The Discovery of Communication

Pierre-Yves Oudeyer and Frédéric Kaplan
{oudeyer, kaplan}@csl.sony.fr
Sony Computer Science Laboratory,
6, rue Amyot,
75005 Paris

Abstract

What kind of motivation drives child language development? This article presents a computational model and a robotic experiment to articulate the hypothesis that children discover communication as a result of exploring and playing with their environment. The considered robotic agent is intrinsically motivated towards situations in which it optimally progresses in learning. To experience optimal learning progress, it must avoid situations already familiar but also situations where nothing can be learned. The robot is placed in an environment in which both communicating and non-communicating objects are present. As a consequence of its intrinsic motivation, the robot explores this environment in an organized manner focusing first on non-communicative activities and then discovering the learning potential of certain types of interactive behavior. In this experiment, the agent ends up being interested by communication through vocal interactions without having a specific drive for communication.

Keywords: development, robotics, communication, intrinsic motivation, vocalizations, stages

1 Introduction

What drives child language development? Why does the child spend so much efforts learning the complex know-how necessary to master language? These questions have led to multiple controversial debates. First, it should be stressed that these questions are not always relevant for certain approaches of language acquisition. Certain theories view language acquisition as primarily a passive automatic process. Children learn language simply by being exposed to it. With such explanations, the motivational aspect of language learning is secondary. Passive theories of language acquisition have been exposed to many criticisms arguing that learning language passively (e.g. by statistical induction) is too hard a problem (Gold, 1967). In general, these have led to strong nativist theories supposing that language can actually be learned passively because a language-specific learning machinery (the Language Acquisition Device) is already available at birth (Pinker and Bloom, 1990).

In contrast with passive approaches to language acquisition, another set of explanations argue that language acquisition is fundamentally an active process. Children pay attention to linguistic events and practice skills for expressing themselves because they are motivated to do so. As they progress, they get interested in different types of linguistic skills. Therefore, they learn to master the complexity of language in an incremental progressive manner. However, to be convincing, such explanations must clarify the drives that underlie this development. Is it a general incentive or a motivation specific to communication? Is it extrinsic or intrinsic? We will quickly review these different hypotheses.

Are children born with a specific motivation for social interaction or communication? Such a motivation would be innate, selected by natural selection. This hypothesis raises new kind of questioning at the phylogenetic level: What are the benefits of communication for humans? How can we explain the emergence of the skills underlying language from an evolutionary perspective? Contrary to what is often assumed, the evolutionary advantage of developing linguistic skills is far from being straightforward (Dessalles, 1998; Gintis et al., 2001). Therefore, making the hypothesis that an innate motivation for communication exists because it leads to a selective advantage is not unproblematic. Moreover, the hypothesis of an innate drive for communicating fails to explain why children develop increasingly complex communication skills. Children do not simply learn basic skills for interacting, they develop amazingly complex competence for expressing themselves and understanding others.

Do children communicate because they try to achieve other basic needs? In such a case, the development of communication skills would be the result of various types of extrinsic motivation like getting food or emotional comfort (Rolls, 1999). Again, this functional perspective fails to explain the open-ended nature of children's linguistic development. Children are not satisfied with their initial success in learning to talk. They continuously explore new dimensions of the language faculty and tackle new challenges: how to segment the speech stream into meaningful linguistic units, how to structure the world into linguistically-relevant events and objects, and how to handle the mapping between such perceptual structuring and the lexical and syntactic level. Can such sophistication emerge simply for fulfilling basic needs?

A third kind of explanations would be that children develop communication skills because it is inherently interesting for them. In such a case, their development would be driven by a form of intrinsic motivation (White, 1959; Deci and Ryan, 1985), not specific to communication but that would nevertheless lead to the development of communication skills. Play and curiosity-driven exploration would be at the origins of this development. The hypothesis that we would like to explore in this article belongs to this third kind.

Our approach is based on computational developmental models and robotics experiments. It is different but complementary with models of language formation studied so far. In the last ten years, a growing number of researchers have taken up the interdisciplinary challenge to understand language formation and evolution through computational and robotic models (see Cangelosi and Parisi (2002); Kirby (2002); Steels (2003) for general overviews of the field). For many researchers that participate in this collective effort, one objective is to show that general learning mechanisms could account for several aspects of language without having to suppose the existence of innate specific preadaptations. Successful experimental and theoretical results were obtained in the domains of lexicon formation (e.g. Hutchins and Hazlehurst (1995); Steels (1996); Shoham and Tennenholtz (1997); Ke et al. (2002); Smith et al. (2003); Kaplan (2005)), phonological evolution (De Boer, 2001; Oudeyer, 2005, 2006) and grammatical aspects of language acquisition (Steels and Neubauer, 2005). Other experiments stressed the role of situatedness and social learning for language acquisition using robotic experiments (e.g. (Steels and Kaplan, 2000)). Some experiments combined these two trends in unified set-ups by showing how a population of robotic agents could collectively negotiate a shared lexicon in which each word was grounded in private perceptual categories (e.g. (Steels and Kaplan, 1999)).

These models have allowed us to clarify important aspects of language dynamics, but they were based on two important underlying assumptions. In all these models, (1) communicative and non-communicative behaviors are separated from the start, and (2) agents are (either explicitly or implicitly) motivated for successful linguistic communication. As we have previously discussed, such assumptions raise some issues. Why would an agent be motivated for linguistic communication if it doesn't have the know-how for communicating yet? What would drive it towards developing appropriate skills?

A new family of models, based on the concept of intrinsic motivation, recently developed in the field of developmental robotics and may help us to explore these aspects of language development. Architectures using intrinsic motivation systems are particular types of reinforcement learning architectures in which rewards are provided not from external means but through internal evaluation (e.g. Schmidhuber (1991); Thrun (1995); Herrmann et al. (2000); Barto et al. (2004); Marshall et al. (2004); Kaplan and Oudeyer (2004); Oudeyer et al. (2005a); Schmidhuber (2006)). In these systems, some situations (e.g. particular sensorimotor contexts) are judged to be intrinsically more interesting than others by the system (for instance because they are novel or on the contrary familiar) and the agent seeks them out for that purpose. A major challenge for this research area is to design intrinsic motivation systems that would permit some form of open-ended task-independent organized development (Weng et al., 2001; Oudeyer et al., 2005a). Such a system should result in the autonomous organization of interesting developmental sequences.

In this article, we consider a robotic agent intrinsically motivated towards situations in which it progresses in learning. To experience progress, the agents must avoid situations already familiar but also situations where nothing can be learned. We have shown in another series of experiments how such a model can permit the organization of complex developmental sequences (e.g. Oudeyer and Kaplan (2004); Oudeyer et al. (2005a)). Here we report on an experiment in which the robot is placed in an environment in which both communicating and non-communicating objects are present. We will show that by looking for progress niches, situations in which learning progress is maximal, the robot explores this environment in an organized manner focusing first on non-communicative activities but then discovering the intrinsic richness of certain types of interactive behavior. In this experiment, the robot will end up being interested by communication and will focus its activity on vocal interactions without having a specific drive for communication.

The next section presents in detail the intrinsic motivation system that we use, and the following one describes an experimental setup and the results which are obtained. This system provides a metaphor to articulate the hypothesis that children may *discover communication* simply by being motivated by exploring and playing with their environment. This experiment should be seen as a "tool for thought" for this developmental scenario. Moreover, it is only a first step. In particular, the setup does not yet address issues related with the open-ended nature of linguistic development. As we will argue in the end of the article, many challenges are yet to be tackled to address convincingly this aspect in a robotic setup.

2 The intrinsic motivation system: Intelligent Adaptive Curiosity

In this part we will give a technical overview of the intrinsic motivation system that we use and of the cognitive architecture in which it is integrated. More details are available in some other papers (Oudeyer and Kaplan, 2004; Oudeyer et al., 2005a).

2.1 Curiosity as the search for learning progress

The intrinsic motivation system that we use is called **Intelligent Adaptive Curiosity**, which can be abbreviated IAC. Figure 3 represents the architecture in which it is integrated. It is based on a module which can compute the interestingness of all sensorimotor situations in terms of what we call the "learning progress". Learning progress is defined as the amount of progress in the quality of anticipations that the robot performs in similar situations. So it relies on the one hand on a system of categorization of the sensorimotor space which is able to group situations based on their similarities, and on the other hand on a system which is able to evaluate the evolution of performances in anticipations for each group of situations. An interesting feature of the system is that this grouping of situations in what we will call regions is *unsupervised* and *incremental*. This means that initially the robot considers that all situations belong to the same group (there is only one region), and progressively it will split this group/region into several groups/regions to refine its categorization, and so on. The robot is motivated to explore preferentially regions in which its performances in predicting the consequences of its actions increase maximally fast. The exploration of these regions leads the robot to gain experience into them, and when a certain amount of experience is reached in a region, then the region is split.

This mechanism implies that the robot is able to evaluate its own performances in anticipation in a given region of the sensorimotor space. This evaluation is based on the computation of its errors in predicting the sensorimotor flow. Indeed, the robot is equipped on the one hand with a set of sensors and motors, whose values evolve continuously, and on the other hand with a learning system which tries to predict the values of the sensors in the near future given the values of the sensors and of the motors in the close past. After each prediction has been made, the robot waits for the actual measure of the predicted sensors and computes its error in prediction. The smoothed evolution of this error in each of the sensorimotor regions is the basis of the evaluation

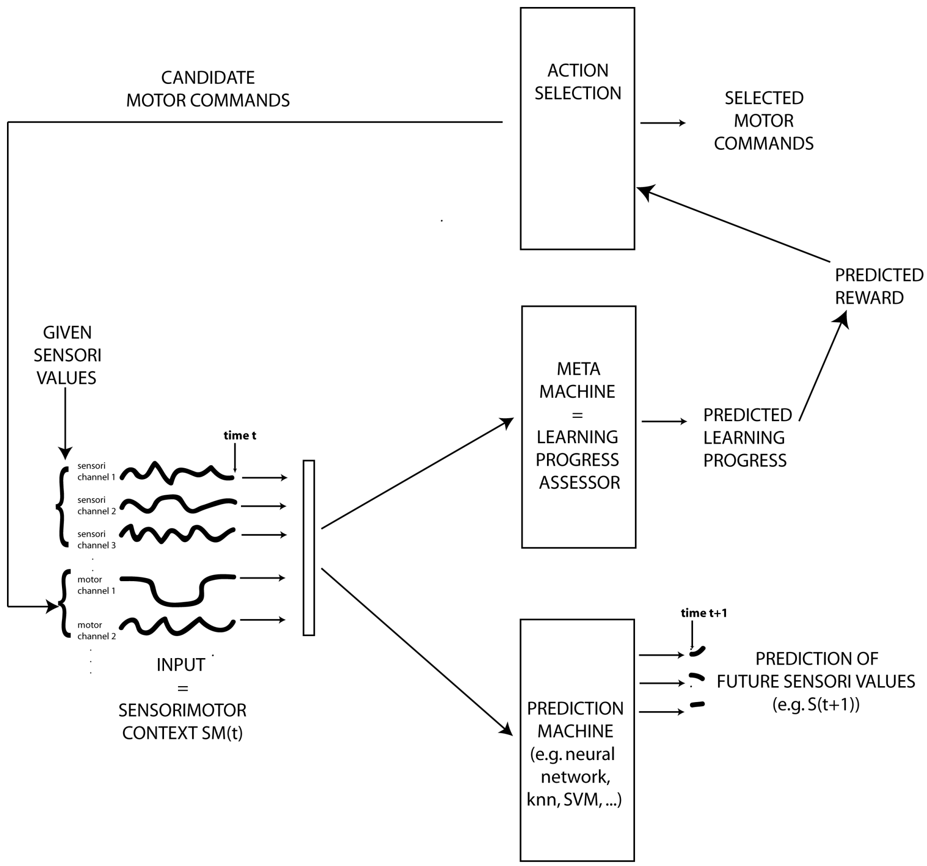

Figure 1: The intrinsic motivation based architecture

of learning progress: learning progress in a given region is defined as the inverse of the decrease of this smoothed error in the close past in this region.

Finally, at each time step the robot decides which values to send to its motors by measuring the current sensory context (which can include proprioceptive information), imagining possible sets of motor values, and for each of these imagined situations it evaluates the corresponding learning progress. Then, most of the time it chooses a set of motor values for which its evaluation of the associated learning progress is maximal, but sometimes it also chooses a random set of motor values, which helps it avoid being stuck in local minima.

It has to be noted that with this system, the interestingness of each situation will possibly evolve with time. Indeed, if for example a situation is not predictable initially, then becomes progressively more predictable thanks to the learning mechanisms, and then reaches a plateau of high predictability, this means for the robot that it will be interesting for a while because its errors in prediction decrease, but then in a second phase the situation will not be interesting anymore since the errors in prediction do not decrease anymore. As a consequence, the value function of the robot is highly non-stationary. This is what will allow the self-organization of behavioral and cognitive stages into a developmental sequence. Moreover, this allows the robot not to be interested by situations which are trivial or too easy from its point of view (where it is always very good at predicting) or by situations which are not learnable given its cognitive abilities (where its predictions will always be bad). This enables the robot to continuously control the complexity of its behavior so that it is increasing progressively. We will now describe in more details the Intelligent Adaptive Curiosity algorithm.

2.2 Technical details: computing similarity-based learning progress

Sensorimotor apparatus. The robot has a number of real-valued sensory channels $s_i(t)$ which are here summarized by the vector $\mathbf{S(t)}$. A sensory channel typically measures the continuous flow of values of a sensor, but it can also be a measure which aggregates the continuous flow of values of a set of sensors. The robot has also a set of motor channels which it can set in order to control its actions. The motor channels take real number values denoted $m_i(t)$, which we summarize using the vector $\mathbf{M(t)}$. A motor channel typically sets the value of a physical motor, but it can also set a value which is used by a low-level controller which itself controls a set of motors and actuators (such as for example the angle of arm movement in the experimental setup which we will describe in the following section). We denote the sensorimotor context $\mathbf{SM(t)}$ as the vector which summarizes the values of all the sensory and the motor channels at time t (it is the concatenation of $\mathbf{S(t)}$ and $\mathbf{M(t)}$). In all that follows, there is an internal clock in the robot which discretizes the time, and the values of all channels are updated at every time step. What is crucial is that the robot does not know the "meaning" of the sensory and motor channels: for the robot, each of these channels is like an unlabeled wire from which it can measure some values (the sensory channels) or send some values (the motor channels). This means in particular that if for example the robot is equipped with one touch and two auditory sensory channel, it does not even know that two of them are of the same kind and correspond to the same modality. This kind of organization will be learned progressively by the robot. It is interesting to note that we tackle here the problem of finding structure in a sensorimotor space about which one knows only the unstructured and unlabeled list of sensors and actuators. Some alternative and complementary solutions to this problem are presented in (Philipona et al., 2003; Hafner and Kaplan, 2005; Stronger and Stone, 2006; Olsson et al., 2006; Kuipers et al., 2006)

Regions. There is a mechanism which incrementally splits the sensorimotor space (defined as the set of possible values for $\mathbf{SM(t)}$) into regions, based on these exemplars. A region is an implicit definition of a sub-part of the sensorimotor space, such as for example "the part which is defined by a value of the third sensor channel lower than 0.7 and a value of the second motor channel higher than 0.9". Moreover, IAC equips the robot with a memory of all the exemplars $(\mathbf{SM(t)}, \mathbf{S(t+1)})$ which have been encountered by the robot. Each region stores all the exemplars which it covers. These exemplars will be the used in both the splitting mechanism and in the predictions made by the experts (see below). At the beginning, there is only one region $\mathcal{R}_1$. Then, when a criterion C_1 is met, this region is split into two regions. This is done recursively. A simple criterion C_1 can be used: when the number of exemplars associated with the region is above a threshold T, then split. T is a parameter of the algorithm. This criterion specifies when the robot splits a region, and so refines its categorization in this part of the sensorimotor space, when it has acquired a certain amount of experience in it. Moreover, it is computationally efficient.

When a splitting has been decided, then another criterion C_2 must be used to find out how the region will be split. A possible criterion is one that splits the set of exemplars into two sets so that the sum of the variances of $\mathbf{S(t+1)}$ components of the exemplars of each set, weighted by the number of exemplars of each set, is minimal. This allows to split the sets in a relatively balanced manner, and in the middle of zones of high non-linearity. This criterion is not crucial to the algorithm and could possibly be replaced by another one.

Recursively and for each region, if the criterion C_1 is met, the region is split into two regions with the criterion C_2. When a region is split into two regions, the parent region is deleted, so that the organization of the set of regions is flat. Yet, each region stores all the cutting dimensions and the cutting values that were used in its generation as well as in the generation of its parent region. As a consequence when one wants to find the region associated with a given sensorimotor context $\mathbf{SM(t)}$, it is easy to find out: it is the one for which $\mathbf{SM(t)}$ satisfies all the cutting tests (and there is always a single region which corresponds to each $\mathbf{SM(t)}$). This is illustrated in Figure 2.

Experts. Each region $\mathcal{R}_n$ is associated with a learning machine $\mathbf{E_n}$, called an expert. A given expert $\mathbf{E_n}$ is responsible for the prediction of $\mathbf{S(t+1)}$ given $\mathbf{SM(t)}$ when $\mathbf{SM(t)}$ is a situation which is covered by its associated region $\mathcal{R}_n$. Each expert $\mathbf{E_n}$ is trained on the set of exemplars which is possessed by its associated region $\mathcal{R}_n$. An expert can be a neural-network, a support-vector machine or a Bayesian machine for example. When a region is split, two child experts are created as fresh experts re-trained with the exemplars that their associated region has inherited. This is not necessarily computationally heavy if one uses memory-based learning methods such as nearest-neighbors algorithms as we will do in the next section. Indeed, in this case no re-training needs to be done.

Evaluation of learning progress. The partition of the sensorimotor space into different regions is the basis of our regional evaluation of learning progress. Each time an action is executed by the robot in a given sensorimotor context $\mathbf{SM(t)}$ covered by the region $\mathcal{R}_n$, the robot can measure the discrepancy between the sensory state $\widetilde{\mathbf{S}}\mathbf{(t+1)}$ that the expert $\mathbf{E_n}$ predicted and the actual sensory state $\mathbf{S(t+1)}$ that it measures. This provides a measure of the error of the prediction of $\mathbf{E_n}$ at time $t+1$:

$$e_n(t+1) = ||\mathbf{S(t+1)} - \widetilde{\mathbf{S}}\mathbf{(t+1)}||^2$$

This squared error is added to the list of past squared errors of $\mathbf{E_n}$, which are stored in association to the region $\mathcal{R}_n$. We denote this list:

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

Note that here t denotes a time which is specific to the expert, and not to the robot: this means that $e_n(t-1)$ might correspond to the error made by the expert $\mathbf{E_n}$ in an action performed at $t-10$ for the robot, and that no action corresponding to this

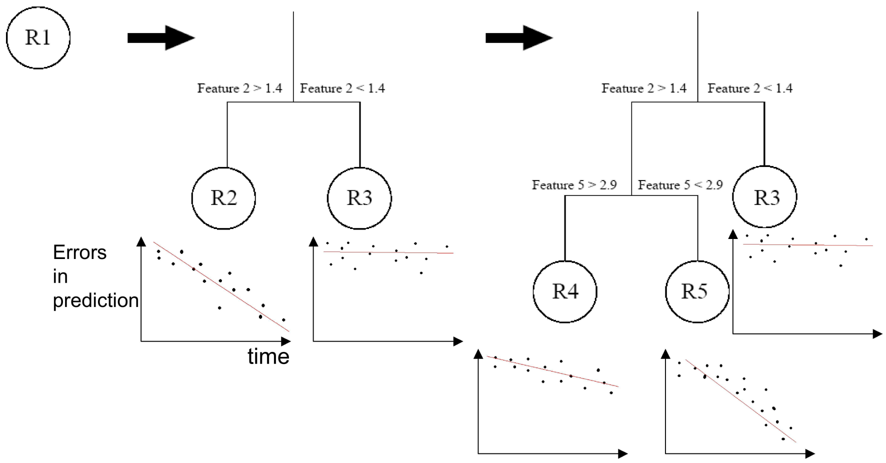

Figure 2: The sensorimotor space is iteratively and recursively split into sub-spaces, which we call "regions". Each region $\mathcal{R}_n$ is responsible for monitoring the evolution of the error rate in the anticipation of the consequences of the robot's actions if the associated contexts are covered by this region. This list of regional error rates is used for learning progress evaluation. As the robot is preferentially interested to explore regions where the learning progress is maximal, these regions will be those that are going to be split most often. For example in this Figure, after the initial region R_1 has been split into R_2 and R_3, the robot discovers that there is more potential learning progress in R_2, and thus will explore this region most of the time. This has the consequence that it will quickly gain experience in it, and because of the criterion C_1, the region R_2 will be split into two new regions R_4 and R_5. This corresponds to a refinement of the categorization system of the robot in this part of the sensorimotor space. Then we see that R_2 was in fact quite heterogeneous in terms of learning complexity, and that initially at least R_5 is more interesting than R_4 in terms of learning progress.

expert were performed by the robot since that time. These lists associated to the regions are then used to evaluate the learning progress associated to each region. Indeed, the learning progress associated with region R_n is defined as the inverse of the derivative of the smoothed curve of the errors in predictions made by E_n in the recent past. Mathematically, the computation involves two steps:

- the mean error rate in prediction is computed at $t+1$ and $t+1-\tau$:

$$< e_n(t+1) >= \frac{\sum_{i=0}^{\theta} e_n(t+1-i)}{\theta+1}$$

$$< e_n(t+1-\tau) >= \frac{\sum_{i=0}^{\theta} e_n(t+1-\tau-i)}{\theta+1}$$

  where τ is a time window parameter typically equal to 15, and θ a smoothing parameter typically equal to 25.

- the actual decrease in the mean error rate in prediction is defined as $D(t+1) =< e_n(t+1) > - < e_n(t+1-\tau) >$. We can then define the actual learning progress as

$$L(t+1) = -D(t+1)$$

Eventually, when a region is split into two regions, both new regions inherit the list of past errors from their parent region corresponding to the exemplars they have inherited, which allows them to make evaluation of learning progress right from the time of their creation.

Action selection. We have now in place a prediction machinery and a mechanism which provides an internal reward (positive or negative) $r(t) = L(t)$ each time an action is performed in a given context, depending on how much learning progress has been achieved. The goal of the intrinsically motivated robot is then to maximize the amount of internal reward that it gets. Mathematically, this can be formulated as the maximization of future expected rewards (i.e. maximization of the returns), that is $E\{\sum_{t \geq t_n} \gamma^{t-t_n} r(t))\}$ where γ $(0 \leq \gamma \leq 1)$ is the discount factor, which assigns less weight on the reward expected in the far future.

This formulation corresponds to a reinforcement learning problem (Sutton and Barto, 1998) and thus the complex techniques developed in this field can be used to implement an action selection mechanism which will allow the robot to maximize future expected rewards efficiently. Yet, the purpose of this article is to focus on the study and understanding of the consequences in terms of behavior of the learning progress definition that we presented. Using a complex reinforcement machinery brings complexity and biases which are specific to a particular method, especially concerning the way they process delayed rewards. While using such a method with intrinsic motivation systems will surely be useful in the future, and is in fact an entire subject of research as illustrated by the work described in (Barto et al., 2004), we will make a simplification which will allow us not to use such sophisticated reinforcement learning methods so that the results we will present in the experiment section can be interpreted more easily. This simplification consists in having the system try to maximize only the expected reward it will receive at $t+1$, i.e. $E\{r(t+1)\}$ This permits to avoid problems related to delayed rewards and it makes it possible to use a simple prediction system which can predict $r(t+1)$, and so evaluate $E\{r(t+1)\}$, and then be used in a straightforward action selection loop. The method we use to evaluate $E\{r(t+1)\}$ given a sensory context $\mathbf{S(t)}$ and a candidate action $\widetilde{\mathbf{M}}\mathbf{(t)}$, constituting a candidate sensorimotor context $\widetilde{\mathbf{SM}}\mathbf{(t)}$ covered by region $\mathcal{R}_n$, is straightforward but revealed to be efficient: it is equal to the learning progress that was achieved in $\mathcal{R}_n$ with the acquisition of its recent exemplars, i.e. $E\{r(t+1)\} \approx L(t-\theta_{R_n})$ where $t-\theta_{R_n}$ is the time corresponding to the last time region $\mathcal{R}_n$ and expert $\mathbf{E_n}$ processed a new exemplar.

Based on this predictive mechanism, one can deduce a straightforward mechanism which manages action selection in order to maximize the expected reward at $t+1$:

- in a given sensory $\mathbf{S(t)}$ context, the robot makes a list of the possible values of its motor channels $\widetilde{\mathbf{M}}\mathbf{(t)}$ which it can set; If this list is infinite, which is often the case since we work in continuous sensorimotor spaces, a sample of candidate values is generated;

- each of these candidate motor vectors $\widetilde{\mathbf{M}}\mathbf{(t)}$ associated with the sensory context $\mathbf{S(t)}$ makes a candidate $\widetilde{\mathbf{SM}}\mathbf{(t)}$ vector for which the robot finds out the corresponding region $\mathcal{R}_n$; then the formula we just described is used to evaluate the expected learning progress $E\{r(t+1)\}$ that might be the result of executing the candidate action $\widetilde{\mathbf{M}}\mathbf{(t)}$ in the current context;

- the action for which the system expects the maximal learning progress is chosen with a probability $1-\epsilon$ and executed, but sometimes a random action is selected (with a probability ϵ). In the following experiments ϵ is typically 0.35.

- after the action has been executed and the consequences measured, the system is updated.

3 Discovering object affordances and communication using the same mechanism

We are now going to present an experimental setup called the Playground Experiment. This is an extension of the setup described in (Oudeyer et al., 2005b) in which the sensorimotor space is larger. This involves a physical developmental robot capable of moving its arms, its neck, its cheeks and to produce sounds, which is installed onto a play mat with various toys as well as with a pre-programmed "adult" robot which can respond vocally to the developing robot in certain conditions. The robots that we use are Sony AIBO robots, as shown on Figure 3. We have developed a web site which presents pictures and videos of this setup: http://playground.csl.sony.fr. We are now going to describe what is the sensorimotor space of the developing robot.

3.1 The sensorimotor space

The motor channels. The robot is equipped with seven continuous motor channels which can control various actuators (see Figure 4):

Figure 3: The Playground Experiment

- two motor channels which control the pan and tilt of the robot's head, whose values at time t are respectively denoted $P(t)$ and $T(t)$, and which are real numbers between 0 and 1;
- two motor channels which control the movement of the front arms: $B_a(t)$ which controls the angle of a bashing movement (the value is a number between 0 and π radians, and when $B_a(t) \leq \pi/2$ the left arm moves, whereas when $B_a(t) \geq \pi/2$ the right arms moves), and $B_s(t)$ which controls the speed of the bashing movement (which is a real number between 0 and 1);
- one motor channel $D(t)$ which controls the angle of the middle joint of the front legs, which allows the robots to kneel down and make an opening-closing cycle of its cheek and which is a real number between 0 and 1;
- two motor channels which control the vocalizations of the robot: $P_{mi}(t)$ which sets the mean pitch of the vocalization, and $L_m(t)$ which controls the length of the vocalization (both are scaled real number between 0 and 1);

Moreover, for each motor channel, a resting position is defined, and the robot can send the value -1 which provokes the setting of the corresponding channels in these rest positions. For example, for the channels $P_{mi}(t)$ and $L_m(t)$, if one of them take the value -1, then no sound is produced. Note also that while some motor channels control directly some motors on the AIBO, like $P(t)$ and $T(t)$, some other channels control coordinations of activations of several motors at the same time, like for example $B_a(t)$ and $D(t)$. Nevertheless, the robot does not know initially that there are different kinds of motor channels, and that they are linked to different kinds of actuators.

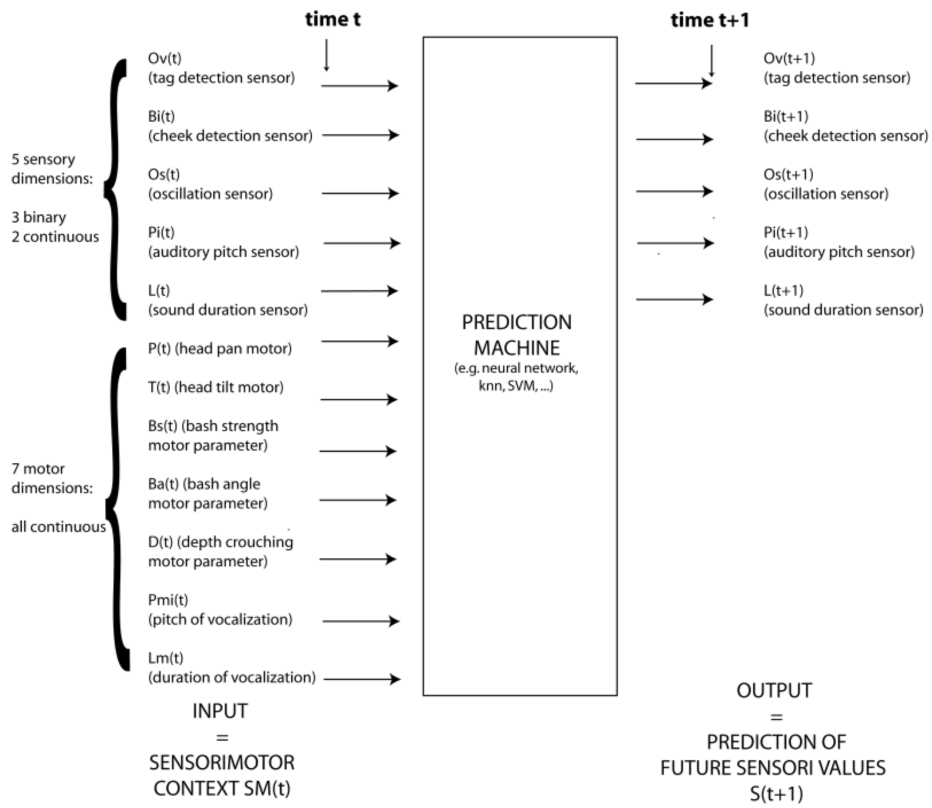

Figure 4: The sensorimotor channels

To summarize, choosing an action consists in setting the values of the 7-dimensional continuous vector $\mathbf{M(t)}$:

$$\mathbf{M(t)} = (P(t),\ T(t),\ B_a(t),\ B_s(t),\ D(t),\ P_{mi}(t),\ L_m(t))$$

The sensory channels. The robot is equipped with five sensory channels, two of which are continuous and three of which are binary, and corresponding to various sensors:

- $O_v(t)$ is a binary sensory channel indicating the presence (or absence) of a visual tag within the field of view. This is based on the use of the video camera which is on the head of the AIBO. These visual tags were put near the objects in the play mat;
- $B_i(t)$ is a binary sensory channel indicating the presence (or absence) of an object in the mouth of the AIBO. This uses the cheek sensor;
- $O_s(t)$ is a binary sensory channel indicating whether something is oscillating or not in the close range of the infra-red distance sensor which is on the nose of the AIBO;

- $P_i(t)$ is a continuous sensory channel which indicates the mean pitch of the sound which is heard when a sound is actually perceived by the AIBO microphone, and which has a continuous value between 0 and 1 when a sound is perceived, and is -1 when there are no sounds;

- $L(t)$ is a continuous sensory channel which indicates the duration of the sound which is heard when a sound is actually perceived by the AIBO microphone, and which has a continuous value between 0 and 1 when a sound is perceived, and is -1 when there are no sounds;

The robot does not know initially that there are different kinds of sensory channels, and for example does not know that some are binary or continuous (it treats them all as continuous a priori) or that some are related to vision and some other to audition for example.

To summarize, the sensory vector $\mathbf{S(t)}$ is the 5-dimensional vector:

$$\mathbf{S(t)} = (O_v(t),\ B_i(t),\ O_s(t),\ P_i(t),\ L(t))$$

The external environment On the play mat, there are two toys and an "adult" robot that the "child" robot can detect as objects using its O_v sensor (but the visual sensor only tells the robot whether or not there is an object in front of it, it does not say whether this is a toy or another robot for example). There is one "elephant ear" integrated in the play mat that the robot can possibly bite but that does not produce perceivable reactions when it is bashed. There is a suspended soft toy that it can bash with the arm but which is too far for biting. Finally, there is one pre-programmed "adult" robot which imitates the sounds produced by the developing robot (with a different voice which shifts the pitch down) only when the developing robot is looking in the direction of the adult robot while it is vocalizing. The adult robot is far enough from the developing robot so that it is not possible to bite or bash it.

Initial ignorance of sensorimotor affordances More generally, 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 $B_i(t)$ or $O_s(t)$ can become 1 only when the values $D(t)$, $B_a(t)$ and $B_s(t)$ are coordinated in a certain manner and when there is an object in the direction of the action. It does not know that some objects are more prone to provoke changes in the values of $B_i(t)$ and $O_s(t)$ when only certain kinds of actions are performed in their direction. It does not know for example that to get a change in the value of the $O_s(t)$ channel, bashing in the correct direction is not enough, because it also needs to look in the right direction (since its infra-red distance sensor is on the front of its head). These remarks make it clear that a random action selection would lead most often to uncoordinated movements which produce no effect on the environment.

The action perception loop. To summarize, the mapping that the robot learns is:

$$f:\ \mathbf{SM(t)} =$$

$$(P(t),\ T(t),\ B_a(t),\ B_s(t),\ D(t),\ P_{mi}(t),\ L_m(t),\ O_v(t),\ B_i(t),\ O_s(t),\ P_i(t),\ L(t))$$

$$\longmapsto \mathbf{S(t+1)} = (\widetilde{O_v(t+1)},\ \widetilde{B_i(t+1)},\ \widetilde{O_s(t+1)},\ \widetilde{Pi(t+1)},\ \widetilde{L(t+1)})$$

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, using the algorithms described in the previous section.

3.2 Results

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 which 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: the head pan/tilt, the arm, the mouth and the sound speakers (used for vocalizing). We also constantly measure the direction in which the robot is turning its head.

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	-	-	-	+

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 5.

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 than two phases of increasing behavioral complexity, i.e. we observe the emergence of several successive levels of behavioral patterns; 4) the robot reaches a behavioral pattern in which it focuses on vocalizing towards the "adult" pre-programmed robot and on listening to the vocal response that it triggers. 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. These behavioral patterns correspond to combinations of clear deviations from the mean in the curves in Figure 5. 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[1]. Here are the different stages which are visually denoted in Figure 5 (see also 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 par-

[1]This definition of a stage is inspired from that of Piaget (Piaget, 1952)

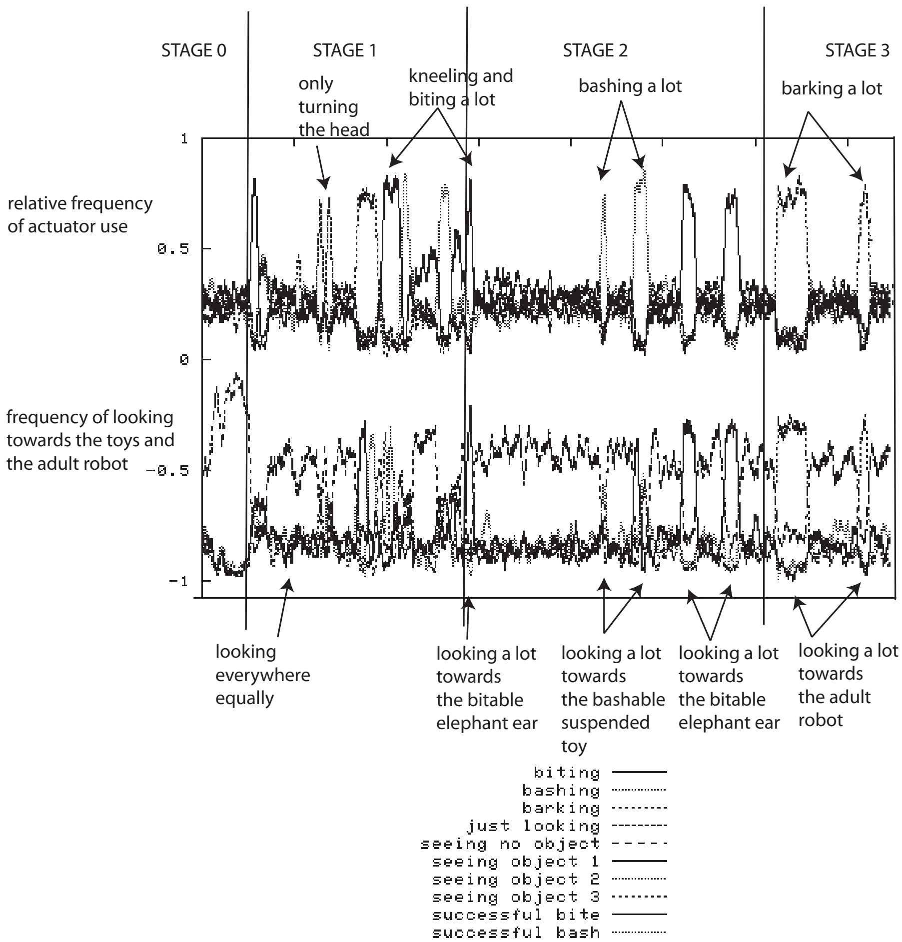

Figure 5: Top curves: relative frequency of the use of different actuators (head pan/tilt, arm, mouth, sound speaker). Bottom curves: frequency of looking towards each object and in particular towards the "adult" pre-programmed robot.

titioned in significantly different areas. During this stage, the robot's behavior is equivalent to the one we would obtain using a 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 robots begins to focus successively on playing with individual actuators, 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 towards 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 kinds of actuators. This results from the fact that using one actuator at a time makes the $\mathbf{SM(t)} \longmapsto \mathbf{S(t+1)}$ 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 action types and certain particular objects: it is now focusing either on trying to bite the biteable object (the elephant ear), and on trying to bash the bashable object (the suspended toy). Furthermore, the trace shows that it does actually manage to bite and bash successfully quite often, which is an emergent side effect of Intelligent Adaptive Curiosity and was not a pre-programmed task. This focus on trying to do actions towards affordant objects is a result of the splitting mechanism of IAC, which is a refinement of its 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 since 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 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 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 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 physical actions with the right affordances, and then to 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 intermediary stages 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 precise statistical measures about the set of developmental sequences that are generated in our experiments (which takes time since one single experiment takes nearly one day) in order to understand better how particular environment and embodiment conditions lead to the formation of recurrent developmental stages.

4 Discussion

The experiment that we presented is not intended to be a model of child development. It must be seen as a metaphorical system that can help us think of the complex dynamics that result from intrinsic motivation like maximizing learning progress. In order to characterize such kinds of developmental trajectories, we have introduced elsewhere the notion of "progress niches" (Kaplan and Oudeyer, 2005). Progress niches are situations, neither too predictable nor too difficult to predict, optimally interesting given the developmental stage reached by the agent. Therefore, progress niches are not intrinsic properties of the environment. They result from a relation between a particular environment, a particular embodiment (sensors, actuators, features 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[2].

In the present case, the notion of progress niche can help us think about the following hypothetical developmental scenario in which "the discovery of communication" takes place. If children indeed show some form of intrinsic motivation for maximizing learning progress, they will get successively interested in playing with and manipulating objects or situations of increasing complexity, the choice of which is determined by

[2]The concept of progress niches is related to Vygotsky's *zones of proximal development*, where the adult deliberately challenges the child's level of understanding. Adult push children to engage in activities beyond their current mastery level, but not too far beyond so that they remain comprehensible (Vygotsky, 1978). We could interpret the zone of proximal development as a set of potential progress niches organized by the adult in order to help the child learn. But it should be clear that independently of adults' efforts, what is and what is not a progress niche is ultimately defined from the child's point view.

their potential returns in terms of learning progress. At a certain point in their development, children should discover that interacting with others, for example by producing and listening for sounds, is an important source of learning progress, a new progress niche. This discovery will make them more and more interested and involved in social activities.

Early communication situations can indeed be considered as progress niches. Parental scaffolding plays a critical role for making the interaction with the child predictable and adapted to the child's competences (Schaffer, 1977). Numerous studies concerning *motherese* can be viewed as support for this hypothesis. As the children progress, they get confronted with updated challenges, new progress niches sustaining their interest for linguistic communication. However, things get more complex at this stage and go beyond the scope of the model presented in this article. For instance, it could be argued that the space that children explore changes as they develop new competences. Children shift their interest from linguistic forms (e.g. when playing to imitate vocalizations) to linguistic content (the meaning of vocalizations). This later aspect is not present in our experiment. Many issues remain to be solved to capture convincingly such increases in complexity in an artificial setup. In particular, in this paper we have shown how an intrinsic motivation system could allow a robot to self-organize its exploration of a given sensorimotor space. But the structure of this sensorimotor space was fixed and did not evolve. Some other work in developmental robotics investigate this question of how new and more complex representations can be bootstrapped by an autonomous robot (Stronger and Stone, 2006; Schlesinger, 2006; Olsson et al., 2006; Gold and Scassellati, 2006; Provost et al., 2006; Kuipers et al., 2006). An interesting continuation of the work presented in this paper could be to use intrinsic motivation systems to guide the exploration of such evolving representational spaces.

What then is the added value of our experiment? The setup illustrates how in the absence of dedicated channel for communication and specific motivation for interaction, primitive linguistic communication can nevertheless be rewarding and actively practiced. However, one could argue that there is actually a dedicated channel for communication as the only possible effect of vocalizations in our setup is to trigger vocal responses by the other robot. This is true, but the crucial point is that the robot does not know about it. Initially, its sound production system is indistinguishable from the rest of its motor system, from its point of view. It has to discover that using this modality leads to qualitatively different result. Moreover, the developing robot does not know that the "adult" robot is a special entity: indeed, it initially considers it to be just another object among others.

Then, can the simple imitation routine demonstrated in this experiment really be considered linguistic communication? No, it is rather a primitive form of vocal interaction, but our point is that more complex linguistic communication shares the same kind of special dynamics that distinguishes it from interaction with simple objects. In our setup, learning to predict the effects of the vocal outputs is different than predicting the effects of the motor commands directed towards non-communicating objects. We believe that communication situations, in general, are characterized by such kinds of different learning dynamics. However, this does not mean that they are always more difficult to learn than learning how to interact with objects (simple forms of imitative behavior have been argued to be present just after birth (Meltzoff and Gopnick, 1993;

Moore and Corkum, 1994)).

Finally, as in this experiment, the environment and the embodiment of the robot have been designed to make communication special given the intrinsic motivation system of the robot. How is this different than an innate motivation for communication? The crucial difference is that in our system the cognitive machinery as well as the motivation system are not specific to communication. From a phylogenetic perspective, the evolution of a motivation for optimizing learning progress, which is essentially linked to behaviors of play and exploration, is much easier to justify than the evolution of a specific motivation for communication.

As illustrated in this experiment, computational and robotic approaches may shed new light on the particular role played by intrinsic motivation in complex developmental processes. We have presented a setup to explore a scenario based on the hypothesis that children may discover communication through a general process of motivated exploration. However, it should be clear that we do not even suggest that maximizing learning progress could be the only motivational principle driving children during their development. Development certainly results from the interplay between a complex set of drives, particular learning biases, as well as embodiment and environmental constraints. Our hope is that this form of experiment can help to develop our intuitions and to better understand the different components that contribute to shaping the dynamics of child development.

5 Acknowledgements

We would like to thank Verena Hafner and Andrew Whyte for their crucial help for building the Playground experiment. In particular, they programmed the low-level control system of the robot. We thank Jean-Christophe Baillie for his URBI system (Baillie, 2004) 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

Baillie, J., 2004. Urbi: A universal language for robotic control. International Journal of Humanoid Robotics.

Barto, A., Singh, S., 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.

Cangelosi, A., Parisi, D., 2002. Simulating the Evolution of Language. Springer.

De Boer, B., 2001. The origins of vowel systems. Oxford University Press.

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

Dessalles, J.-L., 1998. Altruism, status and the origin of relevance. In: Hurford, J., Knight, C., Studdert-Kennedy, M. (Eds.), Approaches to the Evolution of Language: Social and Cognitive bases. Edinburgh University Press, Edinburgh.

Gintis, H., Smith, E., Bowles, S., 2001. Costly signaling and cooperation. Journal of Theoretical Biology 213, 103–119.

Gold, E., 1967. Language identification in the limit. Information and Control 10, 447–474.

Gold, K., Scassellati, B., 2006. Learning acceptable windows of contingency. Connection Science 18 (2).

Hafner, V., Kaplan, F., 2005. Interpersonal maps and the body correspondence problem. In: Demiris, Y., Dautenhahn, K., Nehaniv, C. (Eds.), Proceedings of the Third International Symposium on Imitation in animals and artifacts. Herderfordshire, UK, pp. 48–53.

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

Hutchins, E., Hazlehurst, B., 1995. How to invent a lexicon: the development of shared symbols in interaction. In: Gilbert, N., Conte, R. (Eds.), Artificial Societies: The Computer Simulation of Social Life. UCL Press, London, pp. 157–189.

Kaplan, F., 2005. Simple models of distribued coordination. Connection Science 17 (3).

Kaplan, F., Oudeyer, P.-Y., 2004. Maximizing learning progress: an internal reward system for development. In: Iida, F., Pfeifer, R., Steels, L., Kuniyoshi, Y. (Eds.), Embodied Artificial Intelligence. LNCS 3139. Springer-Verlag, London, UK, pp. 259–270.

Kaplan, F., Oudeyer, P.-Y., 2005. The progress-drive hypothesis: an interpretation of early imitation. In: Dautenhahn, K., Nehaniv, C. (Eds.), Models and mechanisms of imitation and social learning: Behavioural, social and communication dimensions. Cambridge University Press, to appear.

Ke, J., Minett, J., Au, C.-P., Wang, W., 2002. Self-organization and selection in the emergence of vocabulary. Complexity 7 (3), 41–54.

Kirby, S., 2002. Natural language and artificial life. Artificial life 8, 185–215.

Kuipers, B., Beeson, P., Modayil, J., Provost, J., 2006. Bootstrap learning of foundational representations. Connection Science 18 (2).

Marshall, J., Blank, D., 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.

Meltzoff, A., Gopnick, A., 1993. The role of imitation in understanding persons and developing a theory of mind. In: S. Baron-Cohen, H. T.-F., D.Cohen (Eds.), Understanding other minds. Oxford University Press, Oxford, England, pp. 335–366.

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

Olsson, L., Nehaniv, C. L., Polani, D., 2006. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. Connection Science 18 (2).

Oudeyer, P.-Y., 2005. How phonological structures can be culturally selected for learnability. Adaptive Behavior 13 (4).

Oudeyer, P.-Y., 2006. Self-Organization in the Evolution of Speech. Studies in the Evolution of Language. Oxford University Press.

Oudeyer, P.-Y., 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., Balkenius, C. (Eds.), Proceedings of the 4th International Workshop on Epigenetic Robotics. Vol. 117. Lund University Cognitive Studies, pp. 127–130.

Oudeyer, P.-Y., Kaplan, F., Hafner, V., 2005a. Intrinsic motivation systems for open-ended development. Technical report 05-03, SONY CSL PARIS.

Oudeyer, P.-Y., Kaplan, F., Hafner, V. V., Whyte, A., 2005b. The playground experiment: Task-independent development of a curious robot. In: Bank, D., Meeden, L. (Eds.), Proceedings of the AAAI Spring Symposium on Developmental Robotics, 2005. Stanford, California, pp. 42–47.

Philipona, D., O Regan, J., Nadal, J.-P., 2003. Is there something out there? inferring space from sensorimotor dependencies. Neural Computation 15 (9), 2029–2049.

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

Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behavioral and Brain Sciences 13 (4), 707–784.

Provost, J., Kuipers, B. J., Miikkulainen, R., 2006. Developing navigation behavior through self-organizing distinctive state abstraction. Connection Science 18 (2).

Rolls, E. T., 1999. The Brain and Emotion. Oxford UP.

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

Schlesinger, M., 2006. Decomposing infants' object representations: A dual-route processing account. Connection Science 18 (2).

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

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

Shoham, Y., Tennenholtz, M., 1997. On the emergence of social conventions: modeling, analysis, and simulations. Artificial Intelligence 94 (1–2), 139–166.

Smith, K., Kirby, S., Brighton, H., 2003. Iterated learning: a framework for the emergence of language. Artificial Life 9 (4), 371–386.

Steels, L., J. D. B., Neubauer, N., 2005. Linking in fluid construction grammars. Submitted.

Steels, L., 1996. Emergent adaptive lexicons. In: Maes, P., Mataric, M., Meyer, J.-A., Pollack, J., Wilson, S. (Eds.), From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior. The MIT Press, Cambridge, MA, pp. 562–567.

Steels, L., July 2003. Evolving grounded communication for robots. Trends in Cognitive Science 7 (7), 308–312.

Steels, L., Kaplan, F., 1999. Collective learning and semiotic dynamics. In: Floreano, D., Nicoud, J.-D., Mondada, F. (Eds.), Advances in Artificial Life: 5th European Conference (ECAL 99). Lecture Notes in Artificial Intelligence 1674. Springer-Verlag, Berlin, pp. 679–688.

Steels, L., Kaplan, F., 2000. Aibo's first words: The social learning of language and meaning. Evolution of Communication 4 (1), 3–32.

Stronger, D., Stone, P., 2006. Towards autonomous sensor and actuator model induction on a mobile robot. Connection Science 18 (2).

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

Thrun, S., 1995. Exploration in active learning. In: Arbib, M. (Ed.), Handbook of Brain Science and Neural Networks. MIT Press, Cambridge, MA.

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

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

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