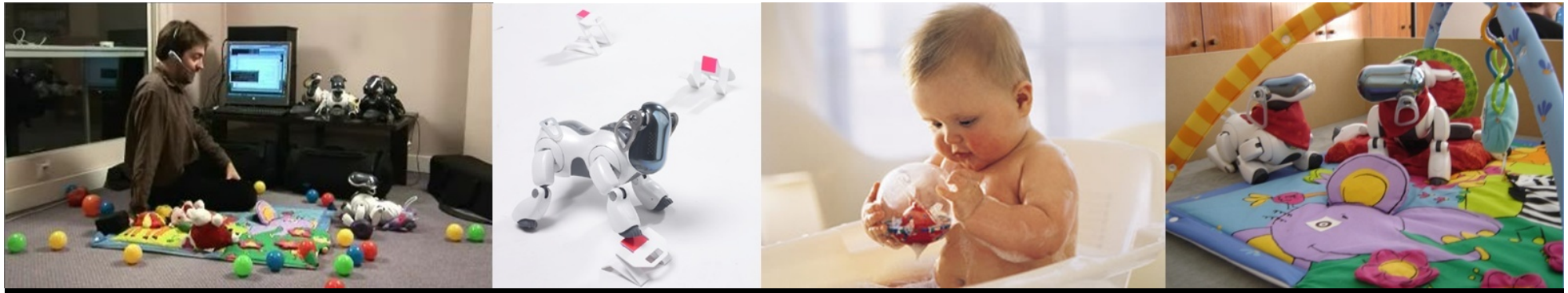


Developmental constraints on active learning for the acquisition of motor skills in high-dimensional robots

Pierre-Yves Oudeyer
INRIA

http://www.pyoudeyer.com

Developmental and social robotics: acquiring new skills « in the wild »

➔**The central target of developmental robotics** is to build machines, which once they are "out of the factory" and arrived "in the wild", are capable of learning by themselves or through interaction with humans a variety of skills and know-how that were not specified at design time;

➔ Exactly what human children are capable to do starting from their innate capabitlites;

➔ Import, formalize, implement and experiment concepts and theories of developmental and social psychology, developmental neuroscience, and linguistics

The challenges of exploration

• The scientific challenges of robot learning do ***not only*** consist in devising powerful (statistical) inference mechanism for building world models or sensorimotor control policies from training data (or rewards);

• Another central issue is to understand ***what kind of training data*** one should consider, ***how it should be encoded/represented***, and extremely importantly ***how it should/can be collected***;

• Most typically, training data will be collected by the robot itself (as opposed to hand prepared by an engineer) through self-experimentation and learning by observation: This takes a lot of time !

➔ Completely impossible to learn all the sensorimotor skills physically possible and learnable during a life-time due to ***HUGE*** (infinite?) size of sensorimotor spaces characterizing the body and its interaction with the external environment;
➔Even for a single kind of motor activity (e.g. playing tennis) life is not long-enough so that we learn everything that is possible to do with one own's body and its interaction with objects (e.g. tennis racket and tennis ball);

➔ How to explore the world in order to learn at least correctly a reasonable collection of motor skills? Obviously random exploration will fail.

Strategy 1: try to avoid the need for exploration

•When one wants a robot to learn a specific task (e.g. learning to walk forward as fast as possible) and allow the engineer to encode a specific reward/target function for this task …

• AND when one allows oneself to make certain assumptions on the analytic form availability of sensorimotor policies and properties of reward function,

➔ Then there are elegant mathematical workarounds (e.g. NAC) that allow us to compute robustly gradients from limited data, hence find good solutions to the problem from relatively little data;

$R(S,A) =$ forward speed of robot

See e.g. Peters and Schaal, 2008; Bhatnagar et al., 2009; Sutton et al., 2009, Theodorou et al., 2010, …

➔ But still for complex motor skills exploration is going to be an issue, and those assumptions might not always be desirable.

Strategy 2: Guide and constrain exploration

• When one is interested in learning fields of motor primitives (e.g. not just walking forward, but all kinds of directions and rotations), or even various kinds of motor skills (e.g. walking + navigating + shooting in balls …), …

• AND even more when one does not allow the engineer to program a specific reward function for each of these tasks, and one does not necessarily want to restrict oneself to motor policies that shall respect certain analytical properties;

➔Then sensorimotor spaces become even bigger and exploration central;

➔ Exploration needs to be guided and constrained in a way that is as task-independent as possible (i.e. one should try to find constraints that are minimally task specific);

➔ Take inspiration from developmental constraints in the human child;

An innate cerebral and morphological equipment …

Innate motivational system that fosters spontaneous BUT organized exploration (intrinsic motivation/curiosity-driven exploration)

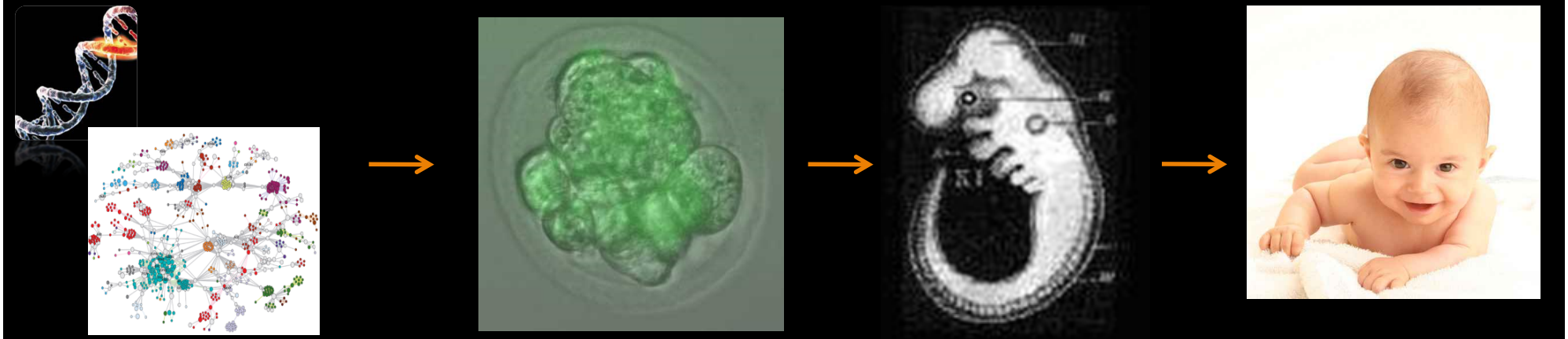
Motor primitives that constrain the space of motor commands and gestures: e.g. muscles are not controlled individually and independently, oscillators, …

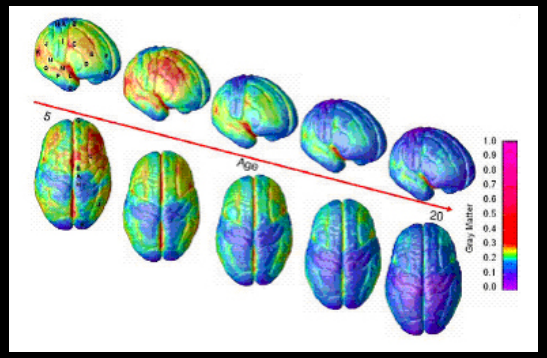
Sensori detectors and trackers that allow the baby to bootstrap its attentional and emotional systems: e.g. movement, high pitch, faces, …

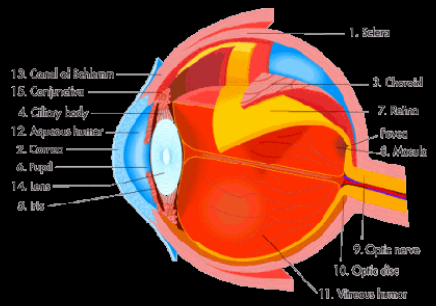
Sensorimotor reflexes: e.g. eye tracking of moving objects, closing hands when objects touched, …

Morphological properties that facilitate the control of the body, …

… built within a maturational program …







e.g. myelination/myelinogenesis progressively building brain regions, connecting them together and to muscles, increasing progressively resolution of senses and motor control, …

… and continuously extended thanks to a generic learning and developmental system



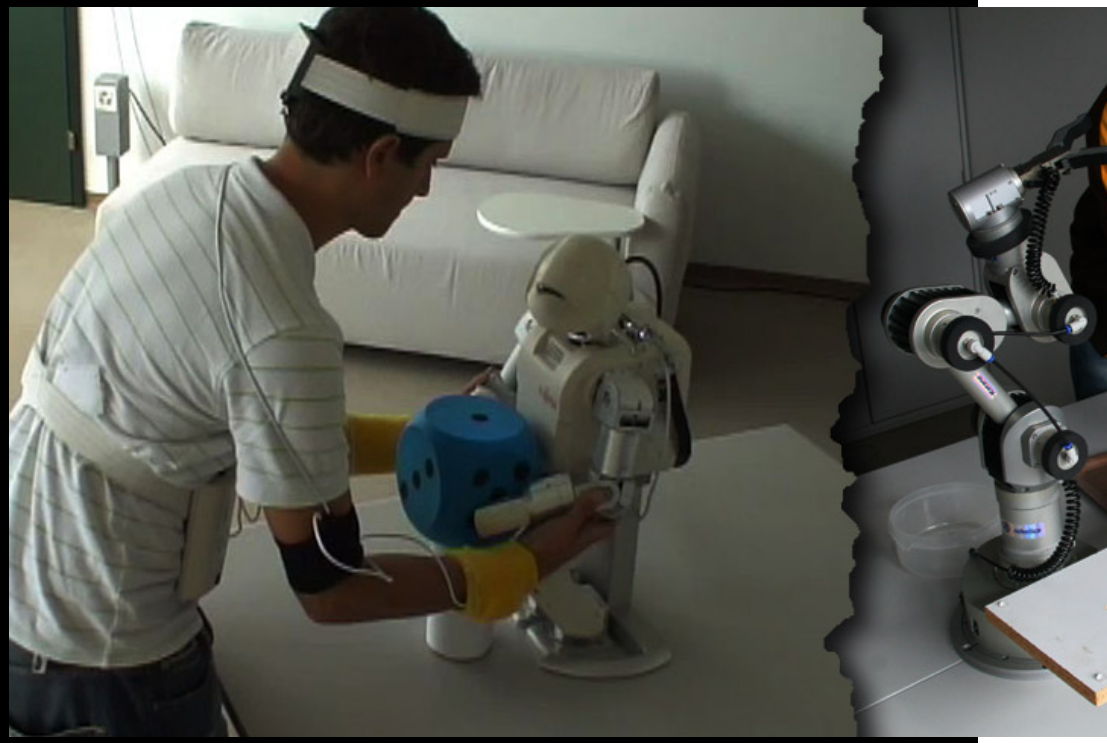








Social guidance: Learning by imitation/ observation



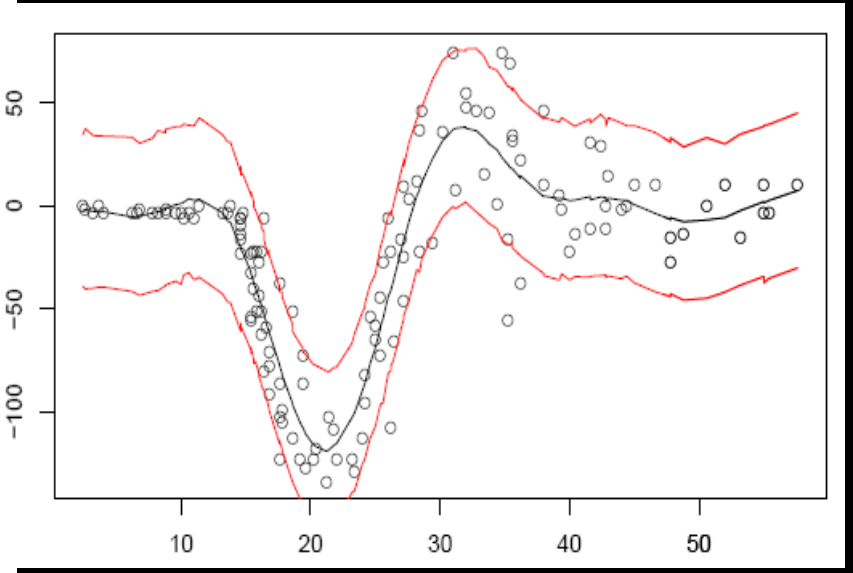
(S. Calinon)

Internal mechanisms that ***directly*** foster spontaneous ***exploration for its own sake***

➔ **INTRINSIC MOTIVATION**

Intrinsically motivated reinforcement learning/ Active learning/Optimal experimental design

Y



f

$(x_1, y_1) \rightarrow \text{model 1}$
$(x_2, y_2) \rightarrow \text{model 2}$
$(x_3, y_3) \rightarrow \text{model 3}$
$\vdots$
$(x_n, y_n) \rightarrow \text{model n}$

X

➔ Which x_{n+1} to experiment?

- **A mapping to learn X → Y** from {(xi,yi)} examplars, where,
  X can be state(t) x action(t) or just action(t)
  Y can be state(t+1)

- **A function of *I(xi)* is defined which measures the "interest"** of getting the yi associated to xi (heuristically or optimally with respect to various criteria related to information gain)
- **Action selection:**

$$x_{choosen} = argmax_{x_i \in X} \sum_{t=n+1}^{\infty} \gamma^t \tilde{I}(x_i)$$

➔ I(xi) is a reward and RL can be used, allowing to address delayed rewards

➔ In both cases, (meta-)exploitation-(meta) exploration dilemna to be addressed

Most frequent measures of “interest”, i.e. heuristic measures of information gain

- Places where we have little data (e.g. Whitehead, 1991);
- Places where prediction errors are high (e.g. Linden and Weber, 1993; Thrun, 1995);
- Places where we have low confidence, or with highest uncertainty (e.g. Thrun and Moller, 1992);
- Places where the variance of is maximal;
- Places where the entropy of is maximal;

…

- **in RL:** Counter-based, recency-based, novelty-based, « exploration bonuses »

(Sutton, 1990; Brafman and M. Tennenholtz, 2001; Strehl et al., 2006; Szita and Lorincz, 2008, …)

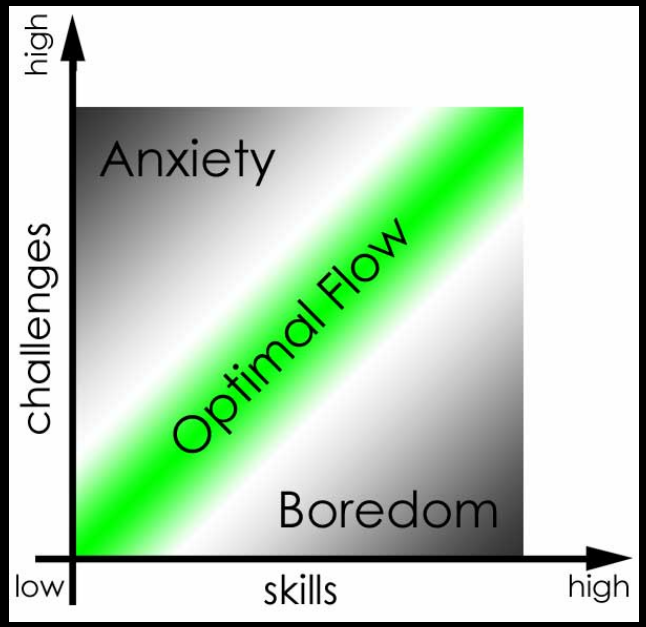
These measures typically make at least one of the following assumptions:
1) It is possible to learn a complete model of the world during the life-time of the learning agent;
2) The world is learnable everywhere;
3) Noise is homogeneous;

➔ These assumptions do not hold in many real-world robotic (= non-toy) sensorimotor spaces (same as for human infants);

➔ These measures are inoperant in these real-world sensorimotor spaces

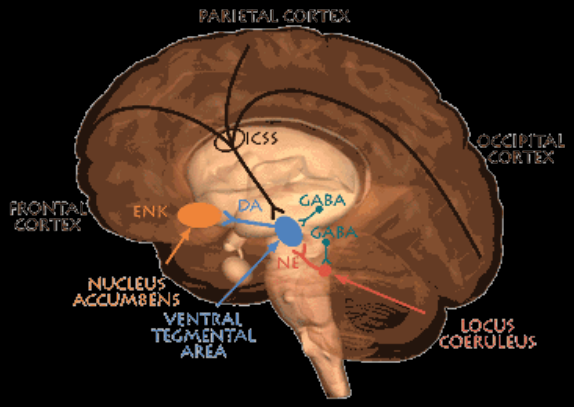
Curiosity driven learning in humans: the search for intermediate complexity

Developmental psychology



White (1959), Berlyne (1960), Csikszentmihalyi (1996)

Neurosciences



Dayan and Belleine (2002), Kakade and Dayan (2002), Horvitz (2000)

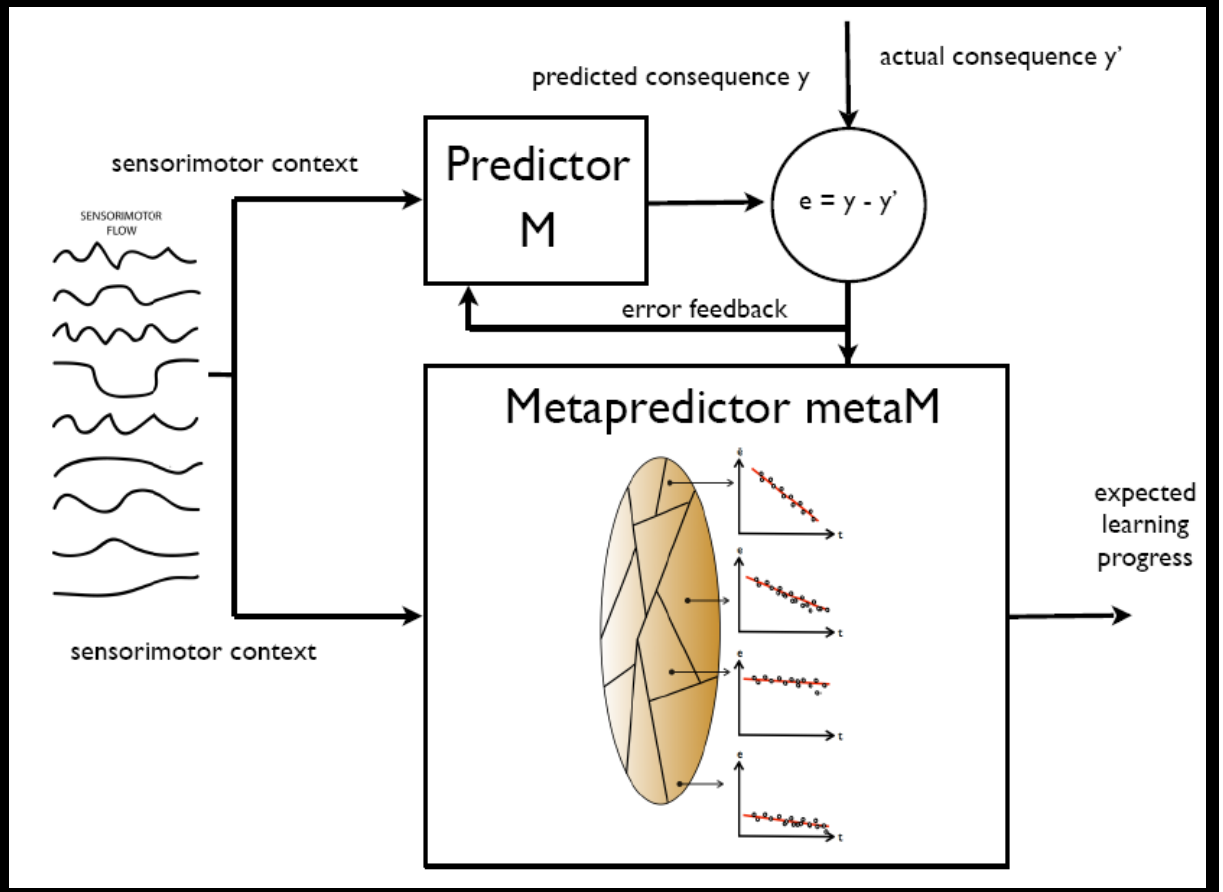
➔Activities of intermediate complexity are intrinsically rewarding

➔ **Mechanisms for regulating the growth of complexity: the importance of starting small**

Active regulation of the growth of complexity in exploration

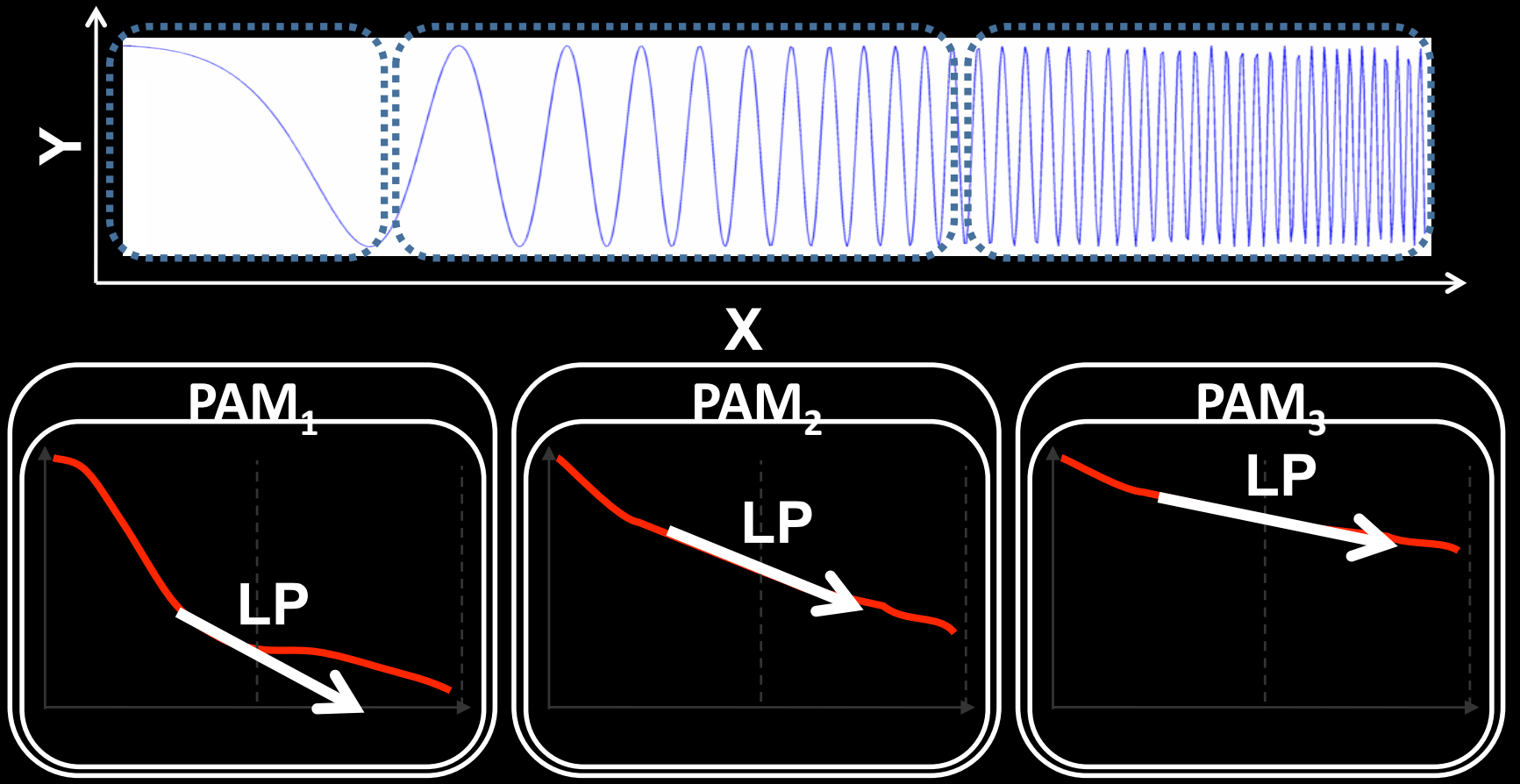
Optimizing learning ***progress***, i.e. the decrease of prediction errors (derivative)

The IAC/R-IAC (Intelligent Adaptive Curiosity) architecture(s)



Oudeyer P-Y, Kaplan , F. and Hafner, V. (2007), Baranes and Oudeyer (2009, 2010)
See also: Schmidhuber (1991, 2006)

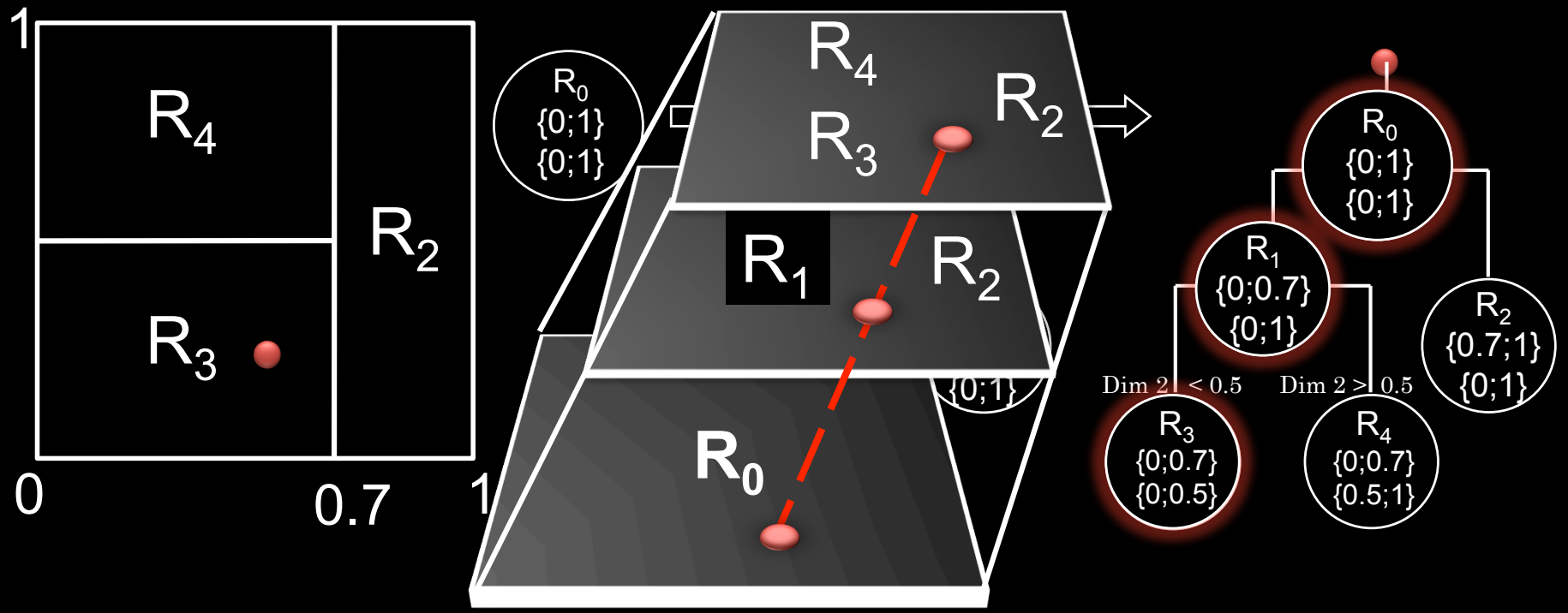
R-IAC: multi-resolution probabilistic region-based learning progress



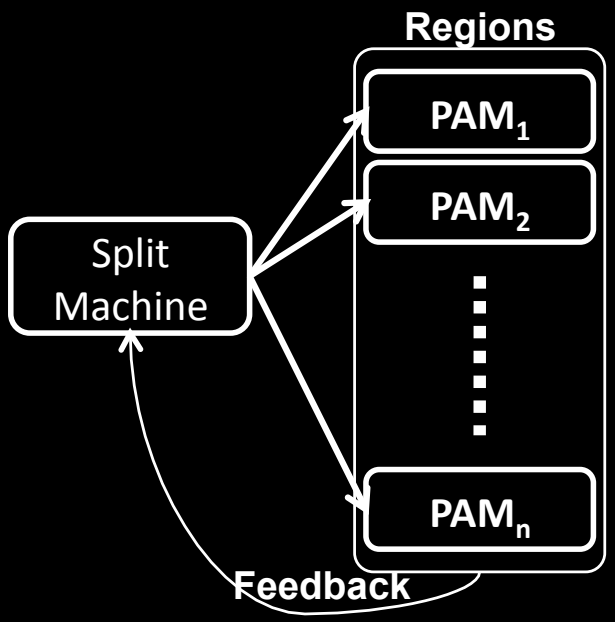
Learning Progress = decrease of mean prediction errors in a region

(Baranes and Oudeyer, 2009)

R-IAC: recursive multi-resolution region splitting



R-IAC: optimized splitting mechanisms



Maximization of dissimilarity of learning progress

$\varphi_n = \left\{ \left(\mathbf{SM}(t), \mathbf{S}(t+1) \right)_i \right\}$ for each region R_n

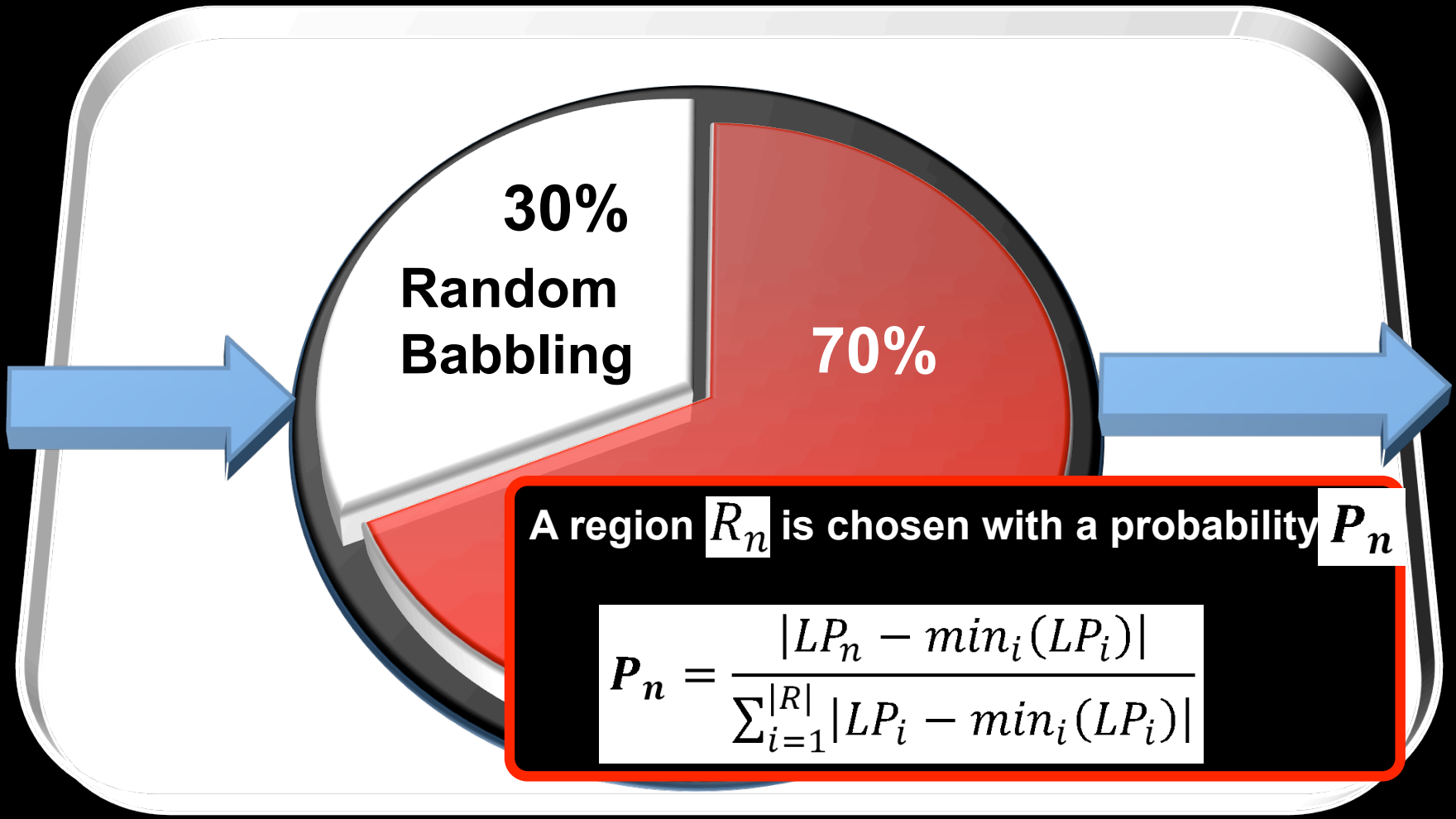
j cutting dimension

v_j associated cutting value

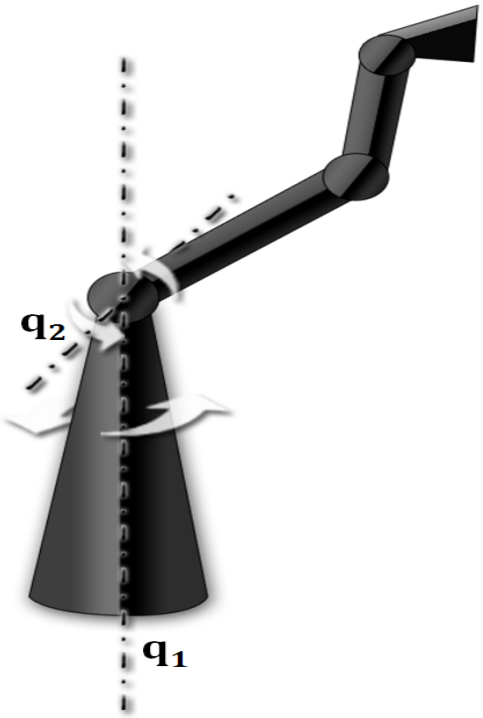
$$Qual(j, v_j) = - \begin{matrix} LP_{n+1}\left(\left\{\mathbf{e}(t+1) | \left(\mathbf{SM}(t), \mathbf{S}(t+1)\right) \in \varphi_{n+1}\right\}\right) \\ LP_{n+2}\left(\left\{\mathbf{e}(t+1) | \left(\mathbf{SM}(t), \mathbf{S}(t+1)\right) \in \varphi_{n+2}\right\}\right) \end{matrix}$$

Where the Learning Progress $LP_k(E) = \dfrac{\sum_{i=1}^{\frac{|E|}{2}} e(i) - \sum_{i=\frac{|E|}{2}}^{|E|} e(i)}{|E|}$

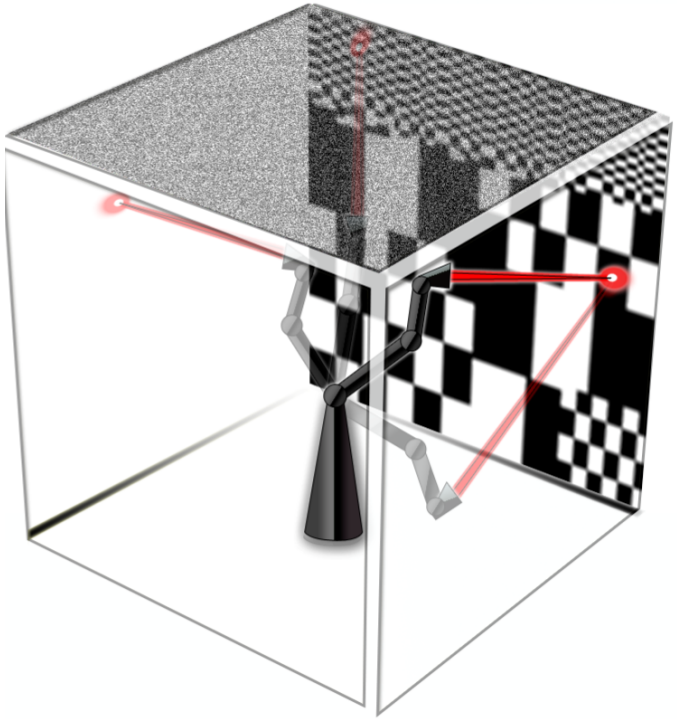
R-IAC: multi-mode probabilistic experiment selection



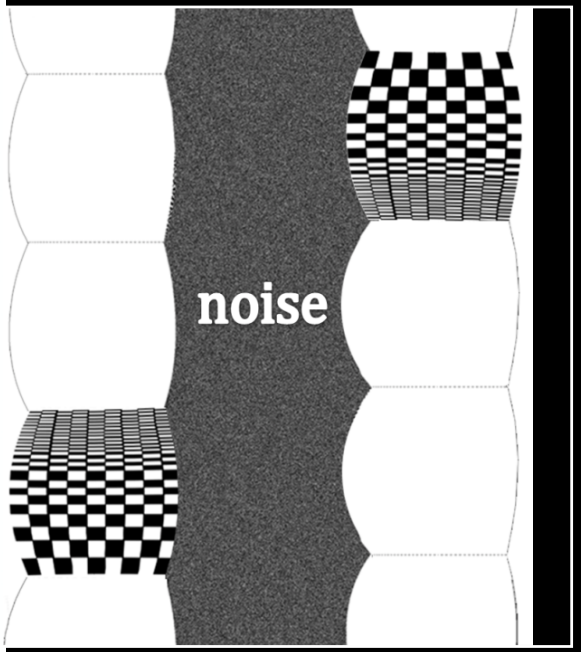
Example in a (not so) simple experiment



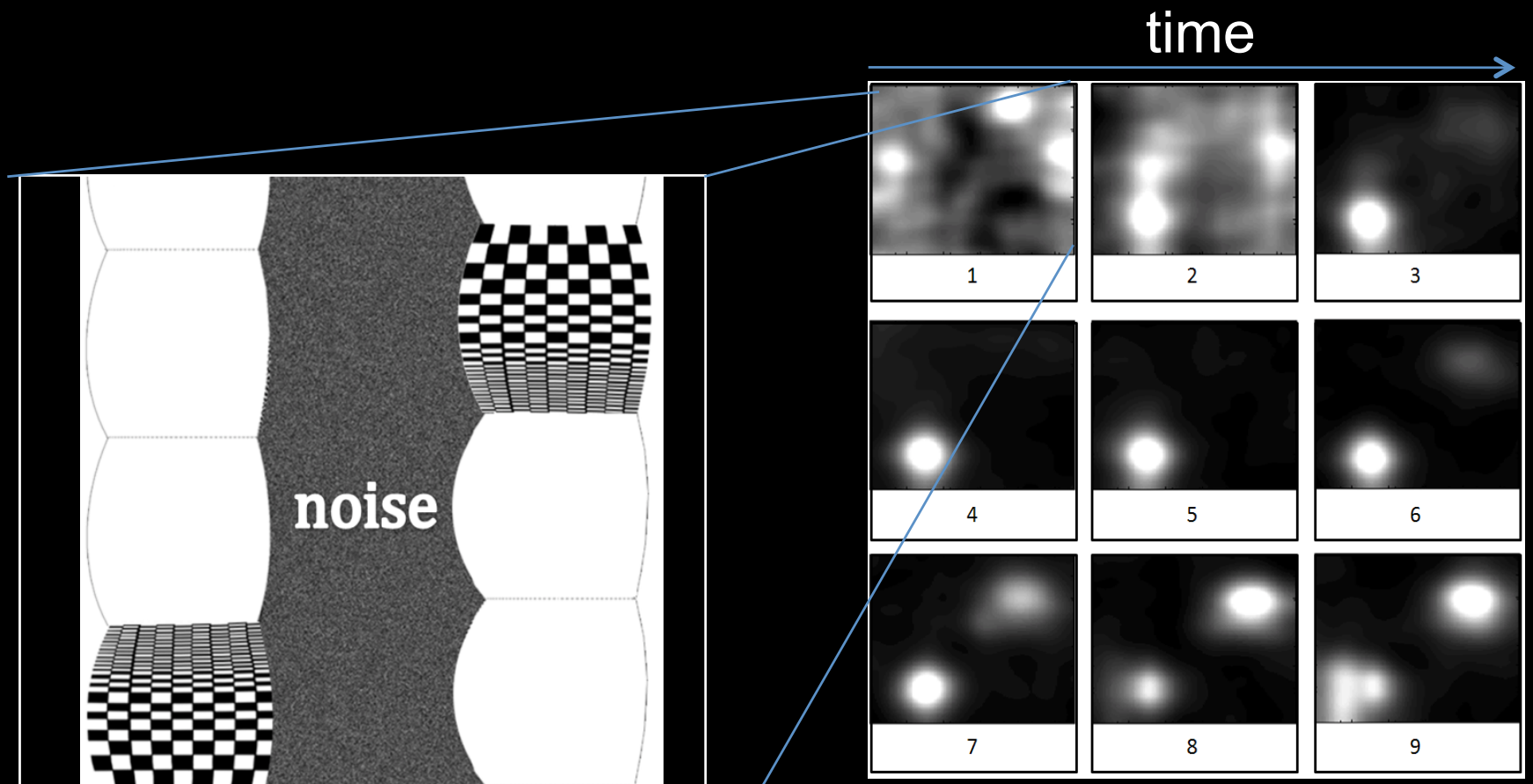
2 DOF redundant robotic arm, with a 1-pixel camera



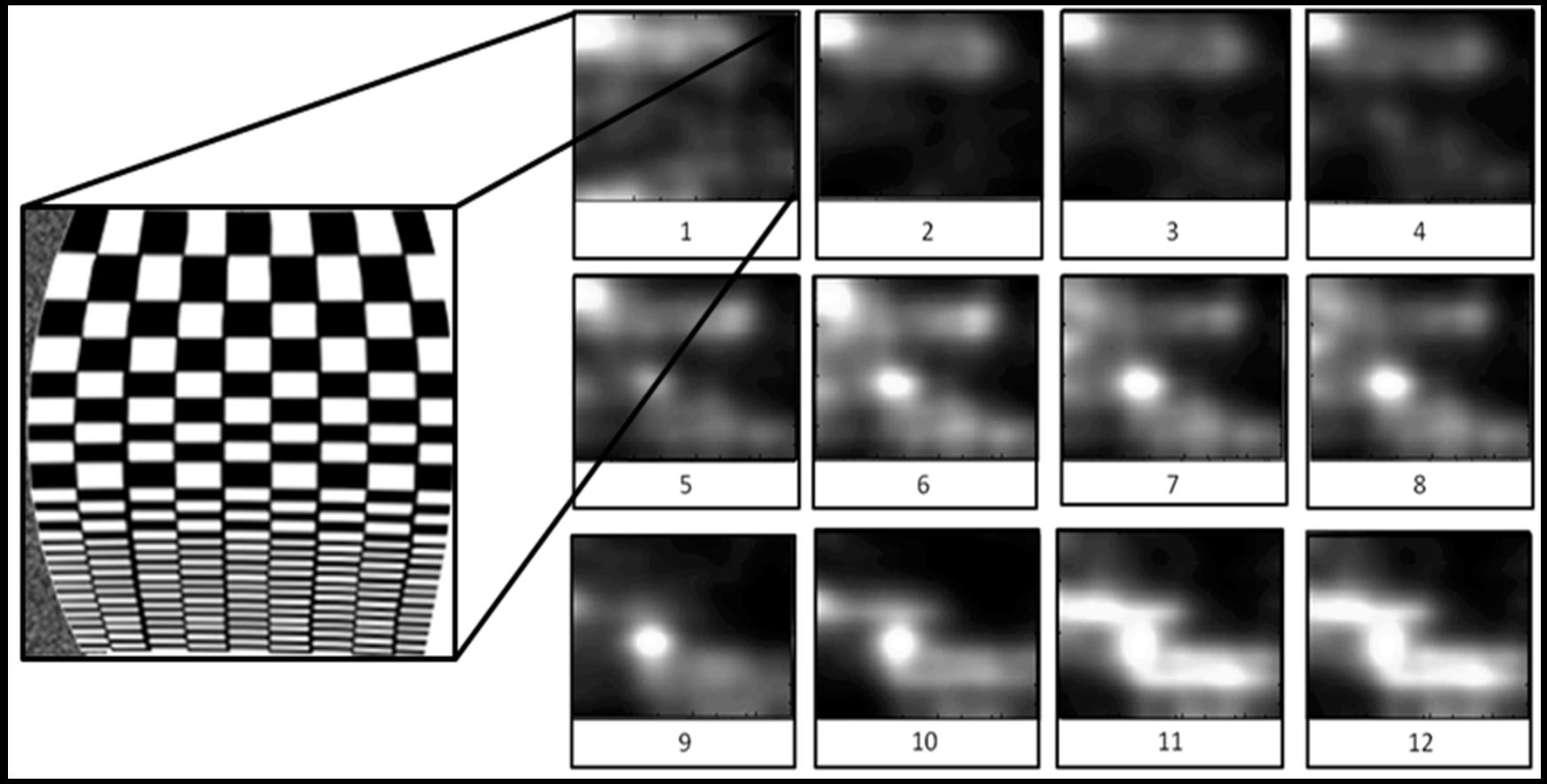
An inhomogeneous space to be explored



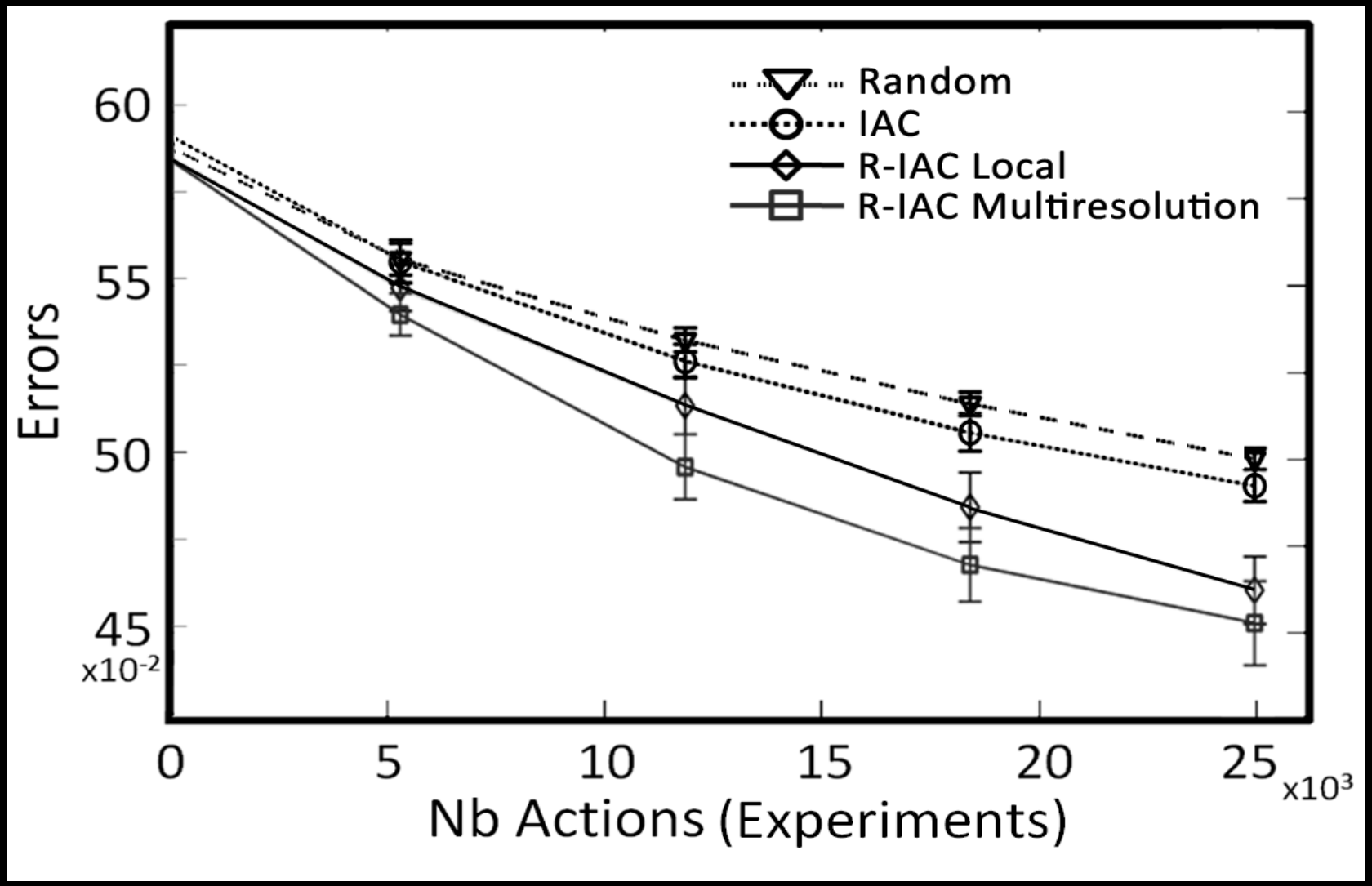
Visualization of the mapping to be learnt



Evolution of exploration focus with R-IAC



Zoomed in exploration focus with R-IAC



(Baranes and Oudeyer, 2009, IEEE Transactions on AMD)

The problem of meta-exploration of « interestingness » in large spaces

• R-IAC like exploration allows to avoid spending too much time on unlearnable or trivial subspaces, and fosters a focus on zones of progressively increasing complexity

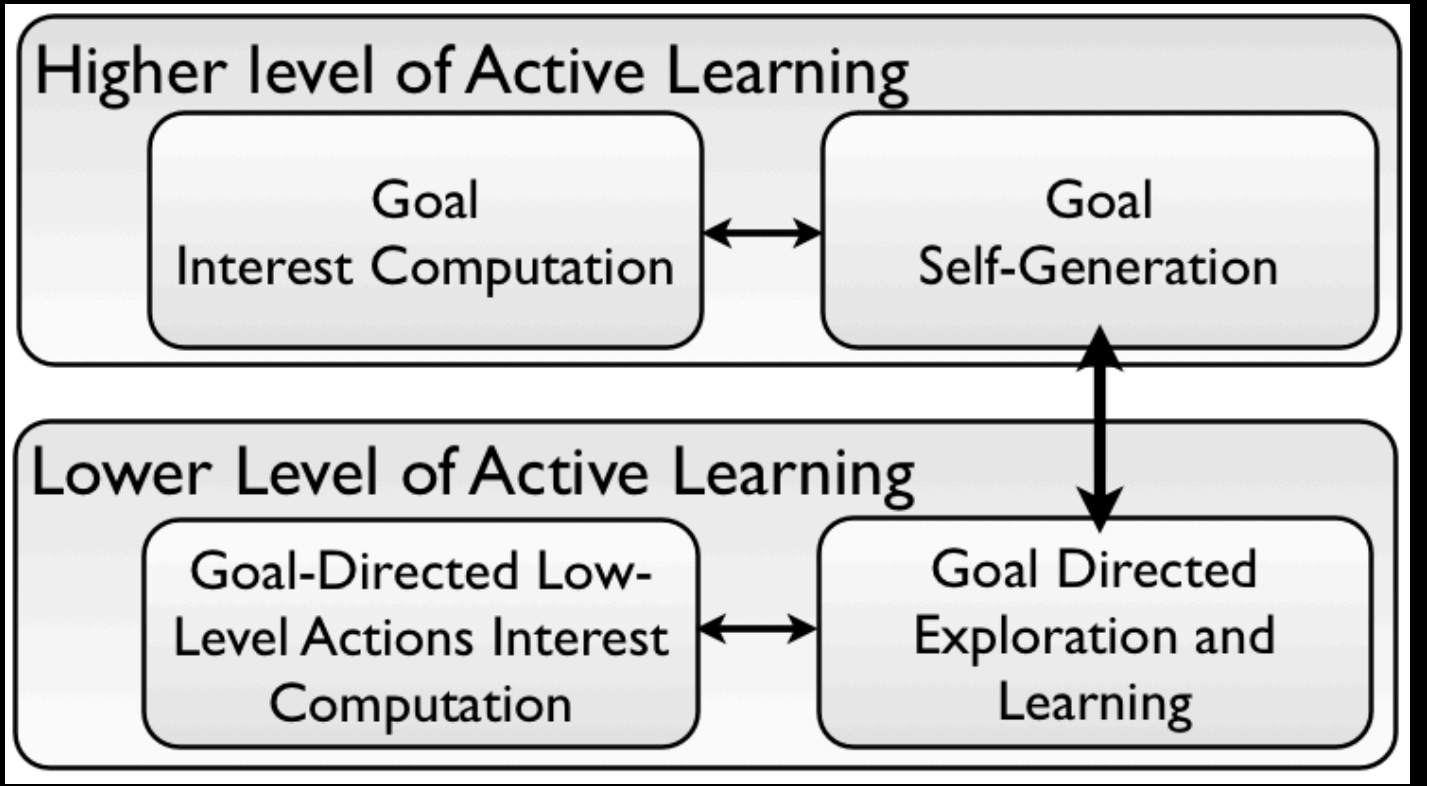
• BUT assessing I(x) still requires a certain amount of exploration in the vicinity of x !

➔ We have a (better but still problematic) meta-exploration problem!

➔ Further constraints on meta-exploration for curiosity-driven learning are needed;

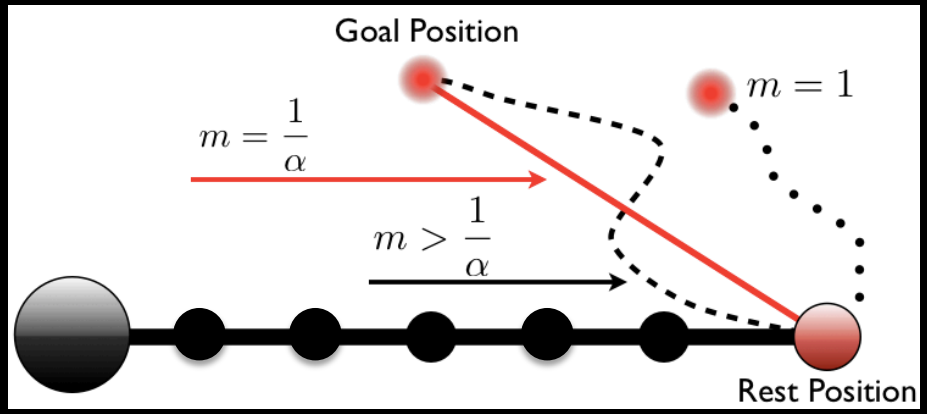
Active learning in the (operational) space of goals (typically much smaller)

Multi-level
Active learning
(SAGG algorithm)



Baranes, A., Oudeyer, P-Y. (2010) Intrinsically Motivated Goal Exploration for Active Motor Learning in Robots: a Case Study, inProceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), Taipei, Taiwan.

Example: learning the inverse kinematics of an n-DOFs arm



SAGG

Algorithm 2 Global Pseudo-Code of SAGG

Input: Current Position: x_0, θ_0
loop
  Goal Self-Generation:
  Selection of a Region R_i according to Interest Values
  Generation of a Random Goal x_g in R_i
  Generation of Sub-Goal x_i
  Goal Directed Learning:
  for each sub-goal, and goal x_j **do**
    while Constraints unrespected and $x_0 \neq x_j$ **do**
      Exploration Phase
      Reaching Phase
    end while
    Interest Update:
    Effectiveness Computation
    Interest Computation
    Region Update
  end for
end loop

Active goal self-generation

(SSA, Schaal et Atkeson, 1994)

Active goal directed learning

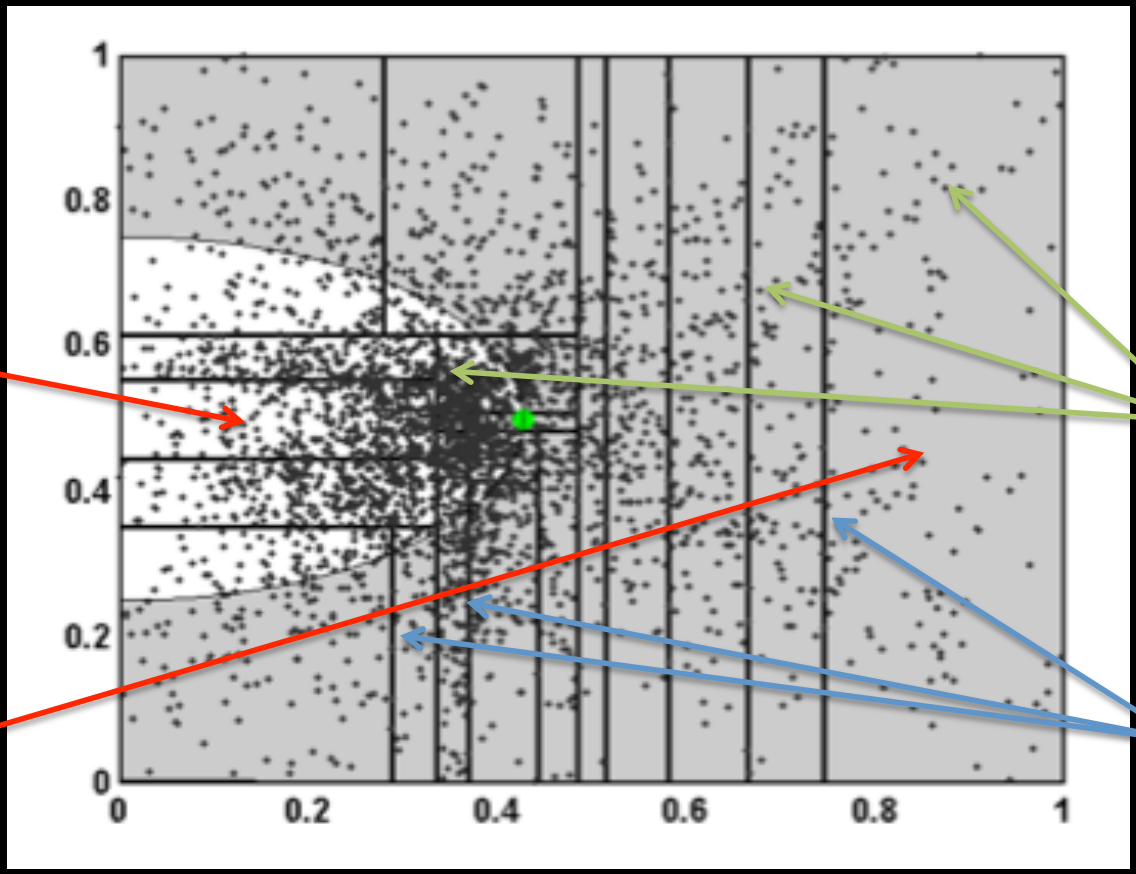
Algorithm 1 Goal Directed Learning

Input: x_g, x_0, θ_0, $m = 0$, m_{max}, ν, γ, ε_{max}
Output: m
while $x_0 \neq x_g$ and $m < m_{max}$ **do**
  $\Delta x_{next} = \nu . \frac{x_0 - x_{goal}}{\|x_0 - x_{goal}\|}$
  J =**get_current_Jacobian**(θ_0)
  $\widetilde{\Delta x_{next}}$ =**move**$(\Delta\theta_{next} = J^{+}.\Delta x_{next})$
  $\varepsilon = \|\widetilde{\Delta x_{next}} - \Delta x_{next}\|$
  if $\varepsilon > \varepsilon_{max}$ **then**
    move$(-\Delta\theta_{next})$
    for $i = 1 : \gamma$ **do**
      $\Delta\theta_{next}$ = **random**
      move$(\Delta\theta_{next})$, **move**$(-\Delta\theta_{next})$
    end for
  end if
end while

Maximizing *competence progress* for active goal exploration allows to focus on *reachable* areas (initially unknown)

Reachable area (in white)

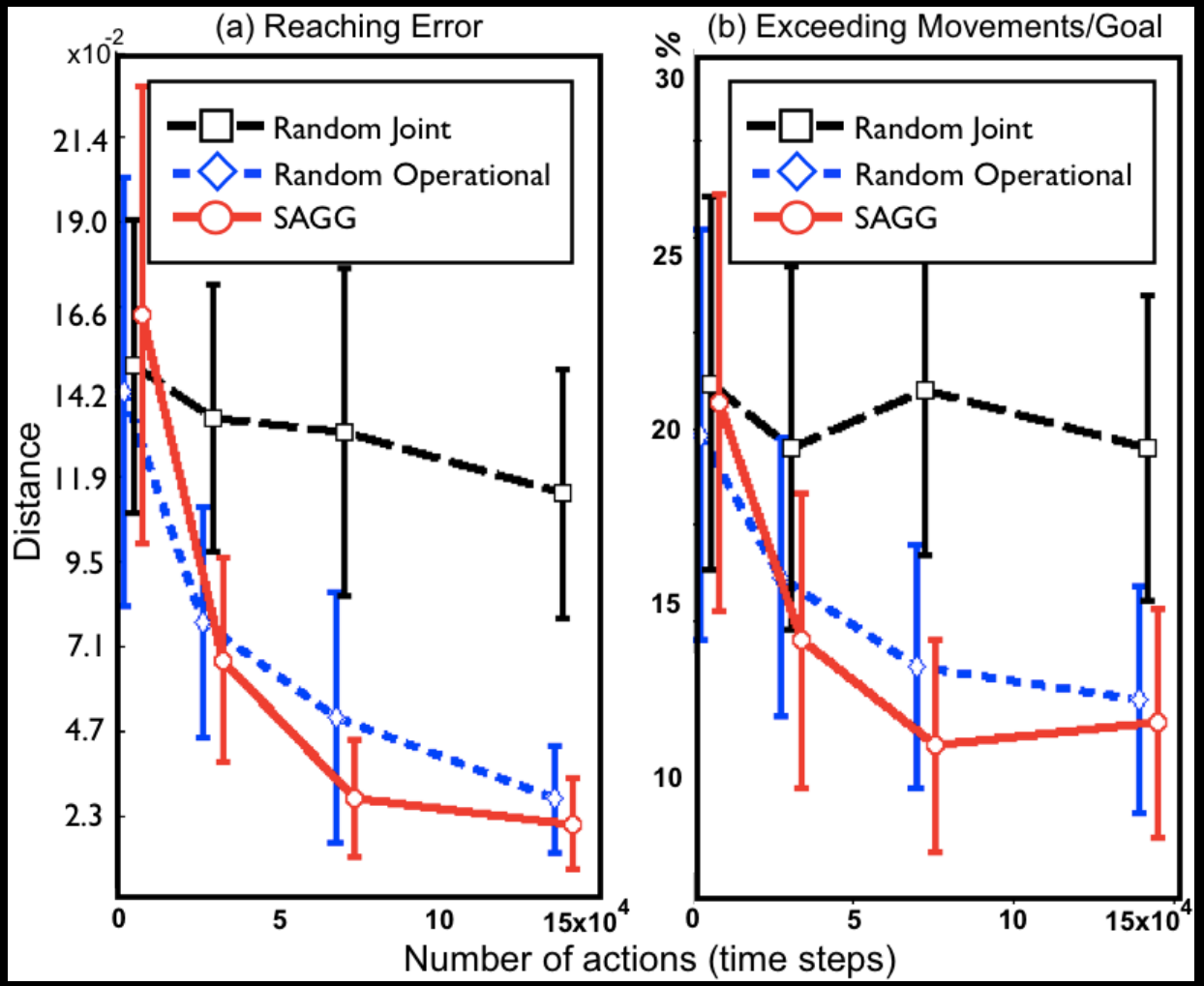
Uneachable area (in grey)



Self-generated goals (dots)

Region splits built by SAGG

Orders of magnitude faster than random exploration and only active exploration in joint space



Developmental constraints on exploration: 1) Motor primitives

Biological organisms CNS do not control muscles individually and at a very low-level, but rather parameters of higher level primitives that encode ***muscular synergies;***

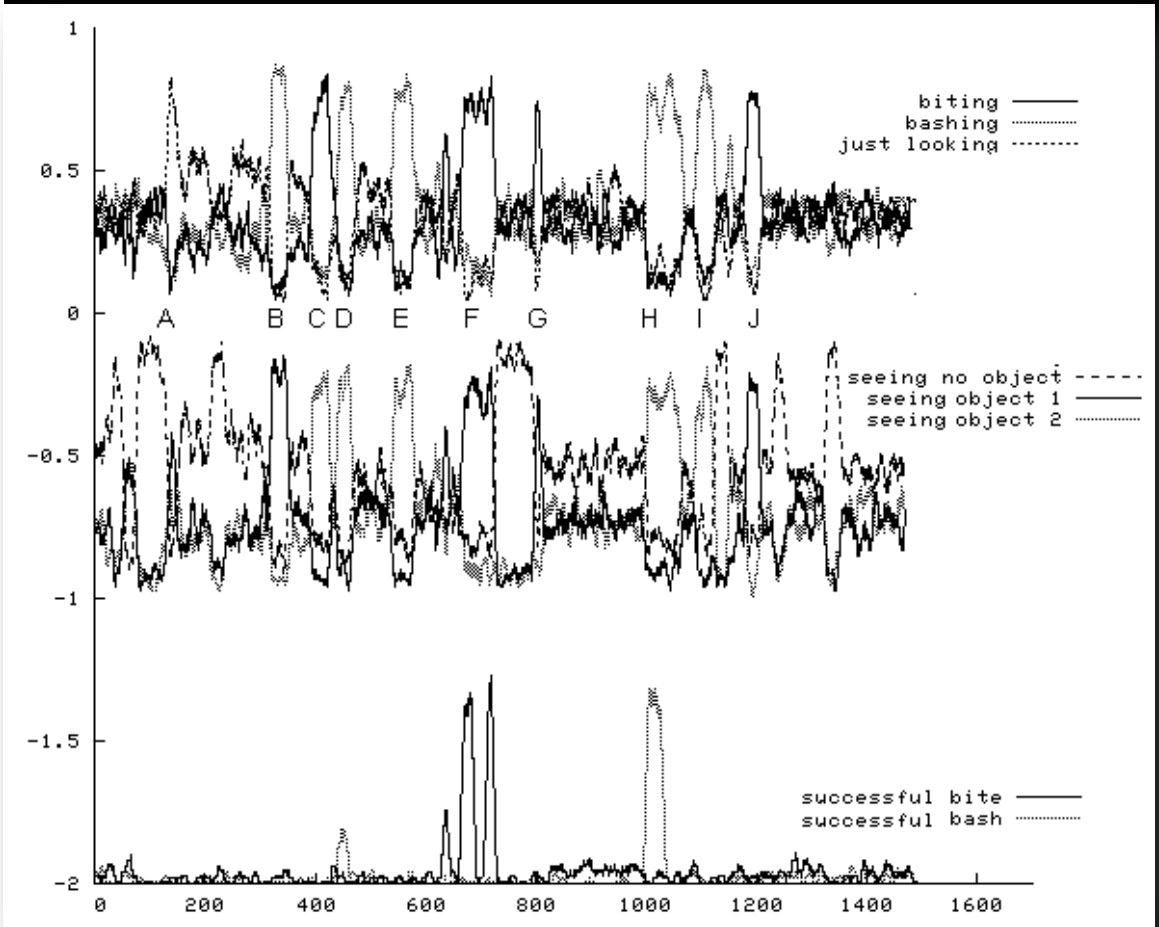
These primitives are often conceived as parameterized dynamical systems;

e.g. CPG, oscillators

The Playground experiment

http://playground.csl.sony.fr

(Oudeyer, Kaplan, Hafner, 2007, IEEE Trans. Evol. Comp.)



Self-organization of developmental patterns

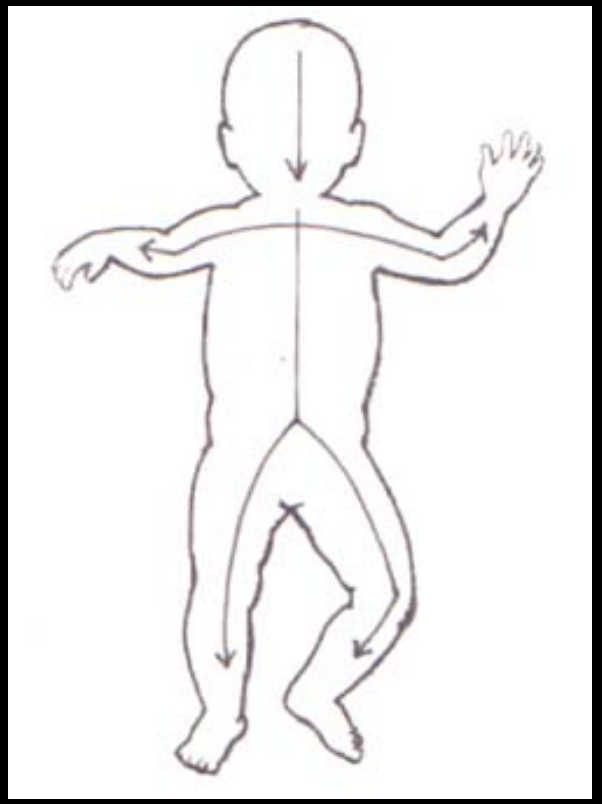
Measure 1 (number of peaks?)	9.67
Measure 2 (complete scenario?)	Yes: 34 %, **No: 66 %**
Measure 3 (near complete scenario?)	**Yes: 53 %**, No: 47%
Measure 4 (non-affordant bite before affordant bite?)	**Yes: 93 %**, No: 7 %
Measure 5 (non-affordant bash before affordant bash?)	**Yes: 57 %**, No: 43 %
Measure 6 (period of systematic successful bite?)	**Yes: 100 %**, No: 0 %
Measure 7 (period of systematic successful bash?)	**Yes: 78 %**, No: 11 %
Measure 8 (bite before bash?)	**Yes: 92 %**, No: 8 %
Measure 9 (successful bite before successful bash?)	**Yes: 77 %**, No: 23 %

Developmental constraints on exploration: 2) Maturation

An important aspect of the maturation of the neural system is the myelination process which only progressively allows the infant's brain to control new muscles.

The corticospinal tract is not functional at birth, but develops extensively over the first year, in a proximo-distal and cephalo-caudal pattern, leading to a gradual development of the infant's ability to control the distal musculature of the arm and hand (Berthier et al., 1999).

For example, in the reaching task, if young infants predominately use the musculature of the proximal arm and trunk, the learning problem become much simpler with the reduction in the functional degrees-of-freedom of the arm.



➔The **MAC-SAGG algorithm**

Baranes, A., Oudeyer, P-Y. (2010) Maturationally-Constrained Competence-Based Intrinsically Motivated Learning, in Proceedings of IEEE International Conference on Development and Learning (ICDL 2010), Ann Arbor, Michigan, USA.

Modeling maturation and its interaction with intrinsic motivation

Maturational clock where maturational time increases as overall competence/ quality of predictions increases

$$\psi(t+1) = \begin{cases} \psi(t) + \lambda.interest(S') & \text{if } interest(S') > 0 \\ \psi(t) & \text{otherwise} \end{cases}$$

Which then controls the growth of:

Time resolution of motor impulses

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

Sensori resolution for state estimation

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

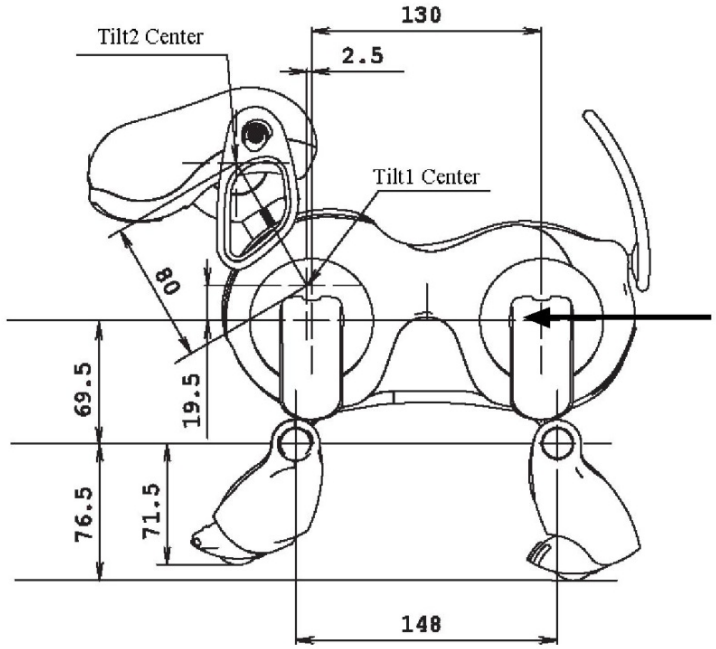
Volume/range of explorable values in motor channels, with proximo-distal law

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

Where k_i represents an intrinsic value determining the difference of evolution velocities between each joint. Here we fix: $k_1 \geq k_2 \geq ... \geq k_n$, where k_1 is the first proximal joint.

Baranes, A., Oudeyer, P-Y. (2010) Maturationally constrained competence based intrinsically motivated learning, in *Proceedings of IEEE ICDL 2010.*

Example: Developmental learning of locomotion with R-IAC, motor primitives and maturational constraints



$$j(t) = m + a.\sin(\omega.t + \phi)$$

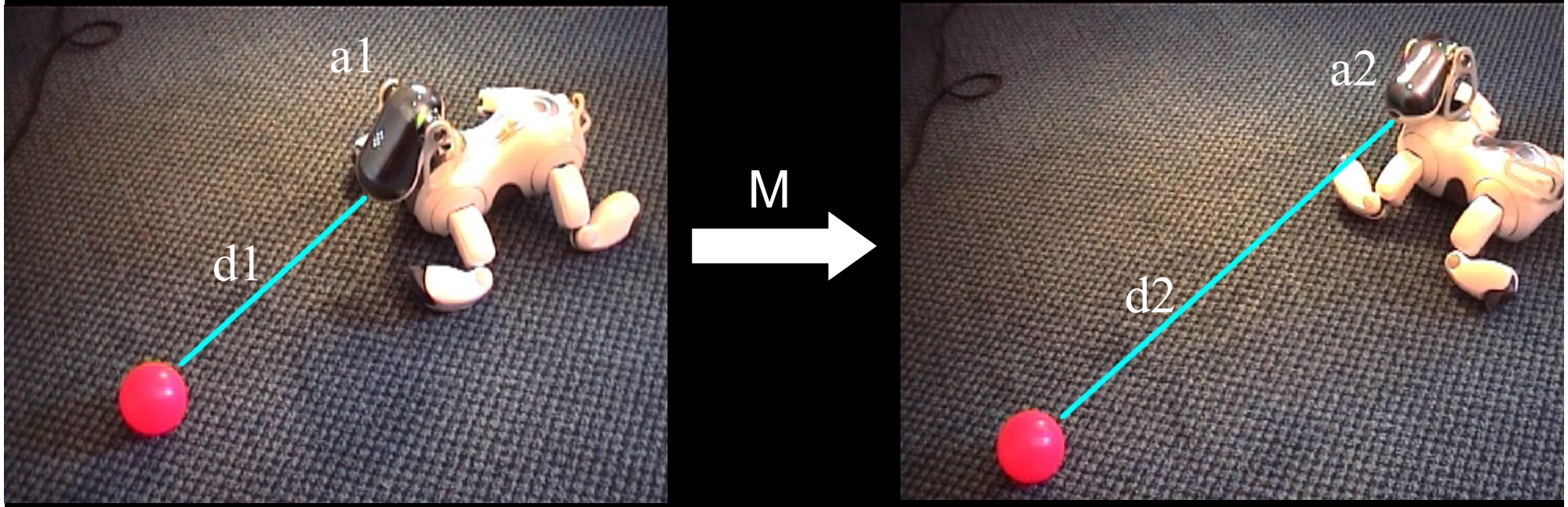
8 joints * 3 parameters

=

Motor vector M with 24 dimensions

+ progressive increase of the range of accessible m, a, phi

Explore the consequence of one's movements



Initial position (d1, a1)

Final position (d2, a2)

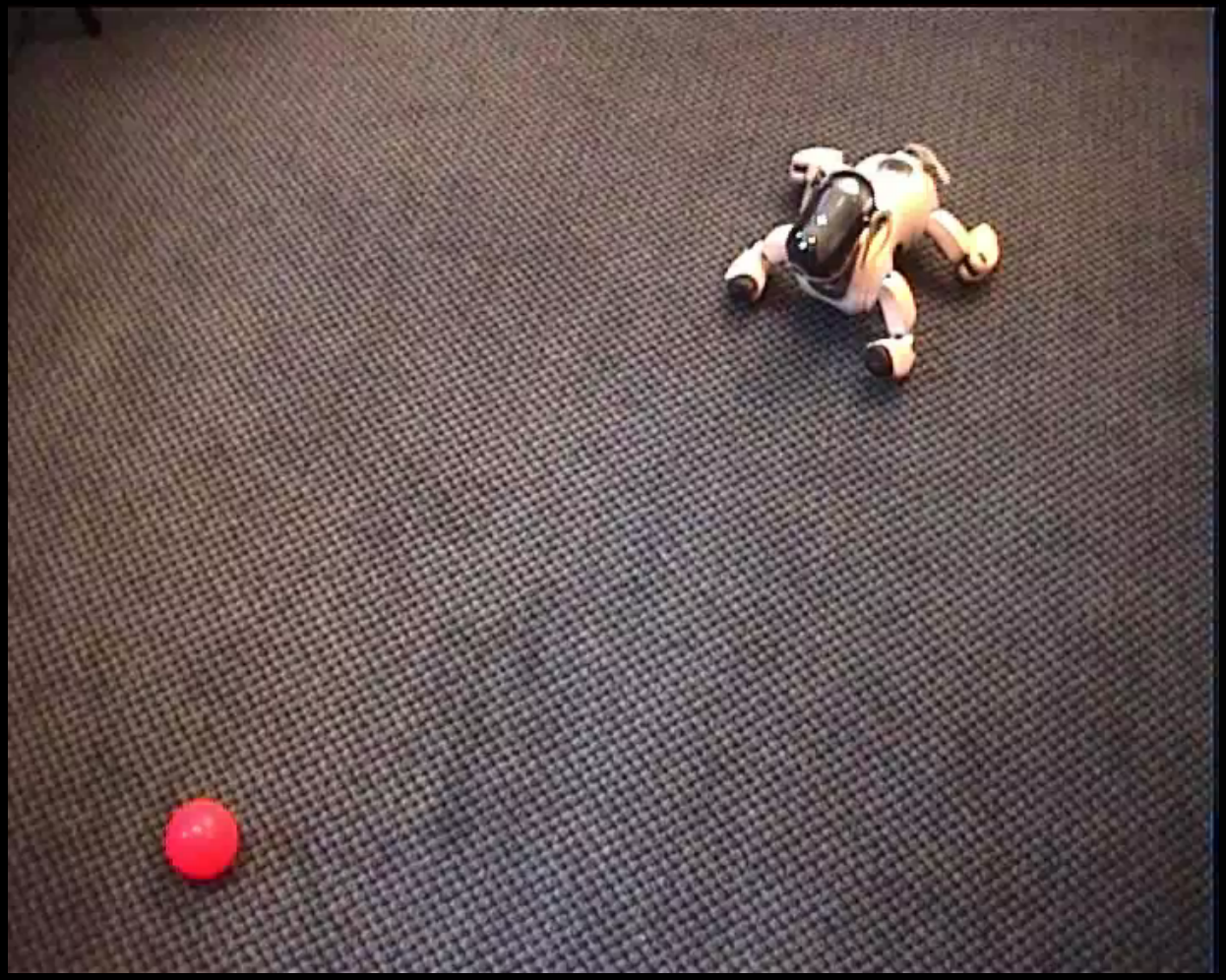
The robot tries to predict:

$f(d1, a1, M) = (d2 - d1, a2 - a1)$

Exploration trajectory

Learnt skills

The robot can re-use its curiosity-driven learnt forward and inverse models to reach any particular location in its field of view



Developmental constraints on exploration: morphological computation and semi-passive dynamics

- **Acroban robot (Olivier Ly), 32 DOFs**
- **Compliant structure that can absorb and store energy (elastic tendons, springs, motors);**
- **Semi-passive torso with a 5 DOFs vertebral column (triple pendulum) for both stabilization and transformation of potential energy into kinetic energy;**
- **Semi-passive feets;**
- **NO MODEL OF DYNAMICS;**
- **General purpose stabilizing motor primitive robust to perturbations;**
- **Walking as a self-perturbation, only 2 parameters for all movements !**
- **Human physical guidance for free !**

Ly, O., Oudeyer, P-Y. (2010) Acroban the humanoid: Playful and compliant physical child robot interaction, SIGGRAPH'2010 *Emergent Technologies. Videos on* http://flowers.inria.fr/acroban.php

Take home message

• Exploration is an essential issue for robot learning of repertoires of complex motor skills;

• Active learning/intrinsically motivated exploration mechanisms are of essential help;

• BUT they cannot alone allow robot to learn real non-trivial motor skills without additional constraints (ideally non task-specific), in particular developmental constraints (motor primitives, maturational constraints, morphological computation, …);

Thank you!

More info on http://www.pyoudeyer.com

References

J. Peters and S. Schaal, “Natural actor critic,” Neurocomputing, no. 7-9, pp. 1180–1190, 2008.

S.Bhatnagar, R.S.Sutton, M.Ghavamzadeh and M.Lee, Natural Actor-Critic Algorithms, Automatica, 2009.

Sutton, R. S., Maei, H. R., Precup, D., Bhatnagar, S., Silver, D., Szepesvari, Cs., Wiewiora, E. (2009). Fast gradient-descent methods for temporal-difference learning with linear function approximation. In Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada.

Theodorou, E., Buchli, J., Schaal, S. (2010). Reinforcement Learning of Motor Skills in High Dimensions: A Path Integral Approach. , *International Conference of Robotics and Automation (ICRA 2010) .*

S. Whitehead, A Study of Cooperative Mechanisms for Faster Re- inforcement Learning Univ. Rochester, Rochester, NY, Tech. Rep. TR-365.

S. Thrun and K. Möller, J. Moody, S. Hanson, and R. Lippmann, Eds., “Active exploration in dynamic environments,” in *Proc. Adv. Neural Info. Process. Syst. 4, Denver, CO, 1992.*

S. Thrun, “Exploration in active learning,” in *Handbook of Brain Science and Neural Networks, M. Arbib, Ed.Cambridge, MA: MIT Press, 1995, pp. 381–384.*

Linden, A. and Weber, F. Implementing inner drive through competence reflection. In Press, MIT (ed.), Proceedings of the second international conference on From animals to animats 2 : simulation of adap- tive behavior: simulation of adaptive behavior, pp. 321–326, 1993.

R. S. Sutton, “Integrated architectures for learning, planning, and re- acting based on approximating dynamic programming,” in *Proc. 7th Int. Conf. Mach. Learn., Washington DC, 1990, pp. 216–224.*

R. I. Brafman and M. Tennenholtz. **R-max - A General Polynomial Time Algorithm for Near-Optimal Reinforcement Learning. In *IJCAI`01, 2001.***

Alexander L. Strehl, Chris Mesterharm, Michael L. Littman, Haym Hirsh: Experience-efficient learning in associative bandit problems. ICML 2006: 889-896

Istvan Szita, Andras Lorincz. The many faces of optimism: a unifying approach. In Proceedings of ICML'2008. pp. 1048~1055

R.White,“Motivationreconsidered:Theconceptofcompetence,”*Psy-chol. Rev., vol. 66, pp. 297–333, 1959.*

D. Berlyne*, Conflict, Arousal and Curiosity.New York: McGraw-Hill, 1960.*

M.Csikszenthmihalyi,*Flow-thePsychologyofOptimalExperience.New York: Harper Perennial, 1991.*

References

P. Dayan and W. Belleine, “Reward, motivation and reinforcementlearning,” *Neuron, vol. 36, pp. 285–298, 2002.*

S. Kakade and P. Dayan, “Dopamine: Generalization and bonuses,”*Neural Netw., vol. 15, pp. 549–559, 2002.*

J.-C. Horvitz, “Mesolimbocortical and nigrostriatal dopamine re-sponses to salient non-reward events,” *Neuroscience, vol. 96, no. 4,pp. 651–656, 2000.*

Oudeyer P-Y, Kaplan , F. and Hafner, V. (2007) Intrinsic Motivation Systems for Autonomous Mental Development, IEEE Transactions on Evolutionary Computation, 11(2), pp. 265--286.

Baranes, A., Oudeyer, P-Y. (2009) R-IAC: Robust intrinsically motivated exploration and active learning, IEEE Transactions on Autonomous Mental Development, 1(3), pp. 155--169.

Baranes, A., Oudeyer, P-Y. (2010) Intrinsically Motivated Goal Exploration for Active Motor Learning in Robots: a Case Study, inProceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010), Taipei, Taiwan.

J. Schmidhuber, “Curious model-building control systems,” in *Proc. Int. Joint Conf. Neural Netw., Singapore, 1991, vol. 2, pp. 1458–1463.*

J. Schmidhuber. Developmental Robotics, Optimal Artificial Curiosity, Creativity, Music, and the Fine Arts. *Connection Science, 18(2): 173-187, June 2006.*

S. Schaal and C. G. Atkeson, “Robot juggling: an implementation of memory-based learning,” Control systems magazine, pp. 57–71, 1994.

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

Baranes, A., Oudeyer, P-Y. (2010) Maturationally-Constrained Competence-Based Intrinsically Motivated Learning, in Proceedings of IEEE International Conference on Development and Learning (ICDL 2010), Ann Arbor, Michigan, USA.

Ly, O., Oudeyer, P-Y. (2010) Acroban the humanoid: Playful and compliant physical child robot interaction, SIGGRAPH’2010 *Emergent Technologies.*