun and Pratt, 1998] or [Cohn et al., 1996]). Interestingly, the mechanisms developed in these papers have strong similarities with mechanisms developed in the field of statistics, where it is called "optimal experiment design" [Fedorov, 1972]. Active learners (or machines that perform optimal experiments) are machines that ask, search and select specific training examples in order to learn efficiently.

A few researchers have started to address the problem of designing intrinsic motivation systems to drive active learning, inspired by research in developmental psychology and neuroscience. The idea is that a robot controlled by such systems would be able to autonomously explore its environment not to fulfil predefined tasks but driven by some form of intrinsic motivation that pushes it to search for situations where learning happens efficiently. One of the first computational system implementing a form of artificial curiosity was described by Schmidhuber in 1991 [Schmidhuber, 1991]. Schmidhuber articulated the idea that in order to learn efficiently a machine should try to reduce prediction error instead of maximizing or minimizing it. More recently, different types of intrinsic motivation systems were explored, mostly in software simulations [Huang and Weng, 2002,Marshall et al., 2004,Steels, 2004]. Most of this research has largely ignored the history of the intrinsic motivation construct as it was elaborated in psychology during the last 50 years and sometimes reinvented concepts that existed several decades before (basically, different forms of optimal incongruity). Technically, such control systems can be viewed as particular types of reinforcement learning architectures [Sutton and Barto, 1998], where rewards are not provided externally by the experimenter but self-generated by the machine itself. The term "intrinsically motivated reinforcement learning" has been used in this context [Barto et al., 2004].

5 Intrinsic motivation and development

Our own research in this field aims at showing how such forms of active learning architectures permit to structure the development of a robot. For this issue,

Piaget remains a reference. Although he has not written extensively about motivation, his view of motivation is implicit throughout his writings [Piaget, 1952]. According to Piaget, children are intrinsically motivated to encounter activities which involve some assimilation and accommodation. Assimilation is a process whereby children incorporate aspects of the environment into their pre-existing cognitive structure, which are called schemata. This means that the child's cognitive structure influences his perception of the environment. During accommodation, the child adapts his cognitive structures to fit the environment. In Piaget's view, learning is simply an aspect of assimilation and accommodation. When the child encounters an informational input from the environment highly discrepant from existing schemata, the input will most likely be ignored. When inputs are completely predictable, children generally lose interest in them. Therefore, Piaget's theory articulates a concept of intrinsic simulation that bears many resemblance to the notion of optimal incongruity developed by Berlyne and others.

Piaget's work has inspired researchers in artificial intelligence for some time. Many artificial intelligence models make use of internal explicit schema structures under names like *frames* [Minsky, 1975] or *scripts* [Schank and Abelson, 1977]. In such systems, there is a one-to-one mapping between these internal structures and the functional operation that the agent can perform. For instance, Drescher describes a system inspired by Piaget's theories in which a developing agent explicitly creates, modifies and merges schema structures in order to interact with a simple simulated environment [Drescher, 1991]. Using explicit schema structures has several advantages: such structures can be manipulated via symbolic operations, creation of new skills can be easily monitored by following the creation of new schemata, etc. Our work differs notably from this approach first because it does not rely on such explicit representations. We use subsymbolic systems, based on continuous representations of their environment. Nevertheless, as we will see, such systems may display some organized forms of behaviour where clear functional units can be identified. Second and most importantly, it is centrally based the idea of an intrinsic motivation to learn.

6 An example of architecture

We have designed a control architecture and performed a series of experiments to investigate how far its intrinsic motivation system, implementing a form of artificial curiosity, can shape the developmental trajectories of a robot [Oudeyer et al., 2007]. The cognitive architecture of our robot can be described as having two modules: 1) one module implements a predictor M which learns to predict the sensorimotor consequences when a given action is executed in a given sensorimotor context; 2) another module is a metapredictor $metaM$ which learns to predict the errors that machine M makes in its predictions: these meta-predictions are then used as the basis of a measure of the potential interest of a given situation. The system is designed to be *progress-driven*. It avoids both predictable and unpredictable situations in order to focus on the ones which are expected to maximize the decrease in prediction error. To obtain such a behaviour, the

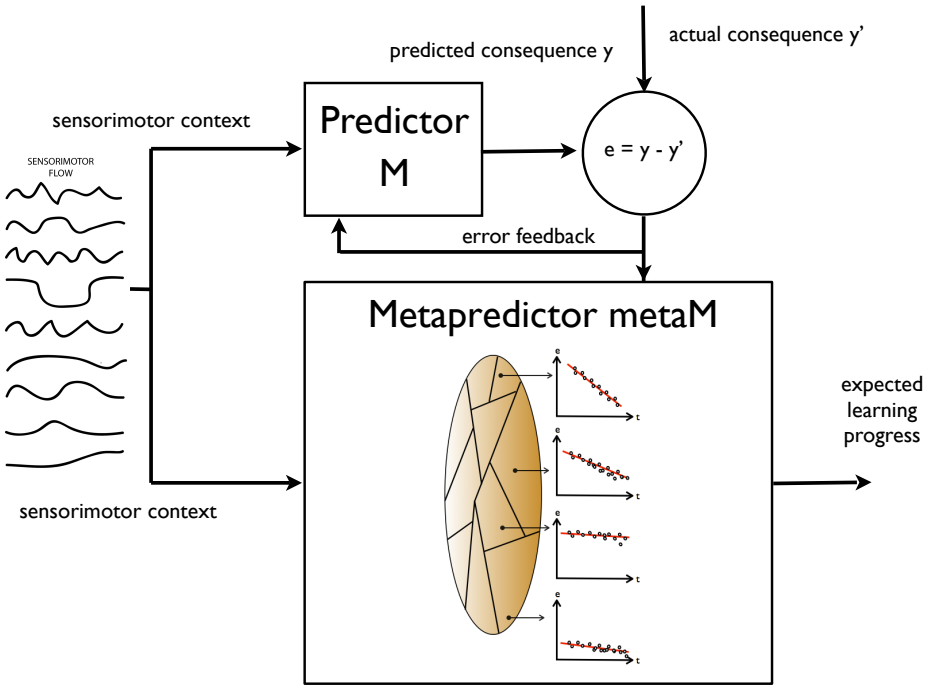


Fig. 1. An intrinsic motivation system including a predictor M that learns to anticipate the consequence $\mathbf{y}$ of a given sensorimotor context and a metapredictor $metaM$ learning to predict the expected learning progress of M in the same context. Once the actual consequence is known, M and $metaM$ get updated. $MetaM$ re-evaluates the error curve linked with this context and computes an updated measure of the learning progress (local derivative of curve). In order to classify similar contexts, $metaM$ includes a hierarchical self-organizing classifier.

metaprediction system computes the local derivative of the error rate curve of M and generates an estimation of the expected learning progress linked with a particular action in a particular context. In order to really evaluate learning progress, error obtained in one context must be compared with errors obtained in similar contexts (if not the robot may oscillate between hard and easy situations and evaluate these changes as progress). Therefore, the metaprediction system must also be equipped with a self-organized classification system capable of structuring an infinite continuous space of particular situations into higher-level categories (or kinds) of situations. Figure 1 summarizes the key components of such progress-driven systems (see [Oudeyer et al., 2007] for more details).

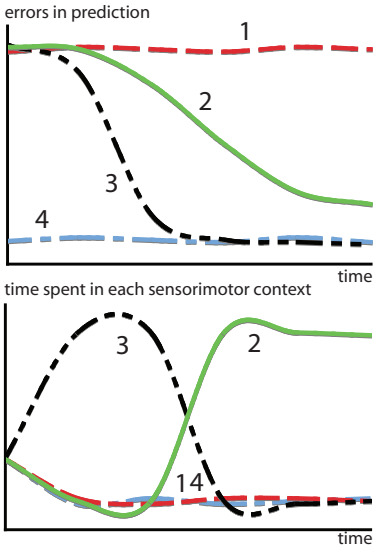


Fig. 2. Confronted with four contexts characterized by different learning profiles, the motivation for maximizing learning progress results in avoiding situations already predictable (context 4) or too difficult to predict (context 1), in order to focus first on the context with the fastest learning curve (context 3) and eventually, when the latter starts to reach a "plateau" to switch to the second most promising learning situation (context 2). This intrinsic motivation system allows the creation of an organized exploratory strategy.

Figure 2 illustrates how progress-driven learning operates on an idealized problem. Confronted with four contexts characterized by different learning profiles, the motivation for maximizing learning progress results in avoiding situations that are already predictable (context 4) or too difficult to predict (context

1), in order to focus first on the context with the fastest learning curve (context 3) and eventually, when the latter starts to reach a "plateau", to switch the second most promising learning situation (context 2). Situations of maximal progress are called "progress niches". Progress niches are not intrinsic properties of the environment. They result from a relationship between a particular environment, a particular embodiment (sensors, actuators, feature detectors and techniques used by the prediction algorithms) and a particular time in the developmental history of the agent. Once discovered, progress niches progressively disappear as they become more predictable.

7 Experiments

We have performed a series of robotic experiments using this architecture. In these experiments, the robot actively seeks out sensorimotor contexts it can learn given its morphological and cognitive constraints. Whereas a passive strategy would lead to very inefficient learning, an active strategy allows the learner to discover and exploit learning situations fitted to its biases. In one of those experiment a four-legged robot is placed on a play mat (for more details, see [Oudeyer and Kaplan, 2006]). The robot can move its arms, its neck and mouth and can produce sounds. Various toys are placed near the robot, as well as a pre-programmed "adult" robot which can respond vocally to the other robot in certain conditions. At the beginning of an experiment, the robot does not know anything about the structure of its sensorimotor space (which actions cause which effects). Given the size of the space, exhaustive exploration would take a very long time and random exploration would be inefficient.

During each robotic experiment, which lasts approximately half a day, the flow of values of the sensorimotor channels are stored, as well as a number of features which help us to characterize the dynamics of the robot's development. The evolution of the relative frequency of the use of the different actuators is measured: the head pan/tilt, the arm, the mouth and the sound speakers (used for vocalizing), as well as the direction in which the robot is turning its head.

Figure 3 shows data obtained during a typical run of the experiment. At the beginning of the experiment, the robot has a short initial phase of random exploration and body babbling. During this stage, the robot's behaviour 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 at or act on objects; it essentially does not interact with the environment. Then there is a phase during which the robot begins to focus successively on playing with individual actuators, but without knowing the appropriate 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 on biting, and finally on bashing in all directions (and stops biting and vocalizing). Then, the robot comes to a phase in which it discovers the precise affordances between certain action types and certain particular objects. It is at

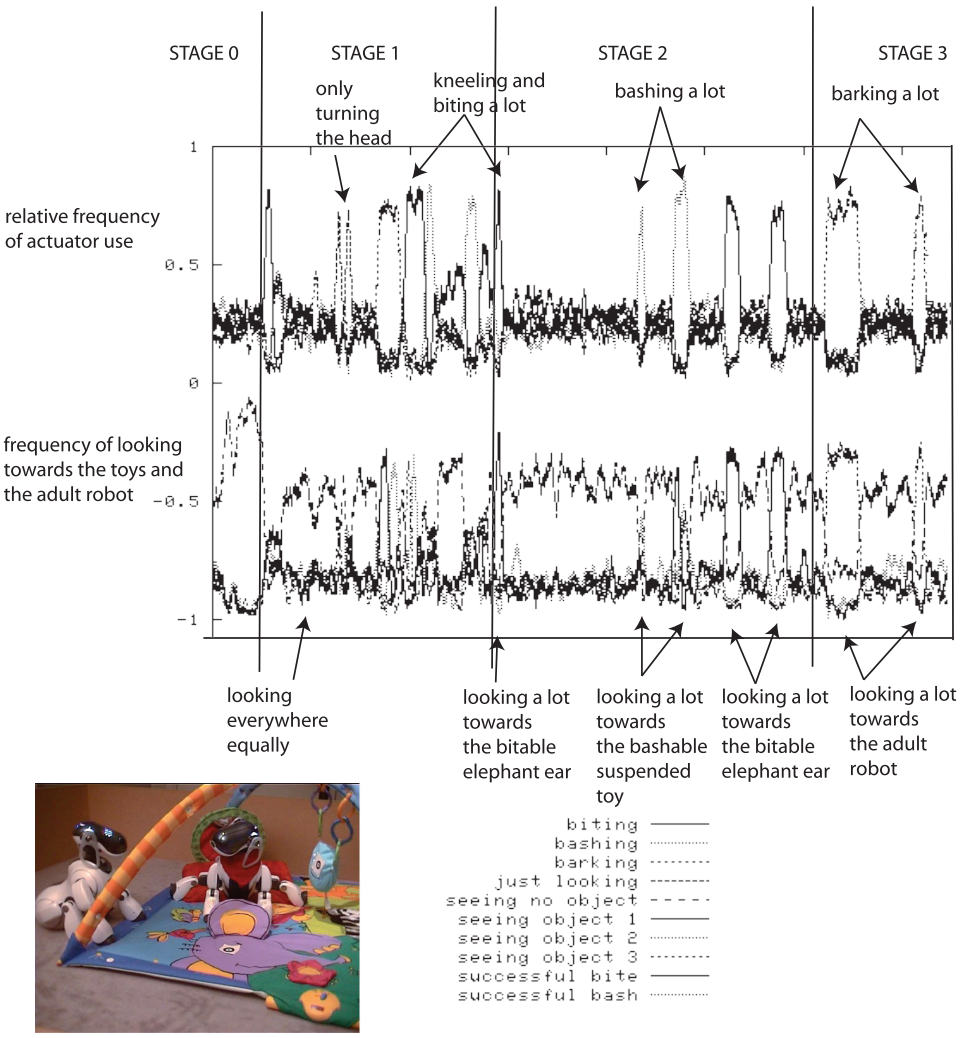


Fig. 3. The robot, placed on a play mat, can move its arms, its neck and mouth and produce sounds. Various toys are placed near the robot, as well as a pre-programmed "adult" robot which can respond vocally to the other robot in certain conditions. Results obtained after a typical run of the experiment are shown. **Top curves**: relative frequency of the use of different actuators (head pan/tilt, arm, mouth, sound speaker). **Bottom curves**: frequency of looking towards each object and in particular towards the "adult" pre-programmed robot

this point focusing either on trying to bite the biteable object (the elephant ear), or on trying to bash the bashable object (the suspended toy). Eventually, it focuses on vocalizing towards the "adult" robot and listens to the vocal imitations that it triggers. 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 developmental trajectories produced by these experiments can be interpreted as assimilation and accommodation phases if we retain the Piagetian's terminology. For instance, the robot "discovers" the biting and bashing schema by producing repeated sequences of these kinds of behaviour, but initially these actions are not systematically oriented towards the biteable or the bashable object. This stage corresponds to "assimilation". It is only later that "accommodation" occurs as biting and bashing starts to be associated with their respective appropriate context of use. Our experiments show that functional organization can emerge even in the absence of explicit internal schema structures and that developmental patterns can spontaneously self-organize, driven by the intrinsic motivation system. Many diverse lines of experimental data can potentially be explained in common terms if we consider that children learn how to focus on what is learnable in the situation they encounter and on what can be efficiently grasped at a given stage of their cognitive and physiological development. For instance, we have discussed elsewhere how progress-driven learning provides an interpretation of developmental sequences in early imitation and sensorimotor development [Kaplan and Oudeyer, 2007a,Kaplan and Oudeyer, 2007b]. What is fundamentally new in these experiments, as compared to what is possible in psychology, is that learning dynamics, embodiment and environmental factors (both social and physical) are *controllable variables*. One experiment can be conducted with the same learning system, but using a different body placed in a different environment. Likewise, the effects of small changes in the intrinsic motivation systems can be studied while keeping the embodiment and environmental aspects similar.

8 The future of intrinsically motivated machines

To conclude, this novel line of research might also provide radically new techniques for building intelligent robots. Indeed, as opposed to the work in classical artificial intelligence in which engineers impose pre-defined anthropocentric tasks to robots, the techniques we describe endow the robots with the capacity of deciding by themselves which are the activities that are maximally fitted to their current capabilities. Intrinsically motivated machines autonomously and actively choose their learning situations, thus beginning by simple ones and progressively increasing their complexity. Of course, many challenges remain to be solved before we could build intrinsically motivated machines capable to learn like children do. One of them is that children's complex behaviour patterns seem hierarchically organized. This aspect is absent from our current architecture but have started to be tackled by other groups, in particular around the

option framework [Sutton et al., 1999,Barto et al., 2004]. This research is very complementary to ours as they experimented the use of a complex reinforcement technique given a simple novelty-based intrinsic motivation system. We believe the future of intrinsically motivated machines lies somewhere between of these two approaches.

References

[Barto et al., 2004] Barto, A., Singh, S., and Chentanez, N. (2004). Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, Salk Institute, San Diego.

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

[Cohn et al., 1996] Cohn, D., Ghahramani, Z., and Jordan, M. (1996). Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145.

[Csikszenthmihalyi, 1991] Csikszenthmihalyi, M. (1991). *Flow-the psychology of optimal experience*. Harper Perennial.

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

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

[Deci and Ryan, 1985] Deci, E. and Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. Plenum Press.

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

[Drescher, 1991] Drescher, G. L. (1991). *Made-up minds*. The MIT Press, Cambridge, MA.

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

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

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

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

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

[Huang and Weng, 2002] 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, 1943] Hull, C. L. (1943). *Principles of behavior: an introduction to behavior theory*. New-York: Appleton-Century-Croft.

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

[Kagan, 1972] Kagan, J. (1972). Motives and development. *Journal of Personality and Social Psychology*, 22:51–66.

[Kakade and Dayan, 2002] Kakade, S. and Dayan, P. (2002). Dopamine: Generalization and bonuses. *Neural Networks*, 15:549–559.

[Kaplan and Oudeyer, 2007a] Kaplan, F. and Oudeyer, P.-Y. (2007a). 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.

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

[Marshall et al., 2004] Marshall, J., Blank, D., and Meeden, L. (2004). An emergent framework for self-motivation in developmental robotics. In *Proceedings of the 3rd International Conference on Development and Learning (ICDL 2004)*, Salk Institute, San Diego.

[Minsky, 1975] Minsky, M. (1975). A framework for representing knowledge. In Wiston, P., editor, *The psychology of computer vision*, pages 211–277. Mc Graw Hill, New York.

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

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

[Oudeyer et al., 2007] Oudeyer, P.-Y., Kaplan, F., and Hafner, V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(1):265–286.

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

[Piaget, 1952] Piaget, J. (1952). *The origins of intelligence in children*. Norton, New York, NY.

[Schank and Abelson, 1977] Schank, R. and Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum Associates, Hillsdale, NJ.

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

[Steels, 2004] Steels, L. (2004). The autotelic principle. In Fumiya, I., Pfeifer, R., Steels, L., and Kunyoshi, K., editors, *Embodied Artificial Intelligence*, volume 3139 of *Lecture Notes in AI*, pages 231–242. Springer Verlag, Berlin.

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

[Sutton et al., 1999] 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.

[Thrun and Pratt, 1998] Thrun, S. and Pratt, L. (1998). *Learning to Learn*. Kluwer Academic Publishers.

[Turing, 1950] Turing, A. (1950). Computing machinery and intelligence. *Mind*, 59:433–460.

[White, 1959] White, R. (1959). Motivation reconsidered: The concept of competence. *Psychological review*, 66:297–333.