The challenges of active exploration and learning in high-dimensional continuous spaces

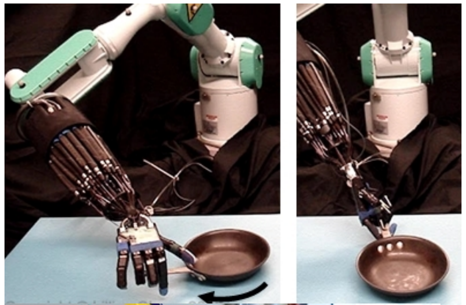
Pierre-Yves Oudeyer
Project-Team INRIA-ENSTA-ParisTech FLOWERS

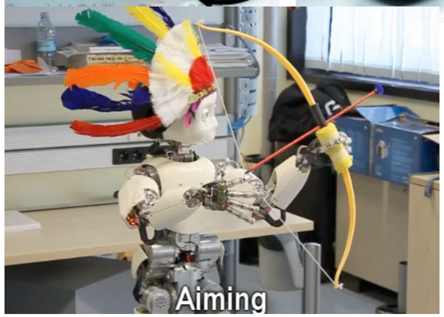
http://www.pyoudeyer.com
http://flowers.inria.fr

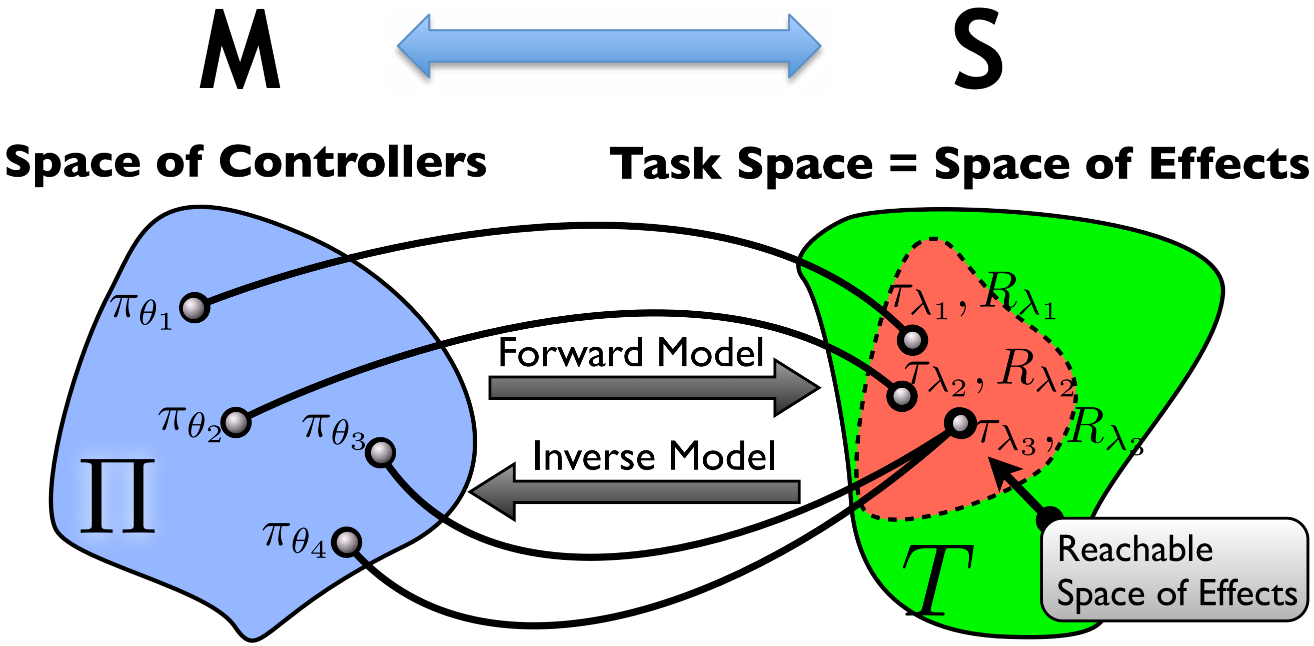
informatics mathematics

European Research Council

Learning (generalized) sensorimotor mappings







Parameterized by

$$\theta_i \in \mathbb{R}^n$$

Parameterized by

$$\lambda_j \in \mathbb{R}^m$$

Probabilistic models P(S|M)
Joint mapping P(M,S): notion of forward and inverse model depend on use

Why difficult to learn?

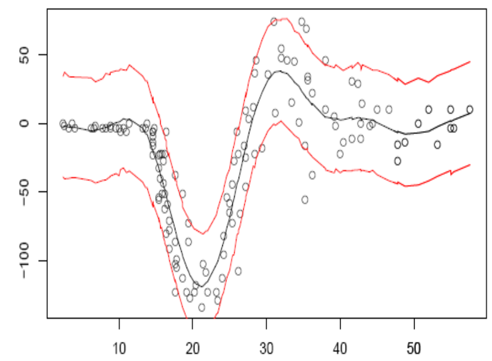
- High-dimensional and continuous
- Redundant
- Stochastic, inhomogeneous in terms of learnability
- Physical experiments time + limited life time = limited number of training data

➔ Guide actively collection of data/experiments to maximize what can be learnt within a life-time;

➔ Whole joint P(M,S) mapping, even P(S|M), even M→S, cannot be learnt (data too sparse)

Apprentissage actif de modèles

$S(t+\Delta)$



$(S(t),\pi_\theta)$

$(S(0),\pi_{\theta,1},proj(S(1)))$
$\Rightarrow pred_1$

$(S(1),\pi_{\theta,2},proj(S(2)))$
$\Rightarrow pred_2$

$\vdots$

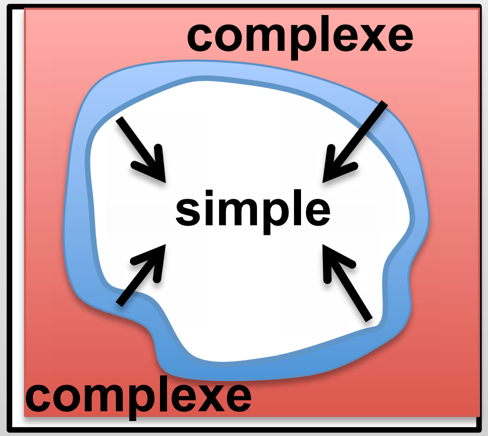
$(S(n-1),\pi_{\theta,n},proj(S(n)))$
$\Rightarrow pred_n$

➔ Quel $(S(n+1),\pi_{\theta,n+1})$ expérimenter ?

$(S(t),\pi_\theta)$

Explorer zones:
- Incertitude/erreurs maximales
- Les moins explorées

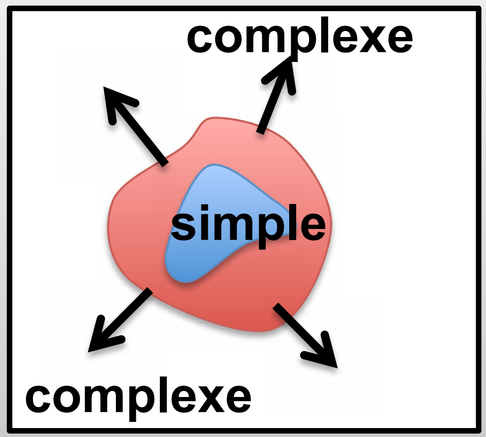
Suppose:
- Stationarité spatiale et temporelle
- Tout est apprenable
- Modèle de la fonction d'erreur



$(S(t),\pi_\theta)$

Approche développementale

Explorer zones:
Progrès en apprentissage empiriquement mesuré est maximal



Spontaneous active exploration, artificial curiosity


Intrinsic Motivation

Berlyne (1960), Csikszentmihalyi (1996)
Dayan and Belleine (2002)

$$\underset{(SVR,GPR,NN,...)}{predict} : (S(t), \pi_\theta) \rightarrow \tilde{S}(t+\Delta)$$

$$\varepsilon(S(t), \pi_\theta) = \left|\tilde{S}(t+\Delta) - S(t+\Delta)\right|$$

$$R(S(t), \pi_\theta) = \varepsilon \quad ?$$ Non !

Quelle fonction de récompense générique ?

$$R : \pi_\theta \rightarrow r \in \Re$$

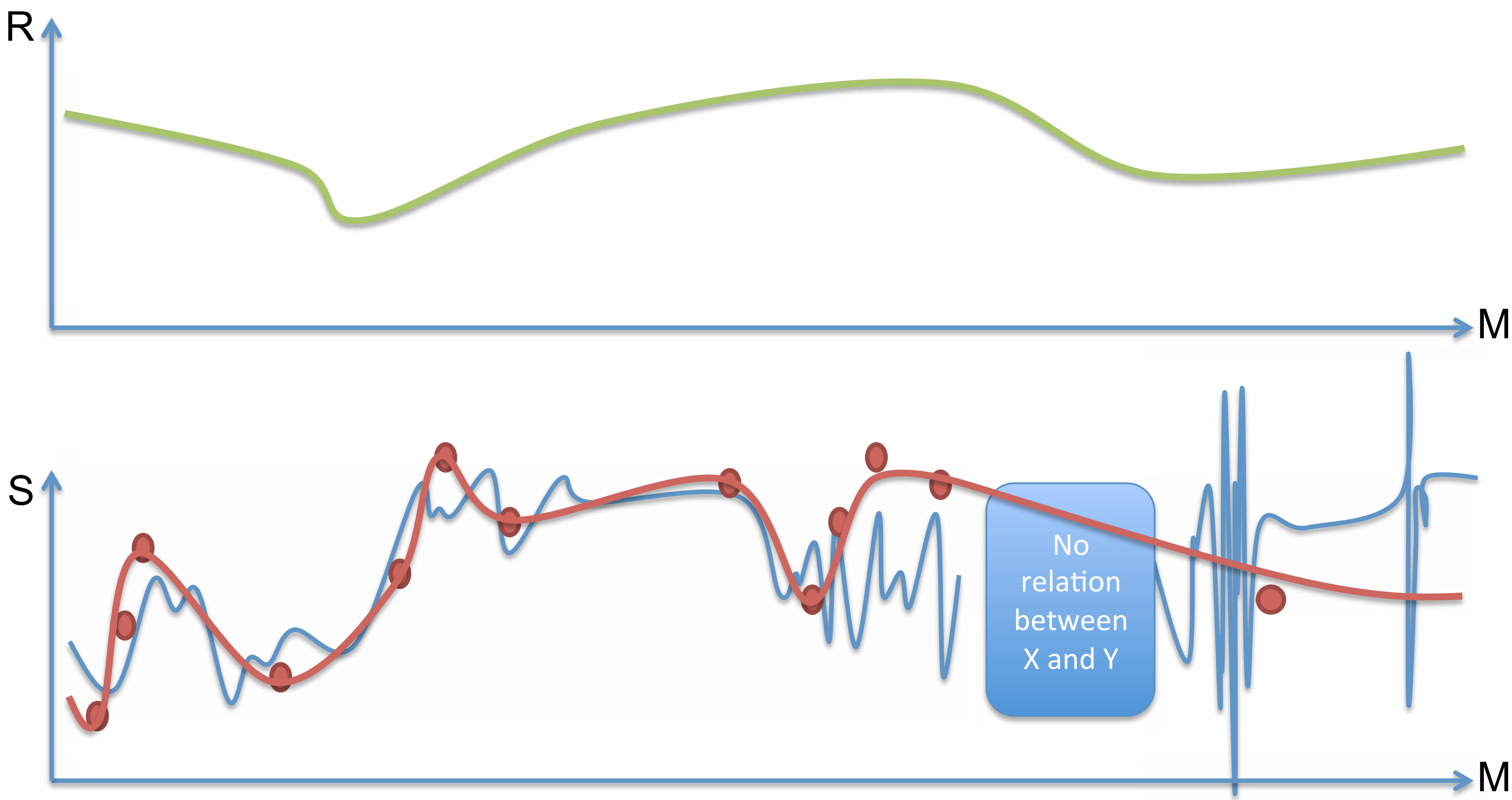
$R(S(t), \pi_\theta) = -\frac{d\varepsilon}{dt}$ in the vicinity of $(S(t), \pi_\theta)$

➔ Non-stationary function, difficult to model

➔ Algorithms for empirical evaluation of de/dt with statistical regression

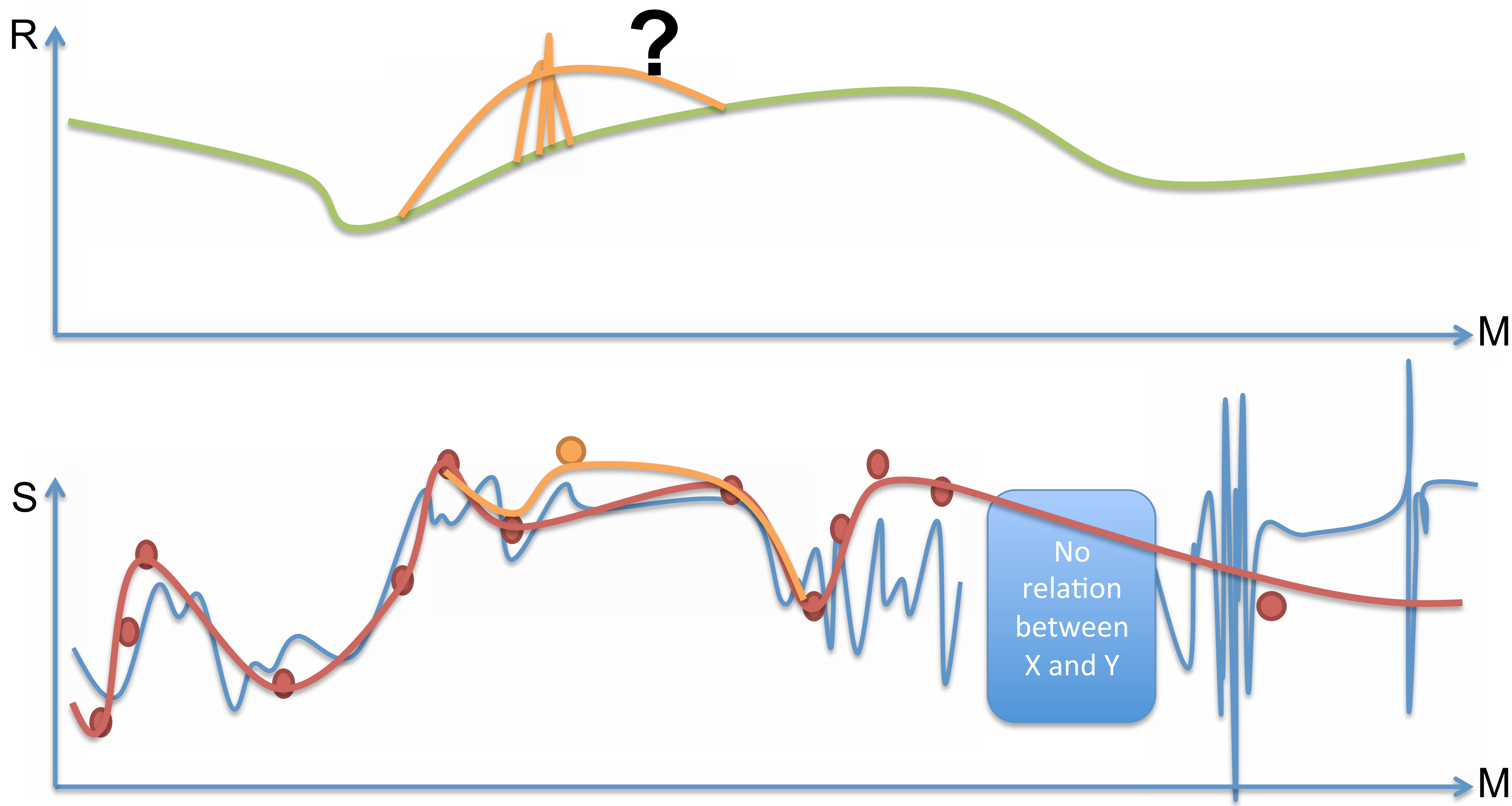
➔ IAC (2004, 2007), R-IAC (2009), SAGG-RIAC (2010) McSAGG-RIAC (2011), SGIM (2011)

Estimating Learning Progress: a Regression Problem



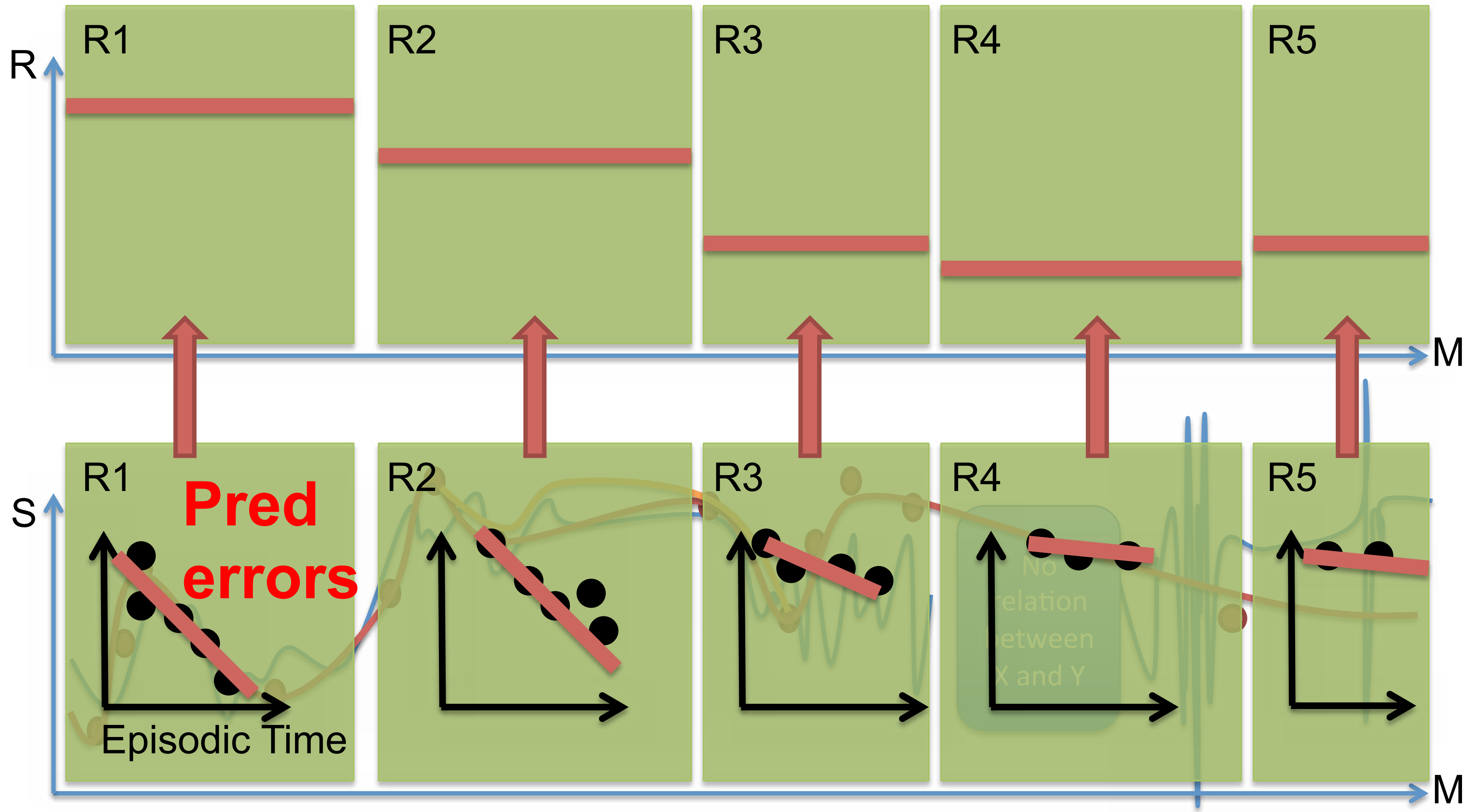
Challenge 1) Estimate R with few samples
Challenge 2) Meta-exploration/meta-exploitation problem

How to estimate it usefully for future experiments?



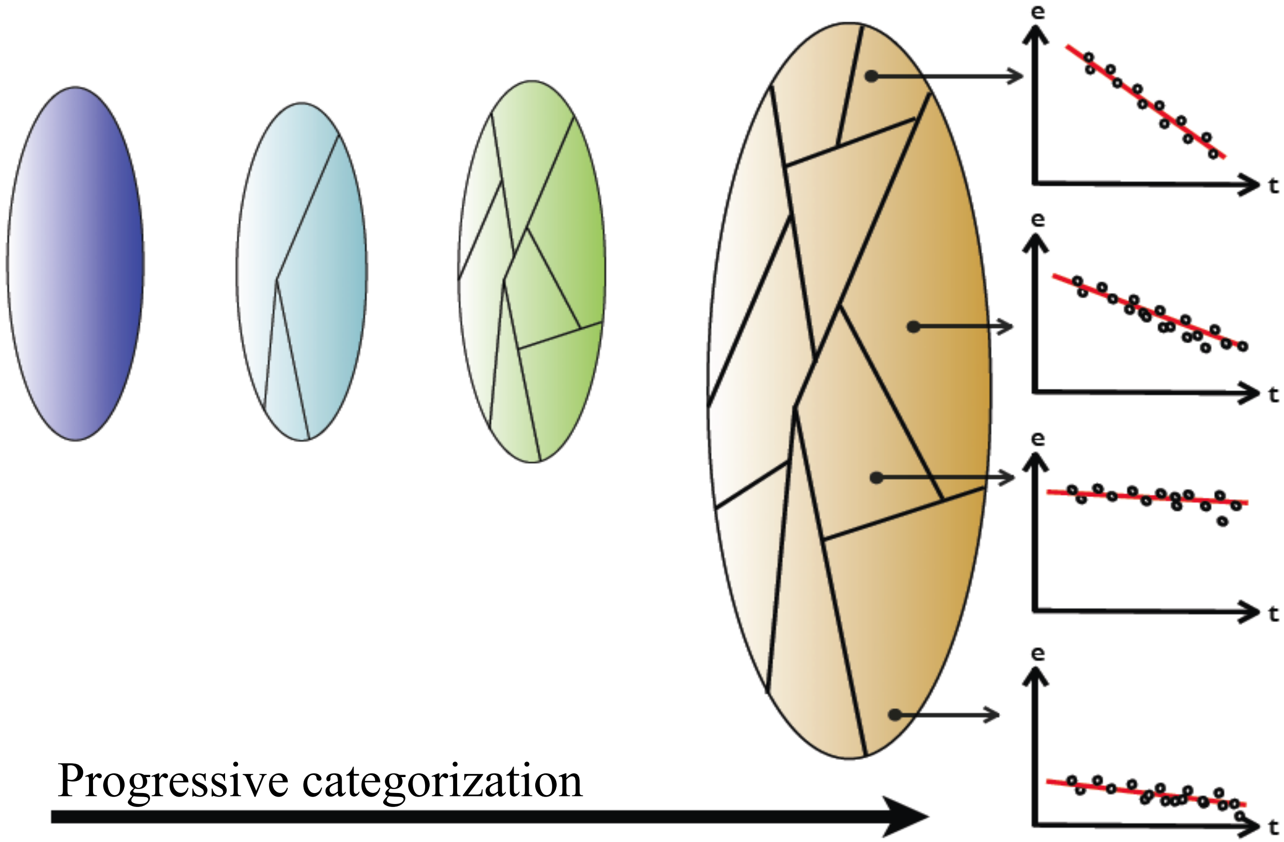
Update locally, but not too much for generalization

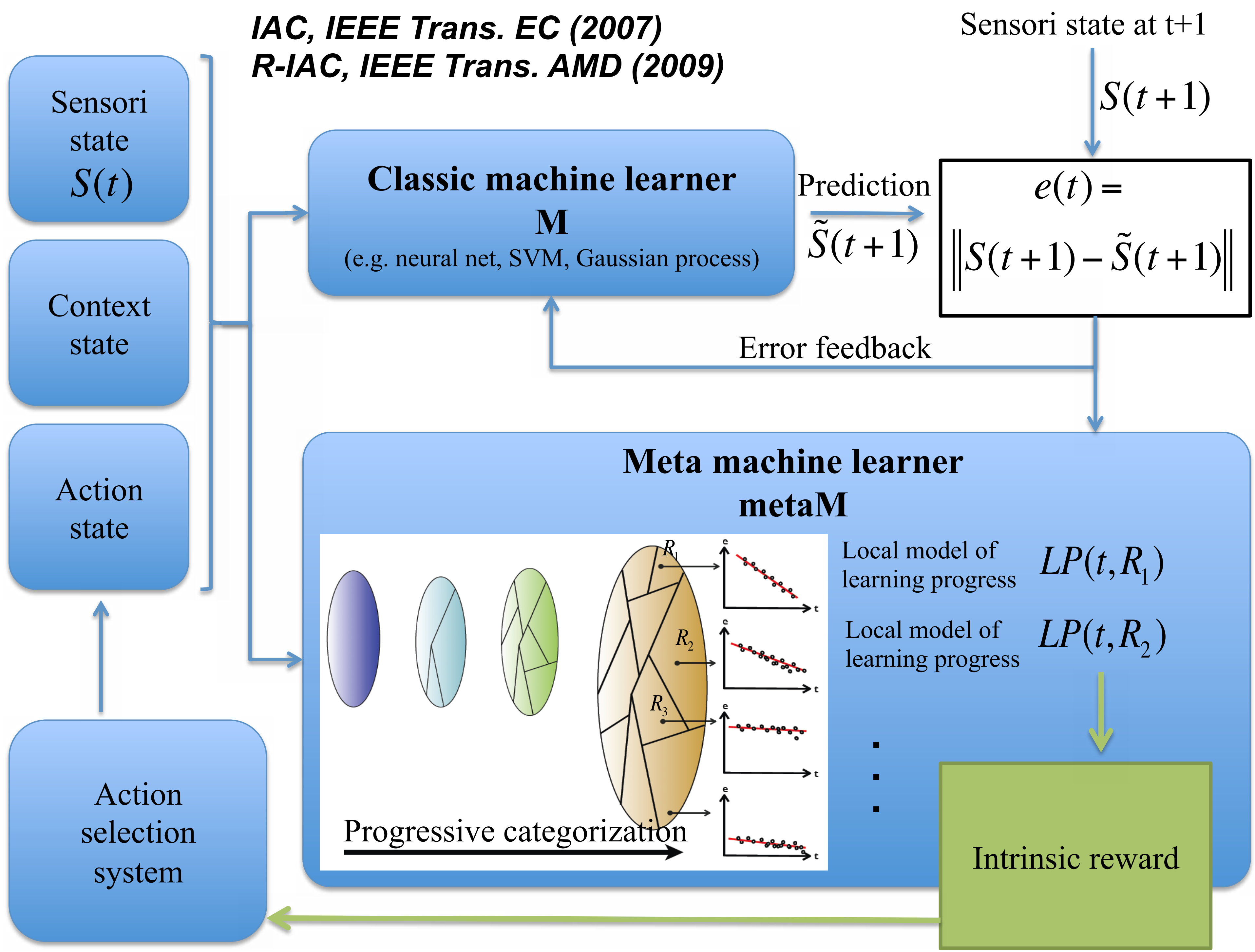
Region-based evaluation of LP



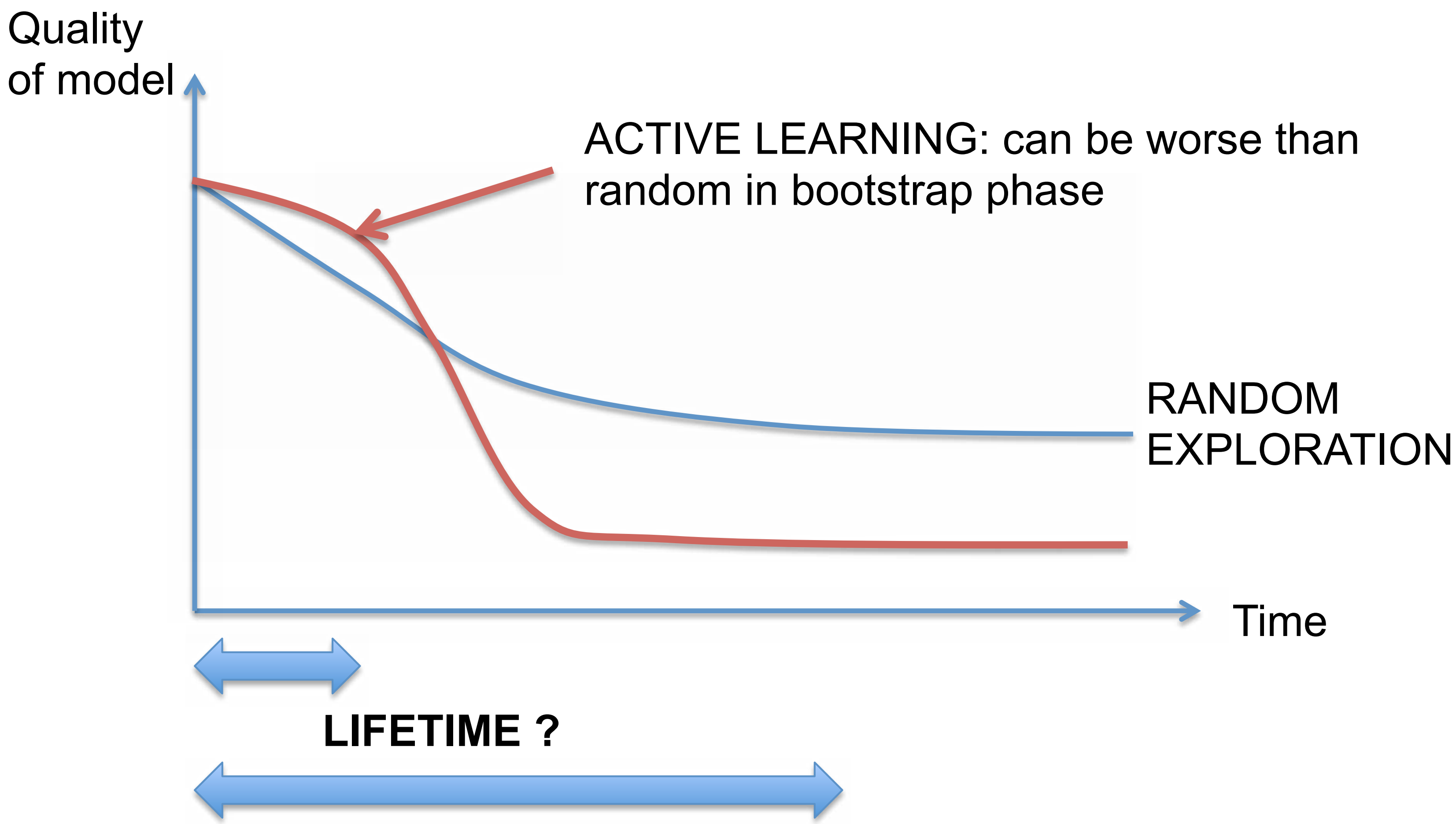
Coarse representation of R: Fast to estimate (sample and comp. complexity)

From coarse to fine regions: Sub-divide where it is most interesting/explored

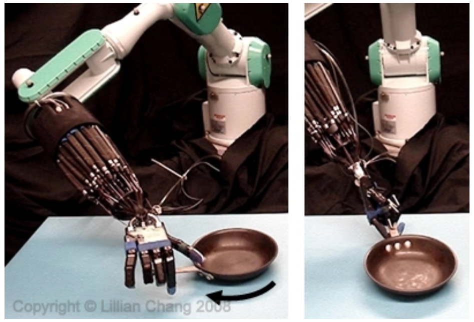



IAC, IEEE Trans. EC (2007)
R-IAC, IEEE Trans. AMD (2009)
Sensori state at t+1
$S(t+1)$
Sensori state $S(t)$
Context state
Action state
Classic machine learner M
(e.g. neural net, SVM, Gaussian process)
Prediction
$\tilde{S}(t+1)$
$e(t) = \left\| S(t+1) - \tilde{S}(t+1) \right\|$
Error feedback
Meta machine learner metaM
R_1
R_2
R_3
Progressive categorization
Local model of learning progress
$LP(t, R_1)$
Local model of learning progress
$LP(t, R_2)$
Intrinsic reward
Action selection system

Bootstrapping a useful model of R (even step-wise) can take time in high-dimensions (or just large domains)

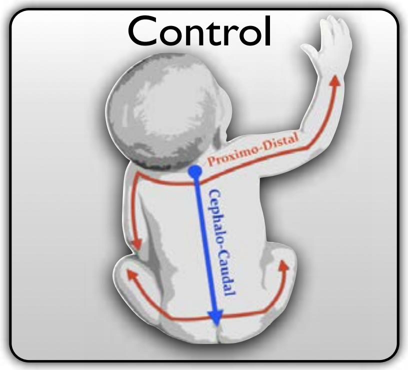


Strategies for scalable active learning in very large spaces



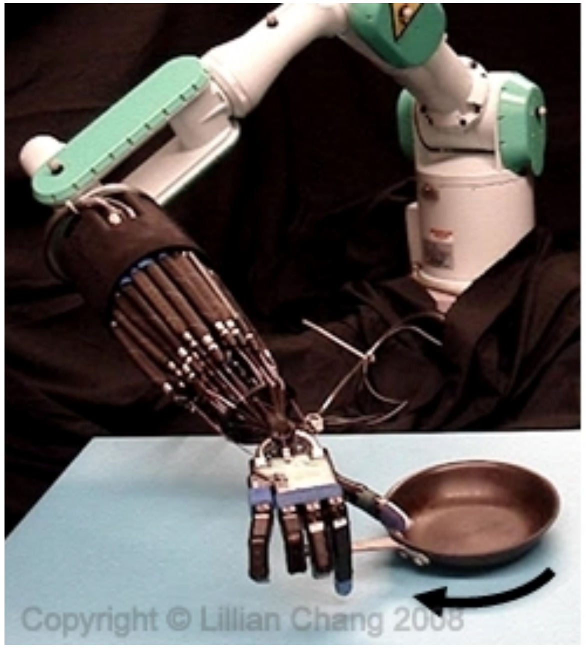
Task space exploration

Social guidance, scaffolding



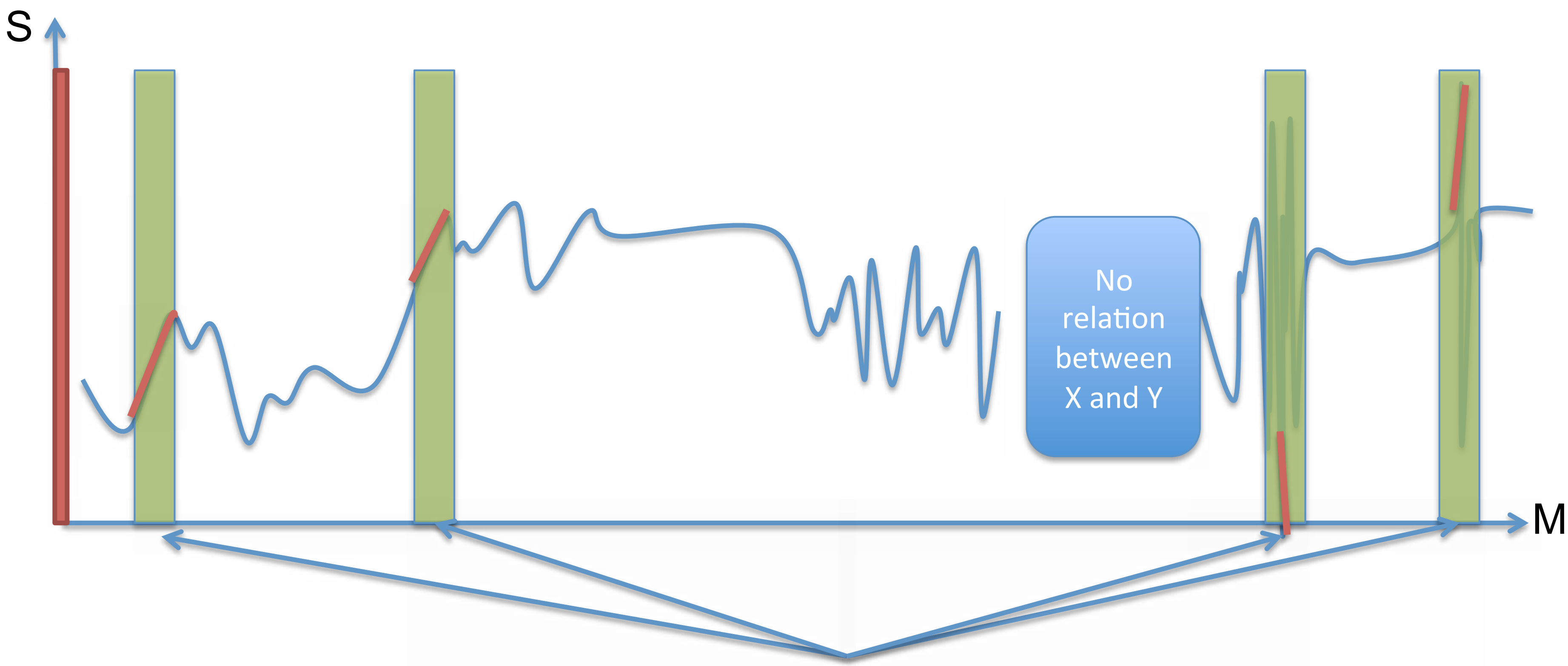
(Adaptive) Maturation and embodiment

Task space active exploration



What is often most useful is knowing P(M|S)

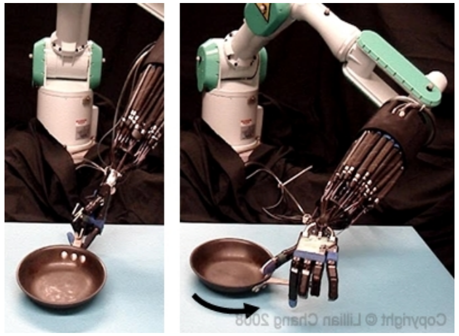
➔ Leveraging of redundancy by learnng only what is enough



Knowing these parts of the forward model is sufficient to know how to produce all possible effects in the Task space

Active learning of inverse models SAGG-RIAC *(RAS, 2012)*

Redundancy of sensorimotor spaces



(Context, Movement) → Effect

From the active choice of action, followed by observation of effect …

$$predict : (S(t), \pi_\theta) \rightarrow proj(\tilde{S}(t+\Delta))$$

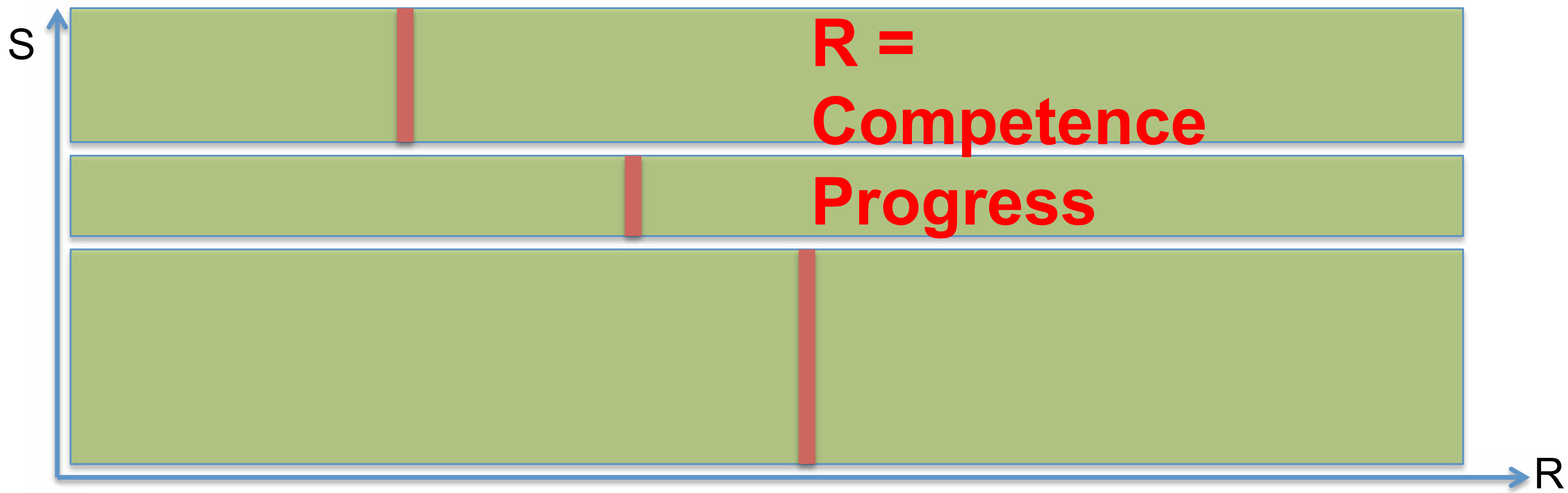
… to the active choice of effect, followed by the search of a corresponding action policy through goal-directed optimization (e.g. using NAC, POWER, PI^2-CMA, …)

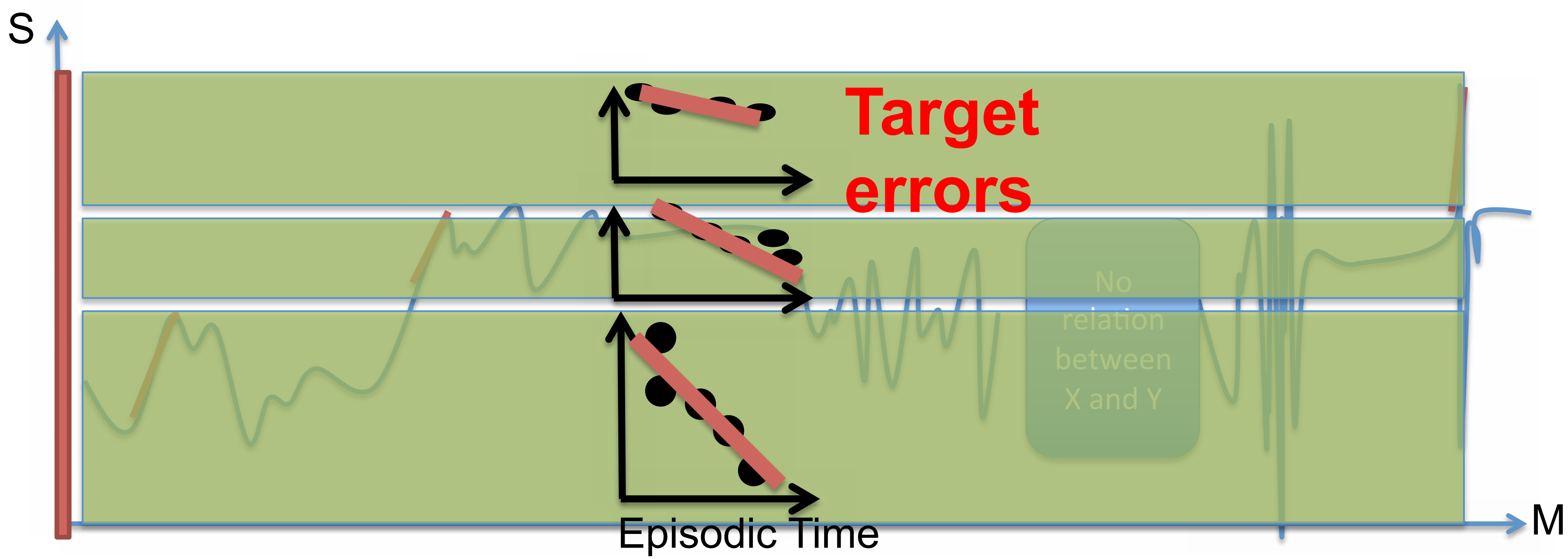
➔ self-defined RL problem $R_\lambda : \pi_\theta \rightarrow \Re$

$$control : (S(t), R_\lambda) \overset{optimisation}{\triangleright} \pi_{\tilde{\theta}}$$

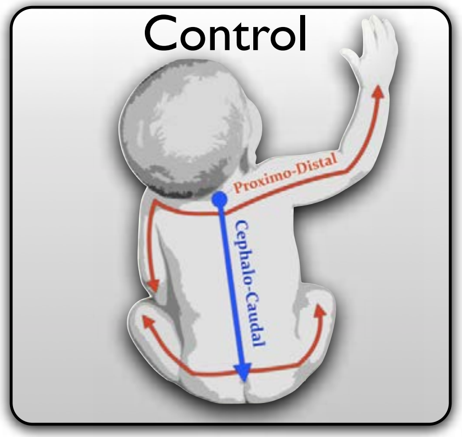
Spontaneous active exploration of a space of fitness functions parameterized by λ where one iteratively chooses the R_λ which maximizes the empirical evaluation of:

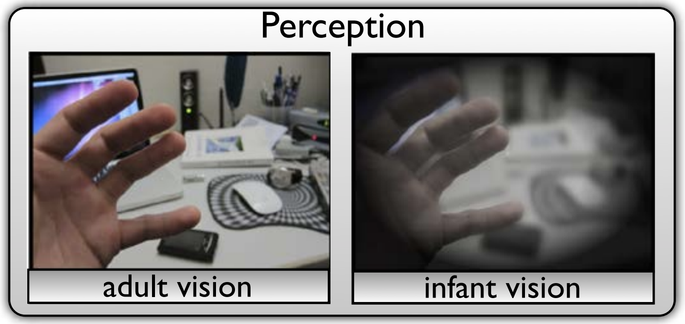
$$competence\ progress : R_\lambda(\pi_{\tilde{\theta},new}) - R_\lambda(\pi_{\tilde{\theta},init})$$


S
R =
Competence
Progress
R


S
Target
errors
No
relation
between
X and Y
Episodic Time
M

Maturational constraints



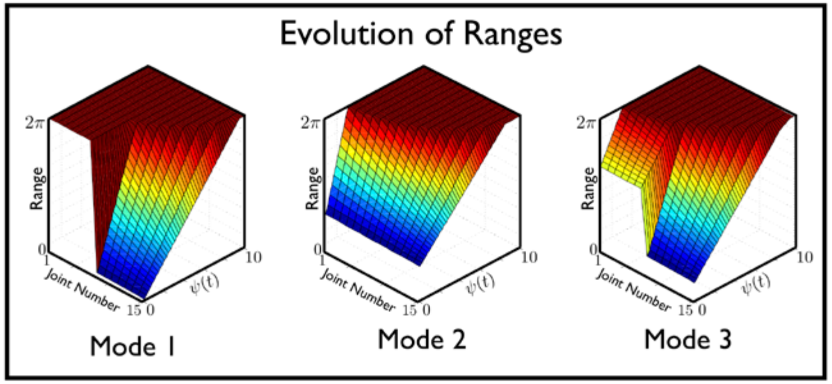


(Bjorklund, 1997; Turkewitz and Kenny, 1985)

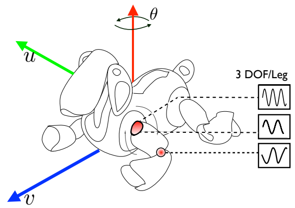
- Progressive growths of DOF number and spatio-temporal resolution

$$S, \pi_\theta$$

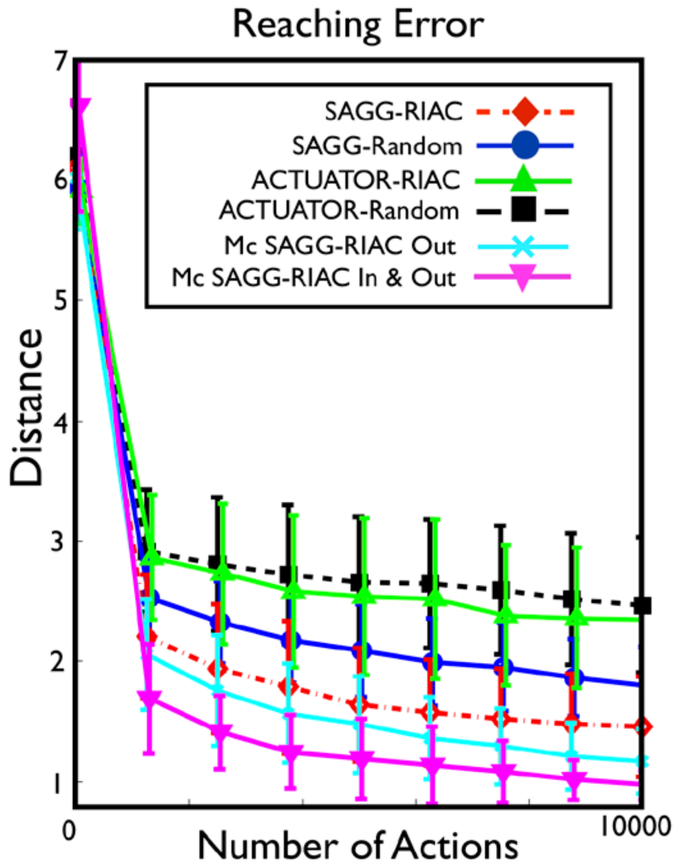
- Adaptive maturational schedule controlled by active learning/learning progress



Experimental evaluation

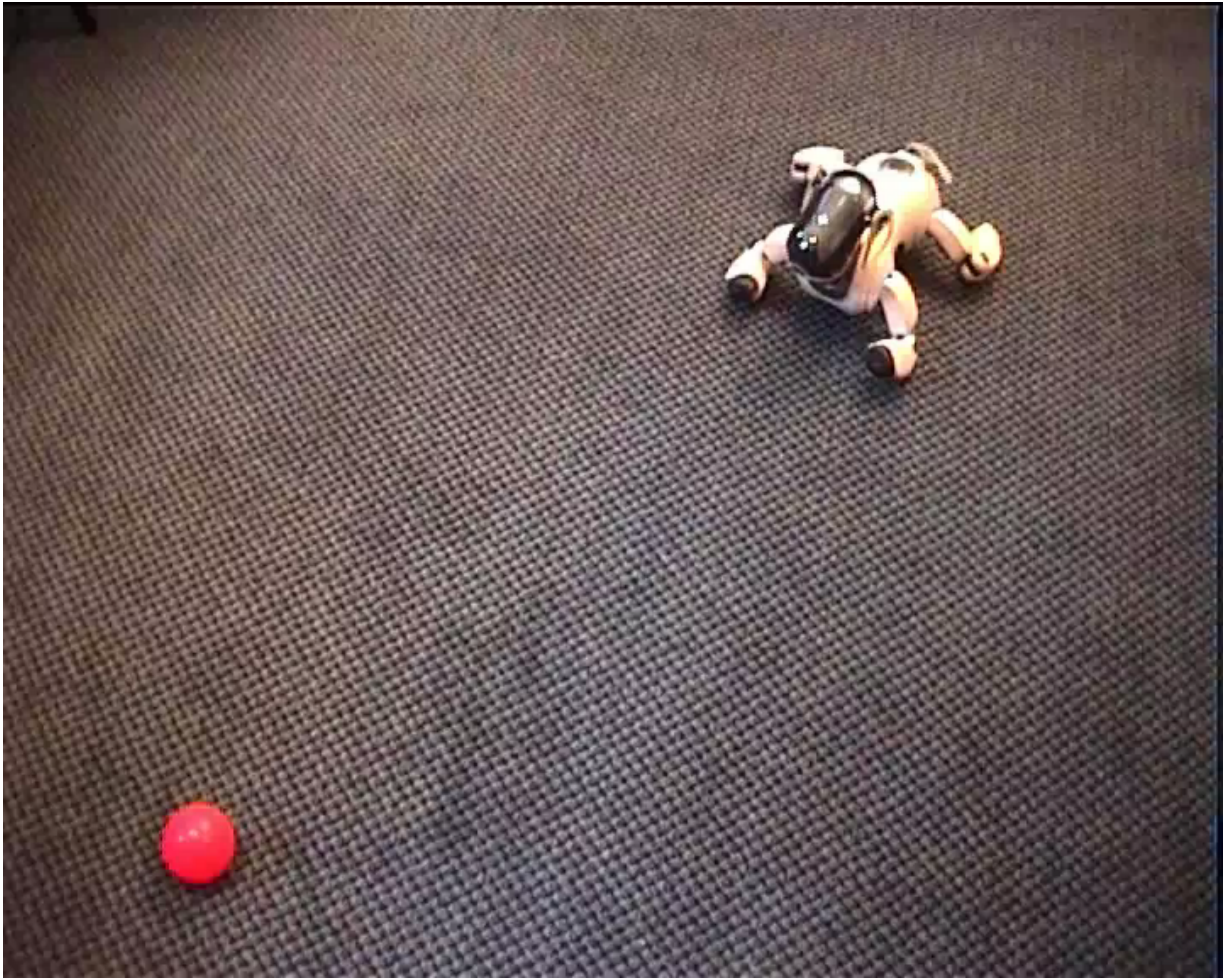


Learning omnidirectional locomotion

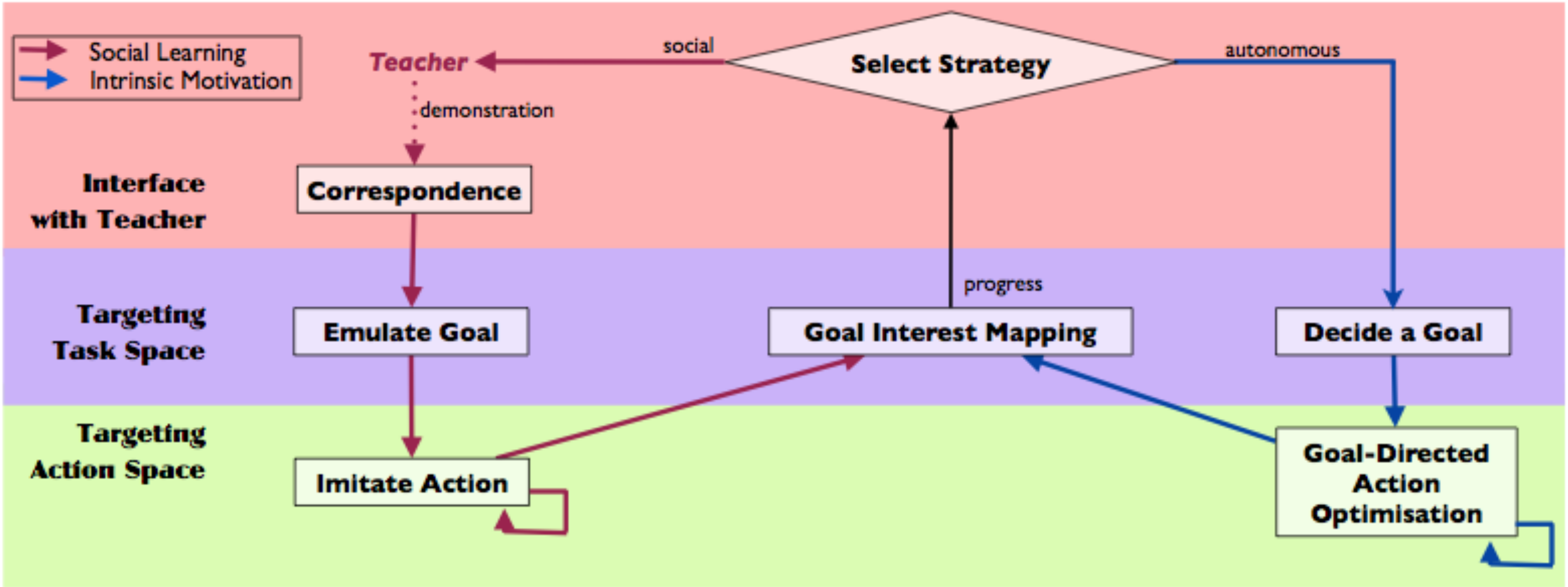


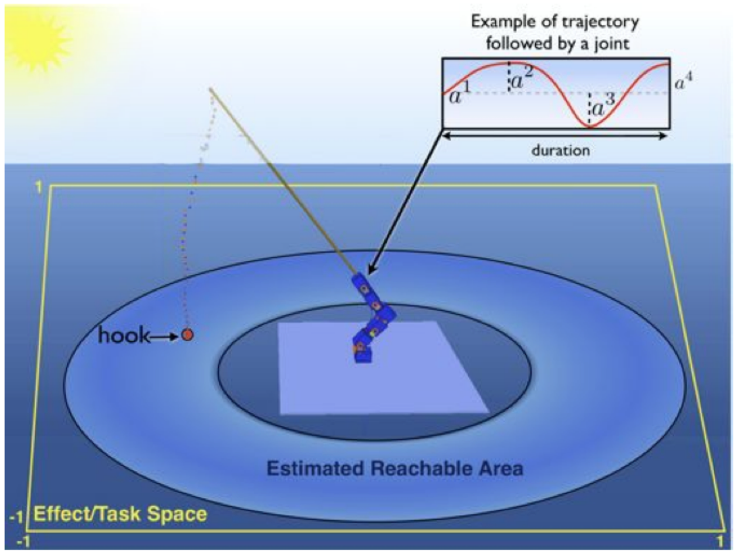
Control Space: $[-1;1]^{24}$ Task Space: $[-1;1]^{3}$

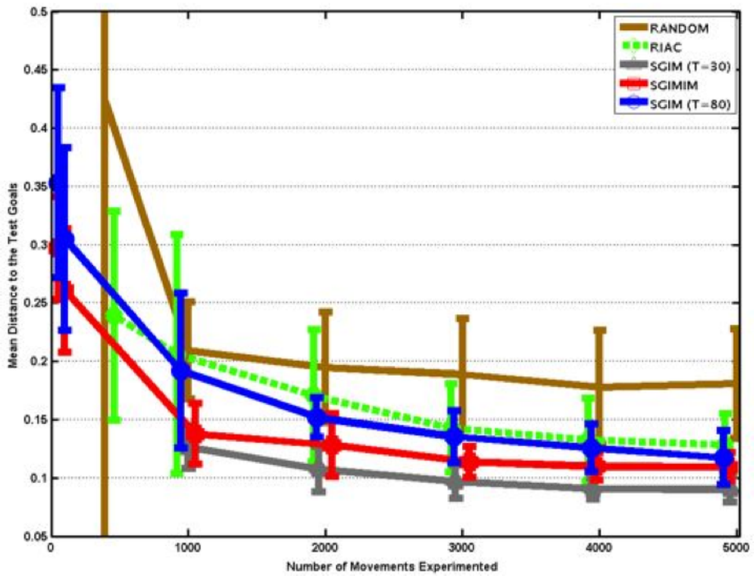
➔ Performance higher than more classical active learning algorithms in real sensorimotor spaces (non-stationary, non homogeneous)
(IEEE TAMD 2009; ICDL 2010, 2011; IROS 2010; RAS 2012)



SGIM: Socially Guided Intrinsic Motivation







ICDL-Epirob, 2011

Baranes, A., Oudeyer, P-Y. (2012) Active Learning of Inverse Models with Intrinsically Motivated Goal Exploration in Robots, Robotics and Autonomous Systems.
http://www.pyoudeyer.com/RAS-SAGG-RIAC-2012.pdf

Baranes, A., Oudeyer, P-Y. (2011) The Interaction of Maturational Constraints and Intrinsic Motivation in Active Motor Development, in Proceedings of IEEE ICDL-Epirob 2011.
http://flowers.inria.fr/BaranesOudeyerICDL11.pdf

Lopes, M., Melo, F., Montesano, L. (2009) Active Learning for Reward Estimation in Inverse Reinforcement Learning, *European Conference on Machine Learning (ECML/PKDD),* Bled, Slovenia, 2009.
http://flowers.inria.fr/mlopes/myrefs/09-ecml-airl.pdf

Nguyen, M., Baranes, A., Oudeyer, P-Y. (2011) Bootstrapping Intrinsically Motivated Learning with Human Demonstrations, in Proceedings of IEEE ICDL-Epirob 2011.
http://flowers.inria.fr/NguyenBaranesOudeyerICDL11.pdf

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.
http://www.pyoudeyer.com/ims.pdf

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.

Exploration in Model-based Reinforcement Learning by Empirically Estimating Learning Progress, Manuel Lopes, Tobias Lang, Marc Toussaint and Pierre-Yves Oudeyer. *Neural Information Processing Systems (NIPS 2012)*, Tahoe, USA. (pdf)

The Strategic Student Approach for Life-Long Exploration and Learning, Manuel Lopes and Pierre-Yves Oudeyer. *under review*, . (pdf)