



HAL
open science

Evolutionary Perspective for Goal-Babbling Exploration

Fabien Hervouet, Eric Bourreau

► **To cite this version:**

Fabien Hervouet, Eric Bourreau. Evolutionary Perspective for Goal-Babbling Exploration. ICDL-Epirob: Development and Learning and Epigenetic Robotics, Oct 2014, Genoa, Italy. pp.337-342, 10.1109/DEVLRN.2014.6983003 . hal-01074295

HAL Id: hal-01074295

<https://hal.science/hal-01074295>

Submitted on 14 Oct 2014

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.

Evolutionary Perspective for Goal-Babbling Exploration

Fabien Hervouet
Université Montpellier 2
fabien.hervouet@lirmm.fr

Eric Bourreau
Université Montpellier 2
eric.bourreau@lirmm.fr

Abstract—Goal-Babbling Exploration consists in the generic task of sensory-motor coordination learning through goal space exploration. A typical drawback in Goal-Babbling Exploration algorithms remains the human involvement during the adjustment phase. In order to tackle this issue, we propose in this article to study an evolutionary perspective. It takes the form of applying an optimization approach to enhance the learning process. Our goal is to disengage the modeler without efficiency loss. We use a simple chromosome describing every parameters on which the model is built. The experimental results obtained demonstrate its feasibility: a self-adjustment of these parameters bringing up good enough developmental trajectories.

I. INTRODUCTION

Goal-Babbling Exploration in developmental robotics has recently become an intensive research field for sensory-motor acquisition. Developmental robotics aims at defining models of developmental mechanisms, allowing a full open-ended incremental learning, compliant and resilient with a real environment, leading to a staged growth of knowledge.

So far, for the vast majority, it still involves a *benevolent supervising human modeler*, seeking to manually optimize his own particular model parameters. Unfortunately this assumption lacks credibility with the rest of the intrinsically developmental paradigm, fundamentally self-organized.

What we propose in this article is basically asking these developments from an evolutionary perspective to find out efficient structural choices, endeavoring to consider the following problem: *how to disengage the modeler in order to give priority to automated adaptation of motivational processes?* We answer by first highlighting the limitations of the external modeling approach, showing possible specific parameters of models for a given incarnation, and then proposing a phylogenetic approach to solve it.

To complete the description of the article, we start by recalling some preliminary background about goal-babbling research and we finish by analyzing *in silico* experimental results¹.

¹This work was realized with the support of HPC@LR, High-Performance Computing Center in Languedoc-Roussillon, France.

II. GOAL-BABBLING EXPLORATION FOR REACHING PROBLEM

A. Goal-Babbling Exploration

Goal-babbling has become a new research field within the last five years. Historically the goal-babbling approach originates from the need to improve inverse models learning for any kind of embodiment through self-exploration. This qualifies the reaching problem, which consists, for example, in moving a robot arm so as to reach a desired target. It remains strongly associated with the developmental robotics field which considers a robot to be seen as pro-active in its own learning process [10], [1].

At the very beginning, researchers discovered a simple way to explore affordances that is generating motor commands to get information so that the agent would be able to build an inverse control model of its particular embodiment. This first step is usually referred as *Motor-Babbling Exploration* (MBE). First works proposed an exploration through randomly generated motor commands, then heuristically guided algorithms appeared to guide the agent by accounting for its weaknesses. For instance, [12] are architected around a MBE approach that is an agent exploring its motor space \mathcal{M} being driven by the *improvement of predictions* of couples (motor command, sensory effect).

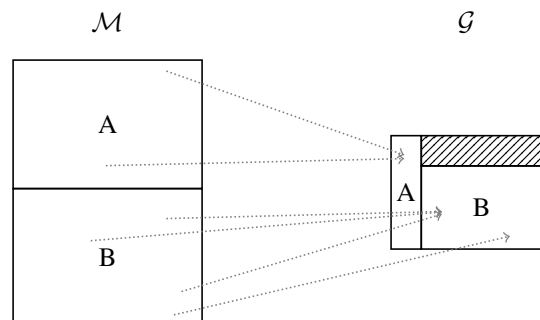


Fig. 1. Illustration of the redundancy within the (surjective) mapping between the motor space \mathcal{M} and the goal space \mathcal{G} . It also shows possible volume dissimilarity between some motor subspaces and their observable consequences in the goal space.

It has since been underlined that exploring a *goal space* \mathcal{G} instead of a *motor space* \mathcal{M} leads to more efficient learning experience. This approach represents the so-called *Goal-Babbling Exploration* (GBE) [2]. The main reason of its important interest is the fact that the agent must learn sensory-motor coordinations in a redundant and high-dimensional motor space in which many different motor commands might result in a sole effect in goal space (cf. figure 1). This means that exploring the huge number of combinations of commands for every actuators cannot be fairly well considered as the right choice given the rather short lifetime of any learning agent [14]. Moreover, it appeared that this computational evidence has also been validated by works in infant developmental psychology: [17], [4]. Recent works investigated bootstrapping approach for coordination skills without pre-specifying, or even representing, a set of goals [13].

B. FIMO: Framework for Intrinsic Motivation

We developed a generic *Framework for Intrinsic Motivation*. It aims at unifying both individual and collective needs to provide a shared experimental frame. It allows to introduce new ideas, implement them and moreover to test their efficiency according to a common experimental protocol [8]. Beyond, *FIMO* holds advantages like its numerous configurable options and the multiple embodiments available. Within this framework, one can run reproducible experiments using specific parameters configuration, and write easily comparison tests for them. The visualization module allows to observe and study experimental differences.

In this framework, an agent is involved in a long-term developmental learning loop, driven by a progress-based metric of interest. Detailed information about model and algorithm is available in previous publications [7], [8]. From an algorithmic point of view, this model is based on two independant kinds of data: *raw data* and *structuring data*.

Raw data represent every low-level sensorimotor transitions experienced by the agent. These are tuples that represent a transition from a start state s^t to a final state s^{t+1} via a motor command m^t . They are used to compute motor commands to execute for reaching a goal.

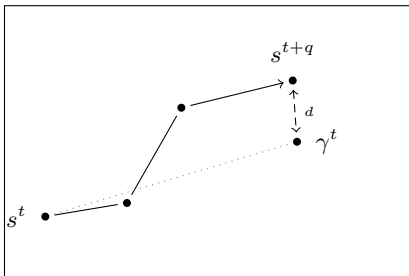


Fig. 2. Graphical explanation of how the agent tries to reach a goal γ^t , starting from s^t , using q sequential attempts.

Structuring data represent high-level experimentations for driving space exploration using the notion of competence progress. These are tuples composed of a start state s^t , a final state s^{t+q} , a goal state γ^t . We can compute the final distance d between s^{t+q} and γ^t , as illustrated on figure 2, and an associated competence value κ^t . These experiments are anchored in a unique region of the exploration space according to the goal location, and are used to compute the interest value of this particular region.

The model describes three nested loops (cf. figure 3). The first one, indicated by the word *Beginning*, consists in generating a particular sensory configuration as a goal. The second one operates so as to try reaching this goal in q attempts (cf. figure 2). After every motor command has been computed and executed, the system broaden its raw data collection. When the goal is considered reached or the agent runs out of reaching attempts, the process goes back to the first level. When a reaching attempt is considered too bad, a third inner loop takes control and leads a local exploration learning phase, in order to acquire p new transitions from the location really reached and thus improve next trial.

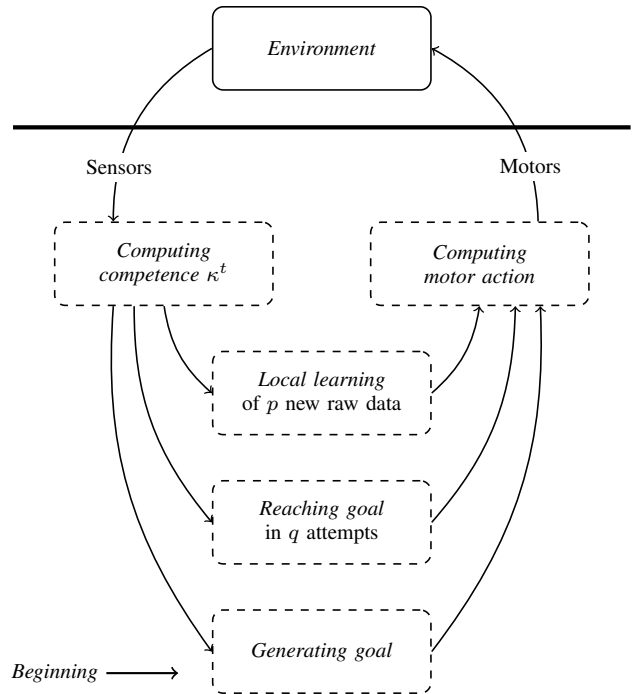


Fig. 3. Algorithmic architecture of the motivational model, composed of three nested loops: goal generation, reaching attempts and local learning.

Unfortunately, contrary to developmental philosophy promoted by this high-level architecture, we must face some low-level manual adjustments.

III. MOTIVATIONS AND PROBLEM

In this section, we first describe particular application of the *FIMO* framework to a simulated robotic arm. We then lead parametric studies on previously presented parameters: the maximum number of attempts to reach a goal (q) and the number of local exploration in learning loop (p).

A. Application to robotic arm

We used for our experiments a simulated setup typically used – in particular in [2], [3] – to test intrinsic motivation algorithm: a robotic arm. The figure 4 illustrates this setup: 15 joints moving 15 equal lengths limbs; θ_i represent angles of each joint, relatively to θ_{i-1} .

The objective is to let the robot discover its own operational space and automatically learn to control its arm in order to be able to reach any specific location, i.e. finding motor commands that move its end-effector s^t towards a particular location γ^t .

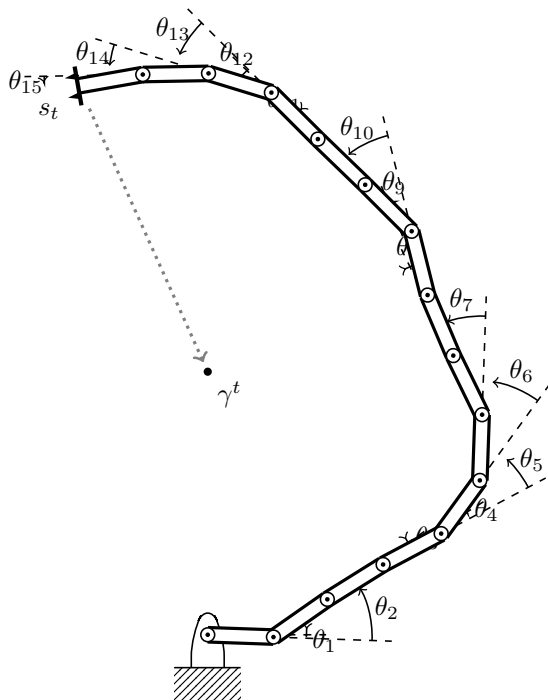


Fig. 4. Schematic representation of a 15 joints robotic arm setup, with equal limbs lengths, moving in 2D space.

B. Parametric study

We faced the classical problem of adjusting these parameters without really being able to predict the consequences. Using our experimental *FIMO* framework, we can easily lead a parametric study. We propose here to focus on two parameters: *exploration trials* (p) and *max reaching trials* (q), as referenced in figure 3.

Monitoring the evolution of the learning progress through time exhibits a particular developmental trajectory. A curve is

composed of the consecutive results for regular examinations. More precisely, an examination is constituted with a set of questions uniformly distributed over the real reachable goal space. A result represents the average distance d to a subset of representative reachable goals (cf. figure 2).

Figures 5 and 6 show different developmental trajectories, where each curve is the average result of 10 runs. X-axis represents the number of motor commands, i.e. motor actions. Y-axis is average distance to questions. AUC is an performance indicator (cf. section IV-C).

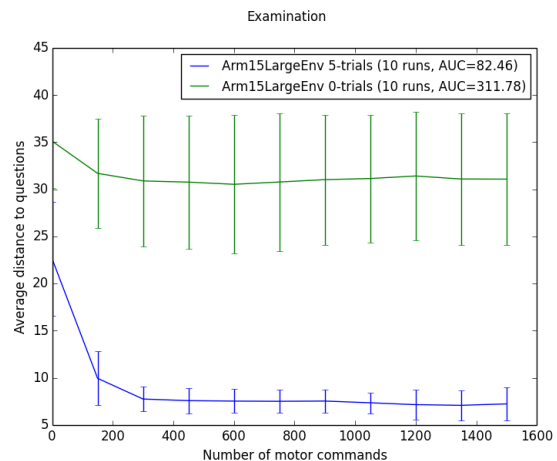


Fig. 5. Experimental results for the number of exploration trials: $p = 0$, i.e. disabling exploration phase (blue) and $p = 5$ (green).

The first studied parameter relates to local learning phase and represents the number p of new transitions to be acquired from the current state. The figure 5 demonstrates clearly importance of such mechanism. Without exploration ($p = 0$), results show bad average distance to the goal of 31.06, while for $p = 5$ average distance equals 7.22, more than 4 times better.

The second one relates to reaching phase and represents the quota q of actions allowed to reach the generated goals. In figure 6, contrary to previous one, every curve exhibits similar behavior: quick decreasing and then plateau. Thus, it is complicated to explain the possible influence of such parameter. Given that the aim of this article is not to analyze deeply the algorithm, we want here to emphasize the difficulty to find out the *good* values.

We introduced in this section only two parameters among more than ten. For instance, we tested different values of the threshold of competence for considering a goal as reached. The presence of so many parameters makes the task of the human tedious. Moreover, this difficulty comes at least from interrelationships characterizing some subsets of parameters required for loop instantiation. These *potential interrelationships* between parameters governing the system

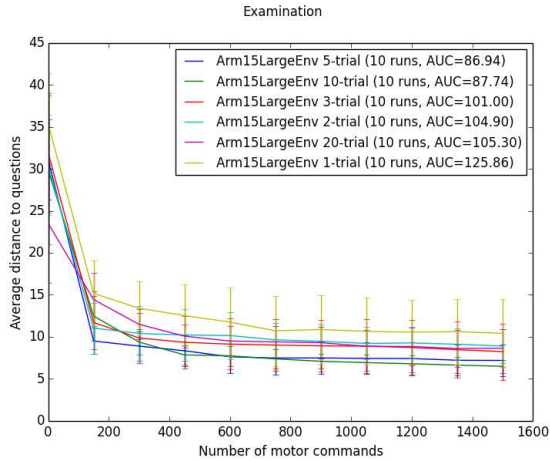


Fig. 6. Experimental results for the quota of reaching attempts: $q = 1$, $q = 2$, $q = 3$, $q = 5$, $q = 10$ and $q = 20$.

may lead to qualify such a system as being complex.

Furthermore, experimenting two different embodiments with the same settings of the main loop implies getting different results or developmental trajectories. This is absolutely not a side-effect but a requirement as we wanted the sensory-motor learning loop to be as generic as possible in order to guide any embodied agent towards experimenting its very own affordant sensory-motor coordinations.

Therefore we propose to use an evolutionary approach for GBE model, in order to perform an automatic human-free parameters ajustement.

IV. EVOFIMO: EXPERIMENTAL RESULTS

In this section we present the experimental results by evolving parameters on a robotic arm. We integrated this new evolutionary perspective in our framework *FIMO*², so that we can call it *EvoFIMO*. The basic evolutionary part is delegated to a library called *pyevolve*. There are many advantages to use this existing piece of software, including facts like: it is written in pure Python like our framework; we can export and visualize easily the evolution statistics through graphs; and more importantly it is open-source.

Evolutionary approaches represent a set of methods and techniques allowing to evolve a population of agents in order to fit them to some point. It has been largely applied to robotics within the last twenty years [11]. But for what concerns us more specifically, there are very few existing relative works [15], [16].

²The Python open source code of the framework is available at <https://info-depot.lirmm.fr/republic/fimo> (public cloning repository) released under GNU GPLv3 license.

A. Principle of evolutionary approach

The first step in evolutionary approaches is the creation of an initial population of solutions for the problem, and then evolve this population along multiple generations. The idea is to improve the solutions taking the best characteristics found in selected parents, and keep this process running until certain stopping criterion is reached and an appropriate solution is obtained.

B. Chromosome

Practically, our chromosome is composed of 13 parameters we want to evolve – as showed in figure 7 – represented as a simple vector. The first two ones are q and p as already explained in previous section. The rest of the chromosome represent other parameters that came up from the previous parametric study we led.

C. Fitness function

As fitness function, we use a numerical integration technique to approximate the *Area Under the learning Curve* of any particular experimental run. The figure 8 illustrates the difference between two developmental trajectories and their respective *AUC* values. This value constitutes a pretty good hint of an agent’s developmental trajectory and performance. With same final examination average distances, we can distinguish different behaviors over time.

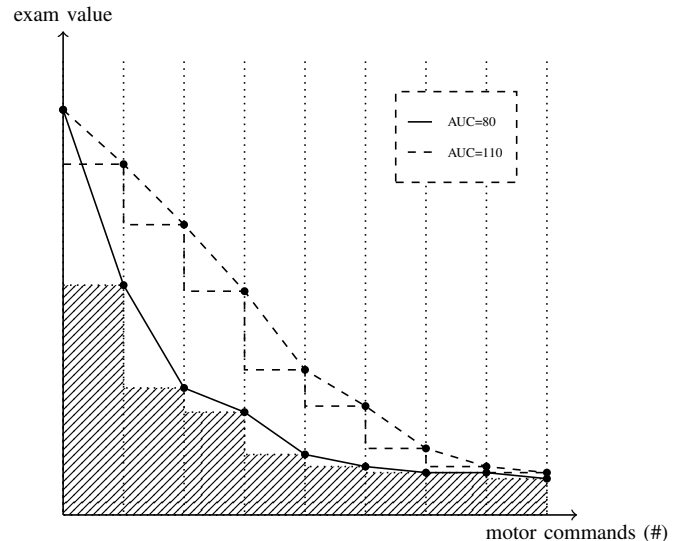


Fig. 8. Illustration of *Area Under Curve* performance indicator, where every dotted line represents a particular exam.

We chose to perform 3 runs, each one during 1500 motor commands, in order to compute the average value of *AUC* for evaluating each individual.

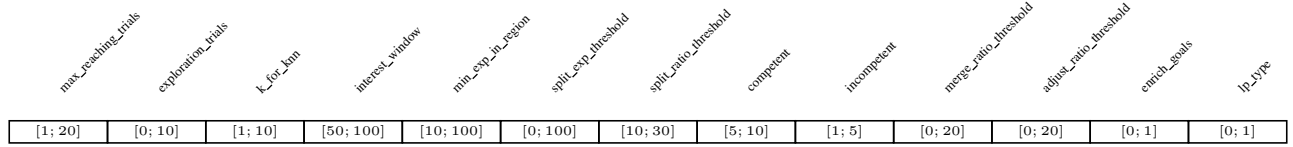


Fig. 7. Chromosomes structure with range values.

D. Population

An initial population of size 10 is generated in a random way by choosing values in different ranges for each parameter, as indicated on figure 7.

We evolve this initial population along 100 generations. The elitism replacement value is 1, which means that we keep the best individual as it is, and generate the others by crossover. To select individuals, we use classical roulette wheel mechanism. The crossover rate is 80%, which means that any new individual results from two parents, each of them contributing respectively with 11 and 2 values. We didn't need any repair mechanism. These values are default ones proposed by *pyevolve* package.

E. First results

Figure 9 shows the fitness evolution of the population over generations. A preliminary observation is the relatively large distribution of the initial population (standard deviation is 91.67) underlining the difficulty of adjusting parameters, as we pointed out in the section III-B.

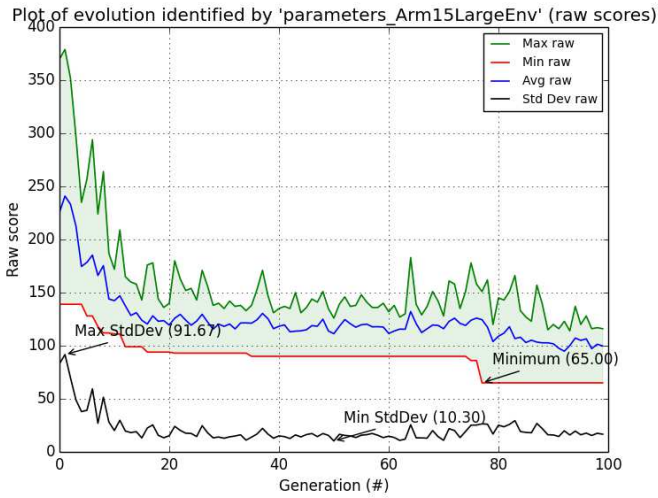


Fig. 9. Fitness evolution for parameters, where raw scores represent AUC values.

What matters is the life-saving decrease of the fitness function from an initial range [139; 375] to a final range [65; 120]. The final average value equals 101.2, which is equivalent to average AUC value for human adjustment, in figure 6. Moreover, the best self-adjusted individual found (65.00) is better than both human-based configuration proposed (82.46)

and any results from all our parametric study. This first result tends to show the interest of this self-adjustment procedure, despite the CPU time requirement (19937 seconds as total elapsed time). The next steps should be to use meta-parameter optimization methods to improve the overall process as proposed and experimented in [5], [6]. Moreover, in a practical way, we could think of alternating *in silico* optimization and real robotic phases in order to decrease the time needed to improve any considered developmental robot.

F. Perspectives

Following the same *credo* we are working on experimenting other criteria to evolve. For instance on the same level, we are interested in experimenting morphological evolution (e.g. arrangement of limbs length, number of limbs for robotic arm) to mimic natural selection.

On another higher level, we are also studying other chromosome structures for using genetic programming [9]. Indeed, apart from the numeric parameters, the model is built on specific mathematically defined metrics that also could be self-adjusted. For instance, the metric we use to compute competence is:

$$\kappa^t = \frac{\|s^{t+q} - \gamma^t\|}{\|s^t - \gamma^t\|}$$

It would be represented as evolvable tree chromosome as described on figure 10.

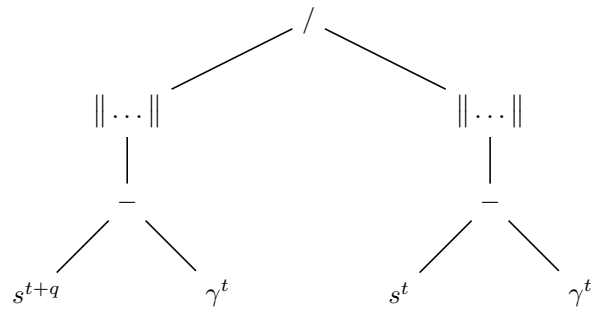


Fig. 10. Example of tree-based metric chromosome.

V. CONCLUSION

Throughout this paper, we proposed an innovative evolutionary perspective for Goal-Babbling Exploration. Our *leitmotiv* is that we need a generic motivational loop, whose overall operation is to push the agent towards reaching goals in accordance with its current level, experience and learning progress margin.

Furthermore, the exact need for different embodiments may obviously be different in order to reveal its very own affordances and improve its very own learning progression. Hence it follows quite naturally to try imitating of *natural selection* in order to fit some particular aspects that may impact the generic loop for a specific embodiment in an evo-devo perspective.

We propose a start for disengaging the benevolent supervising human modeler by self-adjusting parameters. The experiments exhibited results at least as efficient as manually defined configurations.

REFERENCES

- [1] ASADA, M., HOSODA, K., KUNIYOSHI, Y., ISHIGURO, H., INUI, T., YOSHIKAWA, Y., OGINO, M., AND YOSHIDA, C. Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development* 1 (2009), 12–34.
- [2] BARANES, A., AND OUDEYER, P.-Y. Intrinsically-motivated goal exploration for active motor learning in robots: A case study. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)* (2010).
- [3] BARANES, A., AND OUDEYER, P.-Y. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems* 61, 1 (2013), 49–73.
- [4] BERTHIER, N., AND KEEN, R. Development of reaching in infancy. *Experimental Brain Research* 169 (2006), 507–518.
- [5] ELFWING, S., UCHIBE, E., DOYA, K., AND CHRISTENSEN, H. I. Co-evolution of shaping rewards and meta-parameters in reinforcement learning. *Adaptive Behaviour* 16, 6 (2008), 400–412.
- [6] ELFWING, S., UCHIBE, E., DOYA, K., AND CHRISTENSEN, H. I. Darwinian embodied evolution of the learning ability for survival. *Adaptive Behaviour* 19, 2 (2011), 101–120.
- [7] HERVOUET, F., AND BOURREAU, E. Improvement proposals to intrinsically motivational robotics. In *Proceedings of the second joint conference (ICDL-Epirob'12)* (2012).
- [8] HERVOUET, F., AND BOURREAU, E. Fimo: Framework for intrinsic motivations. In *Advances in Artificial Life, ECAL 2013: Proceedings of the Twelfth European Conference on the Synthesis and Simulation of Living Systems, Taormina – Sicily, Italy September 2-6, 2013.* (2013), pp. 997–1004.
- [9] KOZA, J. R. *Genetic programming: on the programming of computers by means of natural selection*, vol. 1. MIT press, 1992.
- [10] LUNGARELLA, M., METTA, G., PFEIFER, R., AND SANDINI, G. Developmental robotics: a survey. *Connection Science* 15 (2003), 151–190.
- [11] NELSON, A. L., BARLOW, G. J., AND DOITSIDIS, L. Fitness functions in evolutionary robotics: A survey and analysis. *Robotics and Autonomous Systems* 57, 4 (2009), 345 – 370.
- [12] OUDEYER, P.-Y., AND KAPLAN, F. Intelligent adaptive curiosity: a source of self-development. In *Proceedings of the 4th International Workshop on Epigenetic Robotics* (2004), L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, Eds., vol. 117, Lund University Cognitive Studies, pp. 127–130.
- [13] ROLF, M. Goal babbling with unknown ranges: A direction-sampling approach. In *IEEE Int. Conf. on Development and Learning and on Epigenetic Robotics (ICDL-Epirob)* (Osaka, Japan, 2013), pp. 1–7.
- [14] ROLF, M., STEIL, J. J., AND GIENGER, M. Goal-babbling permits direct learning of inverse kinematics. *IEEE Trans. Autonomous Mental Development* 2, 3 (2010).
- [15] SCHEMBRI, M., MIROLLI, M., AND BALDASSARRE, G. Evolution and learning in an intrinsically motivated reinforcement learning robot. In *Advances in Artificial Life* (2007), Springer, pp. 294–303.
- [16] SINGH, S., LEWIS, R. L., BARTO, A. G., AND SORG, J. Intrinsically motivated reinforcement learning: An evolutionary perspective. *Autonomous Mental Development, IEEE Transactions on* 2, 2 (2010), 70–82.
- [17] THELEN, E., CORBETTA, D., AND SPENCER, J. P. Development of reaching during the first year: role of movement speed. *Journal of Experimental Psychology: Human Perception and Performance*. 22, 5 (1996), 1059–1076.