

Week 1: Novelty, Curiosity, Surprise

10.02.2016

What, when how to learn

Question: What is the best strategy to learn?

- ▶ Optimize learning to maximize reward? Optimize task-performance?
Learn from mistakes?



Intrinsic Motivation (IM)

Homeostasis is the property of a system in which variables are regulated so that internal conditions remain stable and relatively constant.

IM behaviors have no clear goal, purpose or biological function; performed for their own sake.

- ▶ Serving the need of hunger (homeostatic need) is considered extrinsic.

Advantages:

- ▶ Task-independent, so that in a completely new environment, with no prior-knowledge, it can learn through self-directed exploration.
- ▶ Re-use of skills?
- ▶ IM is open-ended

Questions in IM:

Given possibilities, how to select what to explore next? Novel situations that have not been experienced before. Situations that are easy to learn? Situations that add a lot of information to the world model?

Intrinsic Motivation (IM)

Knowledge based view (focus: properties of the environment)

- ▶ detect novel or unexpected features, objects or events in the environment
- ▶ depends on current state of knowledge
- ▶ motivated to expand the knowledge base
- ▶ *Novelty-based IM:*
 - ▶ *Mismatch between experience and knowledge base*
 - ▶ *Low/moderate/high level of novelty, which one to go?*
- ▶ *Prediction-based IM*
 - ▶ *Explore unfamiliar “edges” of its knowledge base. Curiosity*
 - ▶ *Probe the environment, if does not behave as expected, explore more!*

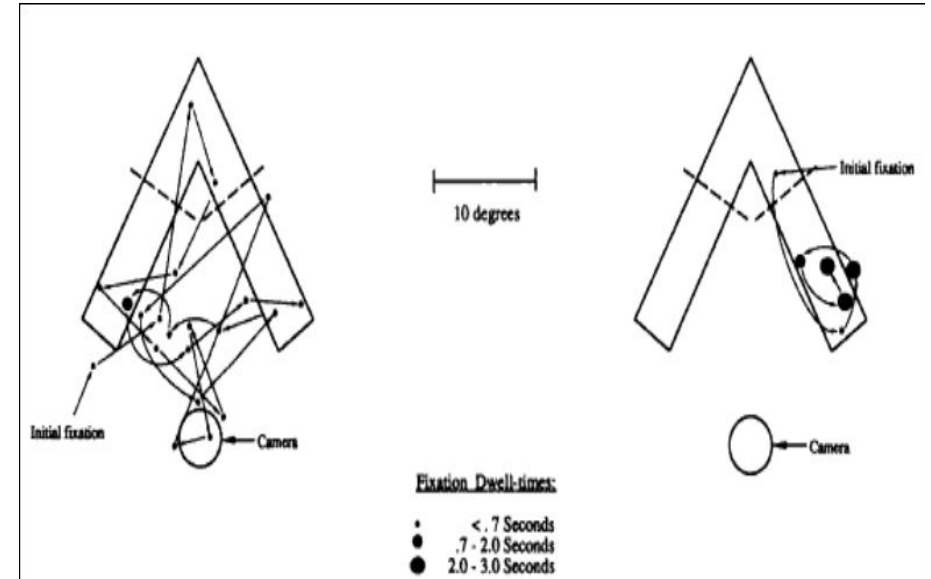
Competence based view (focus: self, abilities, skills it possess)

- ▶ Skill development in challenging environments
- ▶ Piaget's functional assimilation: systematically practice or repeat the newly emerging skills.

Intrinsic Motivation in Infants

Novelty in Infants, 1 - visual exploration

- ▶ Bronson, 1991
- ▶ 2 and 12 month olds
- ▶ Fixations are not distributed and long (2 sec)
- ▶ Interpretation? Only attention to novel areas?
 - ▶ Related to both novelty and control of the eyes.



Intrinsic Motivation in Infants

Novelty in Infants, 2 - novelty detection



- ▶ Habituation: the decrease in response to a stimulus as a result of repeated presentations of that stimulus.
- ▶ The longer infants are exposed to a stimulus, the less time they will spend looking at it. Habituation occurs when there is a substantial decrease in looking time following repeated presentation.
- ▶ How can you detect novelty with this?
- ▶ Habituation – dishabituation
 - ▶ Present male faces repeatedly, infants habituate.
 - ▶ Present female faces in the post-habituation phase.
 - ▶ Significant increase in looking time: something novel!

Intrinsic Motivation in Infants

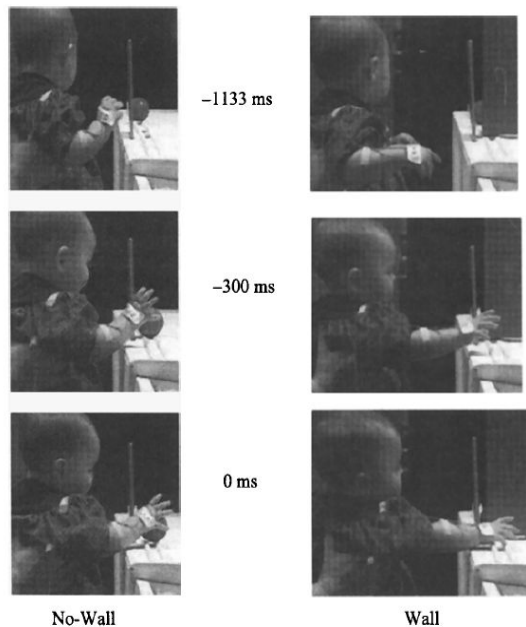
Novelty in Infants, 2 - novelty detection

- ▶ Change in novelty preference in infants:
 - ▶ 0-2 months: prefer familiar objects
 - ▶ 3-6 months: shift toward novel objects
 - ▶ 6-12 months: robust performance for novel objects
- ▶ How to explain familiarity → novelty preference?
- ▶ Evgeny Sokolov, the stimulus-model comparator theory
 - ▶ Gradually create an internal model of the observed stimuli
 - ▶ Compare the stimulus with the model,
 - ▶ If match, then inhibit the response.
- ▶ Limited visual processing does not allow stable model construction

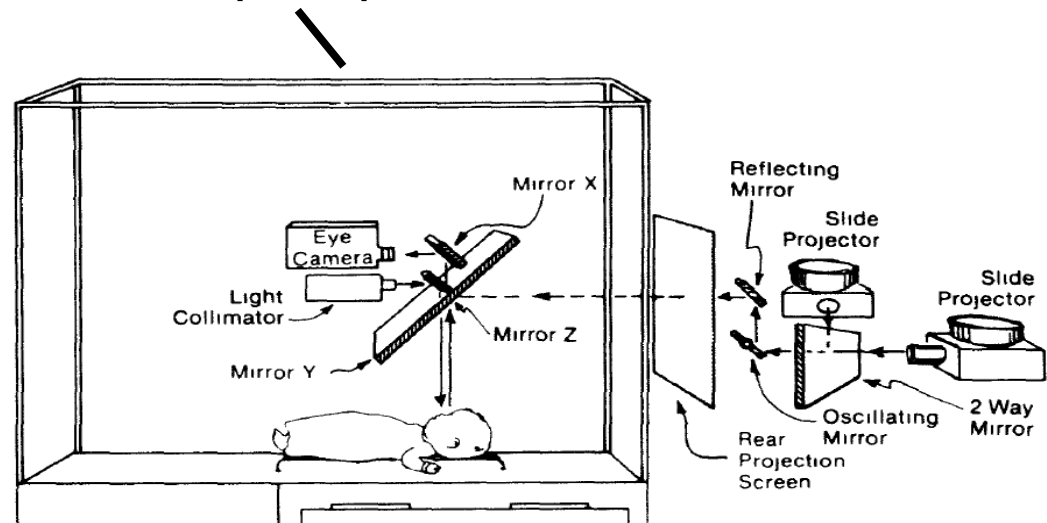
Intrinsic Motivation in Infants

Knowledge-based in Infants: Prediction

- ▶ In anticipation of an event → action based on outcome (e.g. gaze shift)
- ▶ Studied with visual expectation paradigm (VexP)
- ▶ 2-month-olds learn to predict the upcoming locations of the images in regular patterns, cannot deal with complex patterns. But 3-month-olds can



Berthier, Neil E., et al. "Using object knowledge in visual tracking and reaching." *Infancy* 2.2 (2001): 257-284.



Haith, Marshall M., Cindy Hazan, and Gail S. Goodman. "Expectation and anticipation of dynamic visual events by 3.5-month-old babies." *Child development* (1988): 467-479.

Intrinsic Motivation in Infants

Competence-based IM in Infants

- ▶ Reminder: focus is on developing organism rather than acquiring info about environment.
- ▶ Early emerging component: self-efficacy, recognition that one's own behavior has an effect on objects and people.
- ▶ How to study?
- ▶ Contingency perception method – baby in the crib
- ▶ Video display of self&other – e.g. leg
- ▶ Spontaneous play behavior



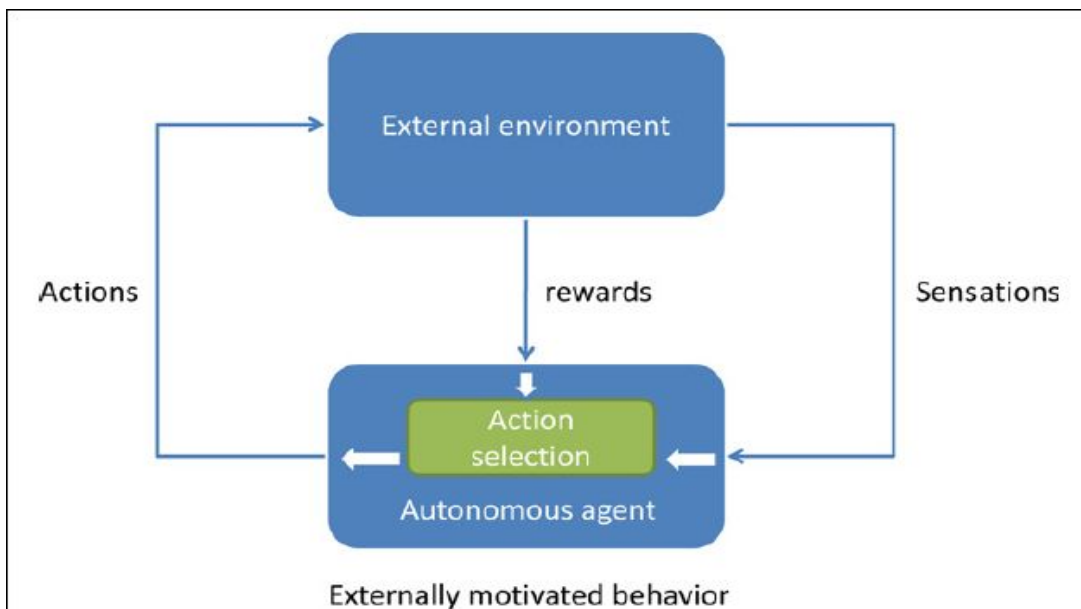
Age	Type of play	Example
0-2 years	Functional or sensorimotor	Gross activity: Running climbing, swinging kicking dropping objects
1-5 years	Constructive	Fine motor skills for building: Stacking blocks, connecting pieces of train rack
2-6 years	Pretend or symbolic	Cooking breakfast
6+ years	Rule-based	Checkers

Rewarding IM-based robots

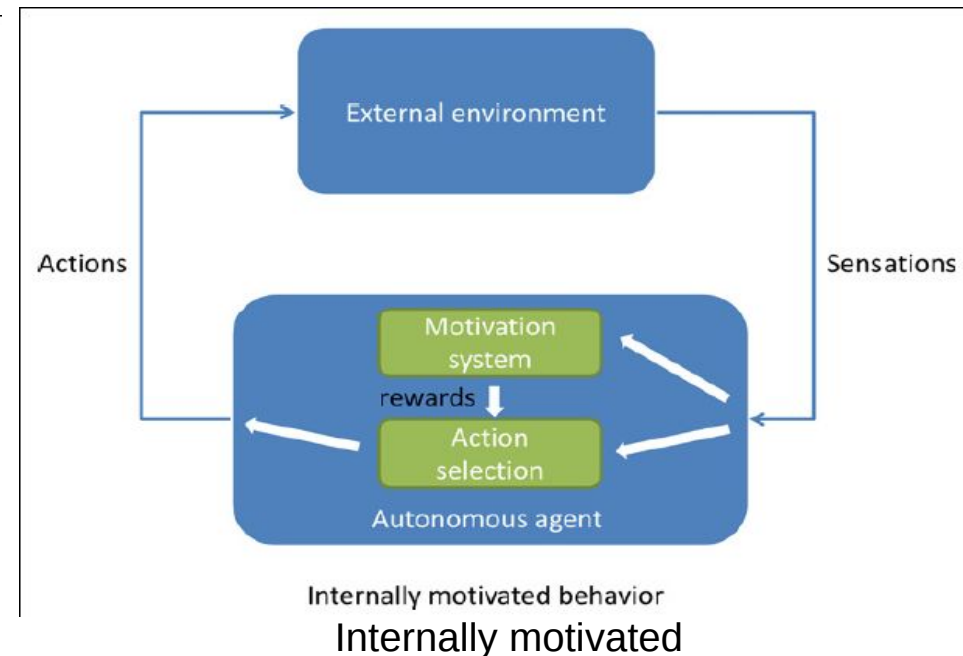


IM behaviors have no clear goal, purpose or biological function; performed for their own sake.

- ▶ Serving the need of hunger (homeostatic need) is considered extrinsic.
- ▶ Battery charge act is extrinsic motivated
- ▶ What might be an intrinsic motivated action?



Reward originates from environment: externally motivated



Internally motivated behavior

Internally motivated

Oudeyer, Pierre-Yves, and Frederic Kaplan. "What is intrinsic motivation? a typology of computational approaches." *Frontiers in neurorobotics* 1 (2007): 6.

Rewarding IM-based robots based on Novelty Amount

- ▶ Choose novel events that the robot is uncertain about

e^k : kth sample from all possible set of events

$P(e^k, t)$: estimated probability of observing event at time t

$$r(e^k, t) = C \cdot (1 - P(e^k, t))$$

- ▶ What if do not change much?
 - ▶ Somehow need to limit maximum novelty (see next slide)

- ▶ Choose events that result in decrease in total uncertainty

$H(E, t)$: total entropy

$$H(E, t) = - \sum_{e^k \in E} P(e^k, t) \ln(P(e^k, t))$$

$$r(e^k, t) = C \cdot (H(E, t) - H(E, t + 1))$$

Rewarding IM-based robots based on Prediction Error

- Reward events that are unpredictable

$SM(t)$: sensorimotor context upto time t

$\Pi(SM(t))$: prediction of next context

$$r(SM(t)) = C \cdot E_r(t)$$

where

$$\Pi(SM(t)) = \tilde{e}^k(t+1)$$

$$E_r(t) = \|\tilde{e}^k(t+1) - e^k(t+1)\|$$

- The robot can get lost! Beyond its “current” limits!

$$r(SM(t)) = C_1 \cdot e^{-C_2 \cdot \|E_r(t) - E_r^\sigma\|^2}$$

Thresholded to moderate novelty

- Reward event that provide maximum change in prediction error

- Maximum Learning Progress

$$r(SM(t)) = E_r(t) - E'_r(t)$$

where

$$E'_r(t) = \|\Pi'(SM(t)) - e^k(t+1)\|$$

Jurgen Schmidhuber:
<http://people.idsia.ch/~juergen/>

Rewarding IM-based robots based on Competence

- ▶ Focus is skills, skills have goals (g_k), goal-directed behaviors occur in episodes (t_g)

- ▶ Difference between expected goal and observed result

$$I_a(g_k, t_g) = ||\widetilde{g_k(t_g)} - g_k(t_g)||$$

- ▶ Reward?

- ▶ (mis) achievement of the intended goal. Favor large differences between attempted and achieved goals

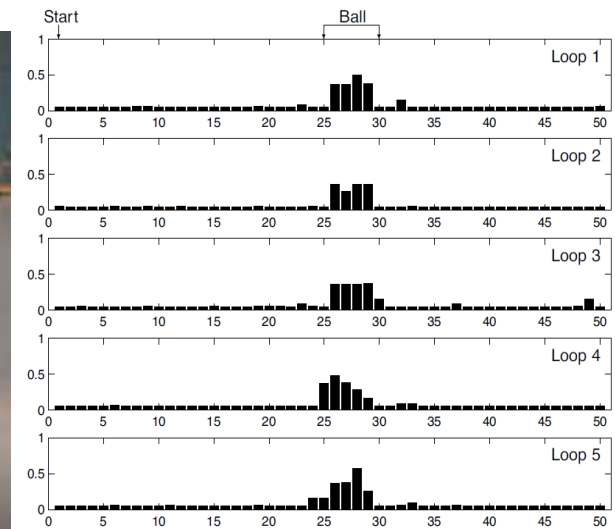
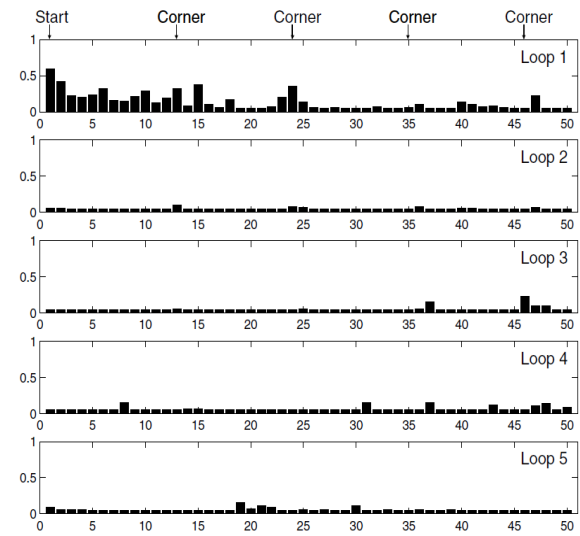
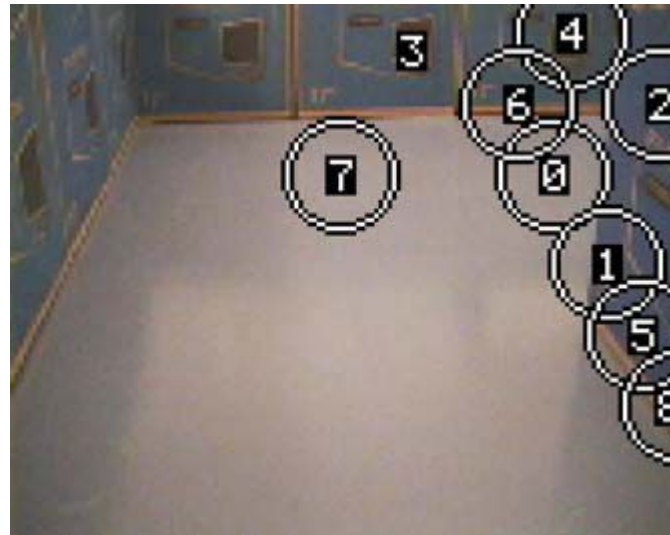
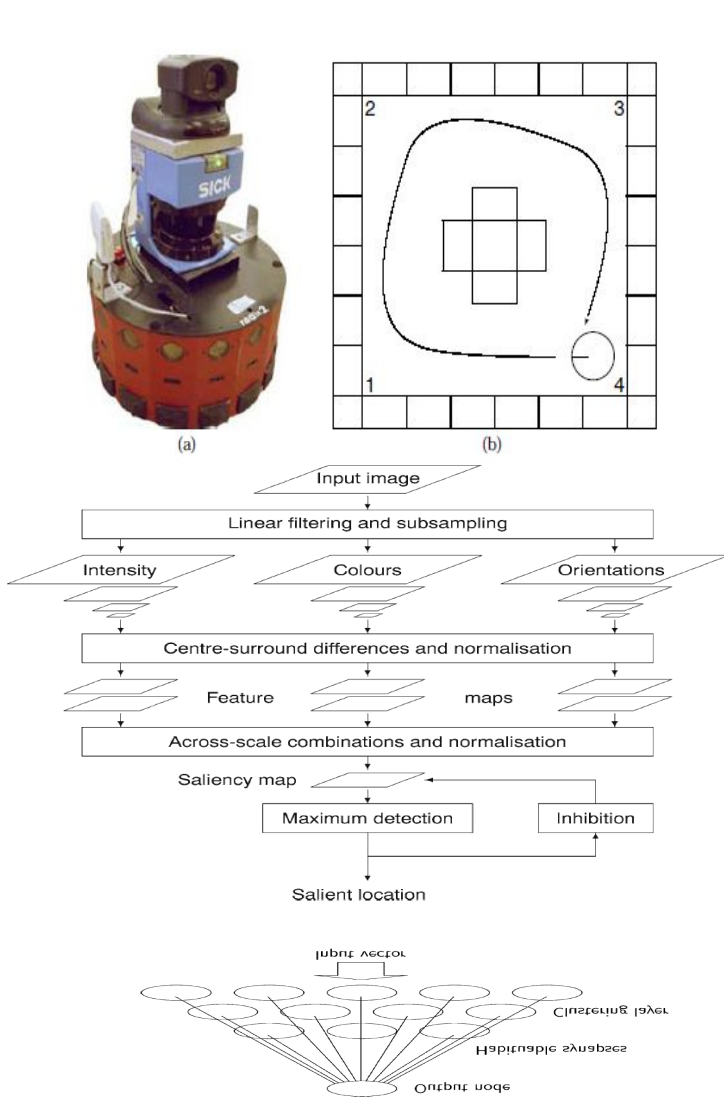
$$r(RM(t), g_k, t_g) = C \cdot I_a(g_k, t_g)$$

- ▶ Maximizing-incompetence motivation? Goals well-beyond skill level.

- ▶ Favor improvement!

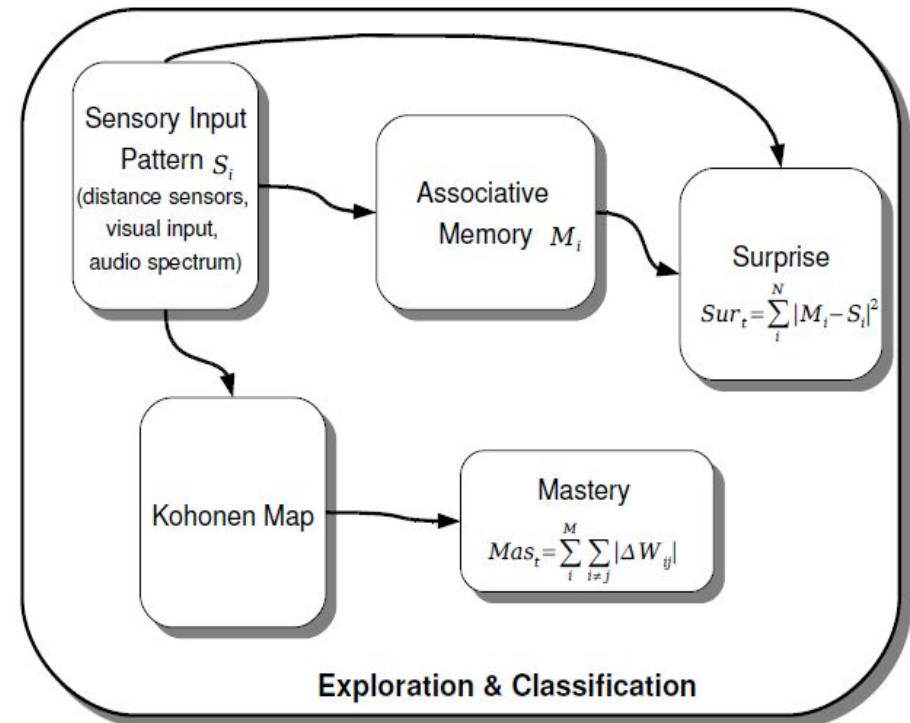
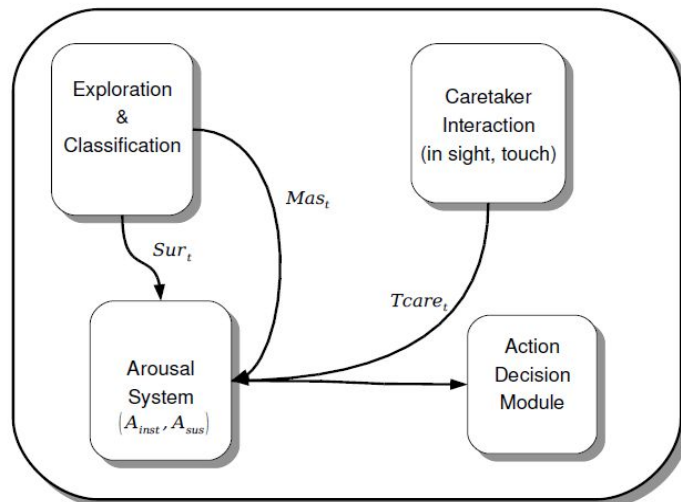
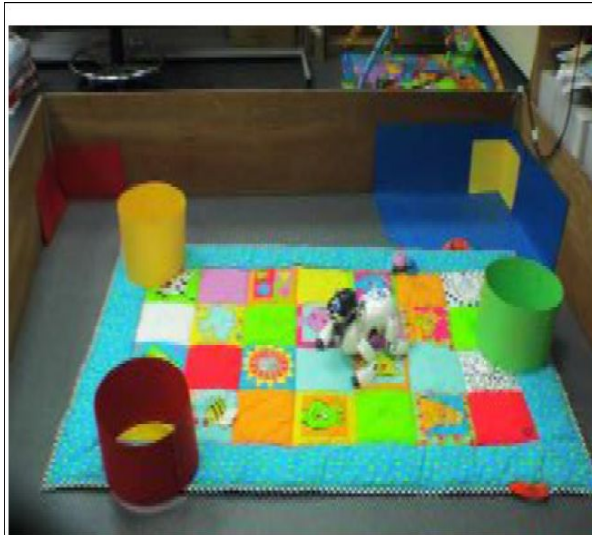
$$r(RM(t), g_k, t_g) = C \cdot (I_a(g_k, t_g - \Theta) - I_a(g_k, t_g))$$

Novelty seeking robot systems



Neto, Hugo Vieira, and Ulrich Nehmzow. "Real-time automated visual inspection using mobile robots." *Journal of Intelligent and Robotic Systems* 49.3 (2007): 293-307.

Novelty seeking robot systems



$\left\{ \begin{array}{l} \text{if } A_{inst} < 0.25 \\ \text{if } A_{inst} > 0.25 \text{ and } A_{inst} < 0.7 \\ \text{if } A_{inst} > 0.7 \\ \text{if } A_{inst} > 0.7 \text{ and } A_{sus} > 0.6 \end{array} \right.$	$\Rightarrow$ turn to explore $\Rightarrow$ stay still and learn $\Rightarrow$ bark to attract attention $\Rightarrow$ search for the caretaker
--	--

Hiolle, Antoine, and Lola Canamero. "Why Should You Care?-An Arousal-Based Model of Exploratory Behavior For Autonomous Robot." ALIFE. 2008.

Robots with prediction driven IM

$$r(SM(t)) = E_r(t) - E'_r(t)$$

where

Recall Schmidhuber's *learning process* based IM

$$E'_r(t) = \|\Pi'(SM(t) - e^k(t+1))\|$$

- ▶ Reward based on changes in the prediction error over time
- ▶ LP = difference in consecutive prediction errors
- ▶ The actions are rewarded because they led improvement in prediction.

Oudeyer et al. Intelligent Adaptive Curiosity

- ▶ Life-long learning in continuous high-dimensional sensorimotor spaces
- ▶ *Playground Experiments*

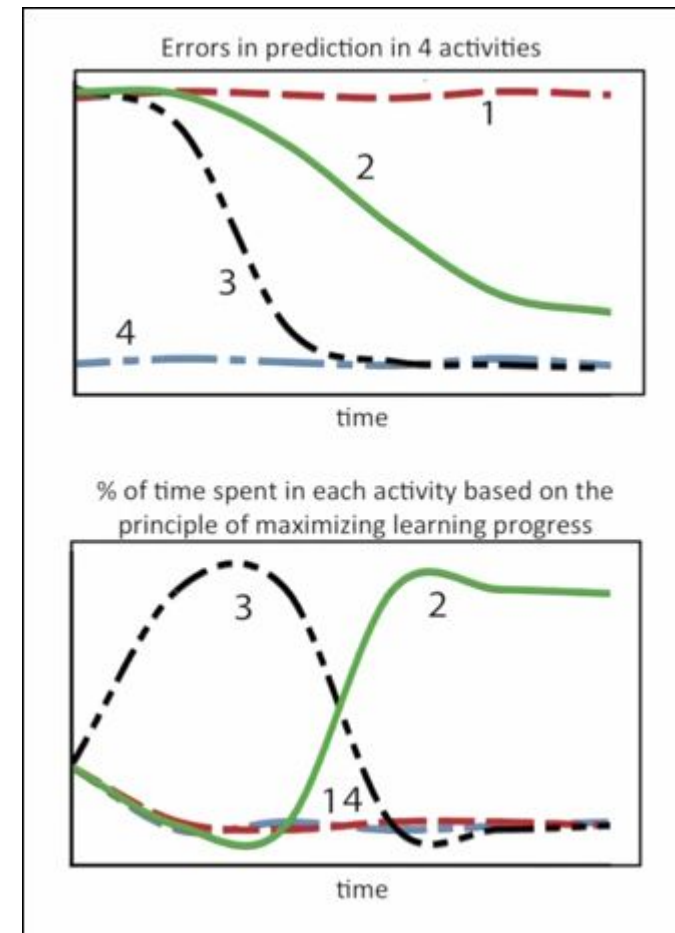
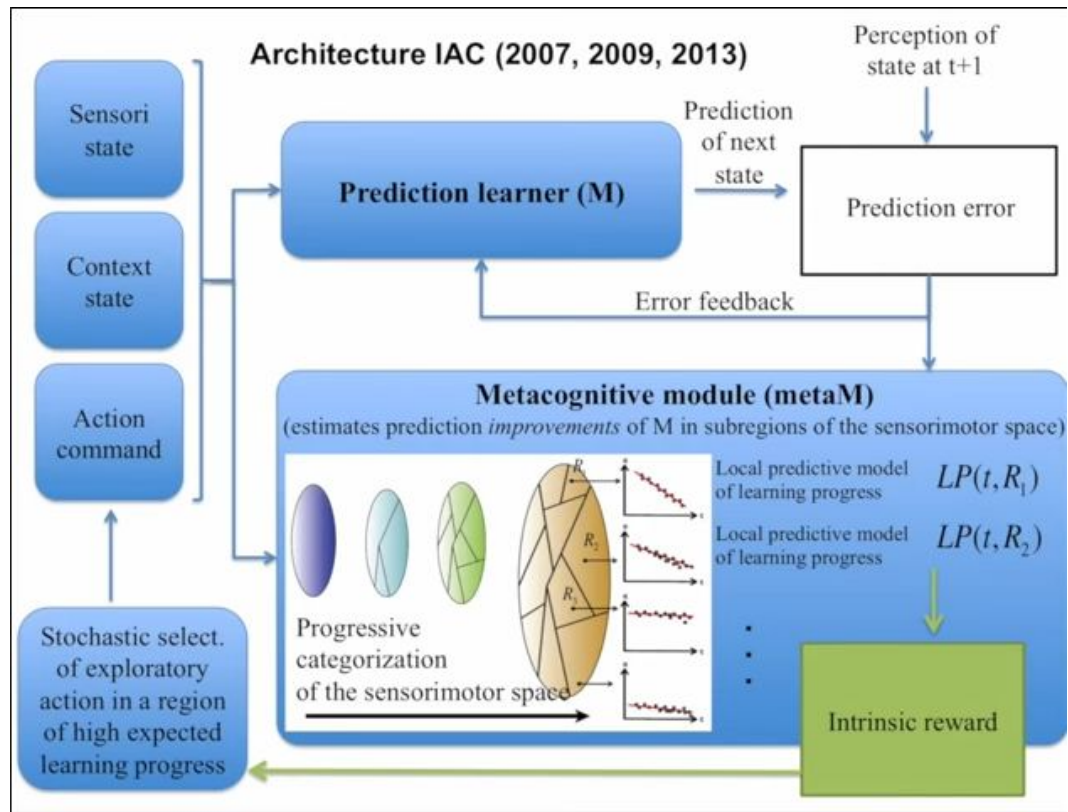
Robots with prediction driven IM

- ▶ Environment
 - ▶ An elephant toy on the ground
 - ▶ A hanging toy
 - ▶ An adult robot which automatically imitates if the infant is looking and speaking
- ▶ 4 parametrized motor primitives
 - ▶ turn head in various directions
 - ▶ Open/close mouth while crouching with varying strengths
 - ▶ Vocalizing with various pitches and lengths
 - ▶ Rocking leg with various angles and speed
- ▶ Sensory perception:
 - ▶ Visual (salient visual properties)
 - ▶ Auditory (other robot)
 - ▶ Tactile (something in mouth)



Oudeyer, Pierre-Yves, et al. "The playground experiment: Task-independent development of a curious robot." Proceedings of the AAAI Spring Symposium on Developmental Robotics. Stanford, California, 2005.

Robots with prediction driven IM



Oudeyer, Pierre-Yves, et al. "The playground experiment: Task-independent development of a curious robot." Proceedings of the AAAI Spring Symposium on Developmental Robotics. Stanford, California, 2005.

Robots with prediction driven IM

No hard-coded strategy

Stage-like development

1. Unorganized babbling (with combined primitives)
2. Explore each primitive randomly
3. Act towards areas with objects, e.g. localize to elephant, rock towards the adult robot
4. Shift focus to contingent action-object pairs

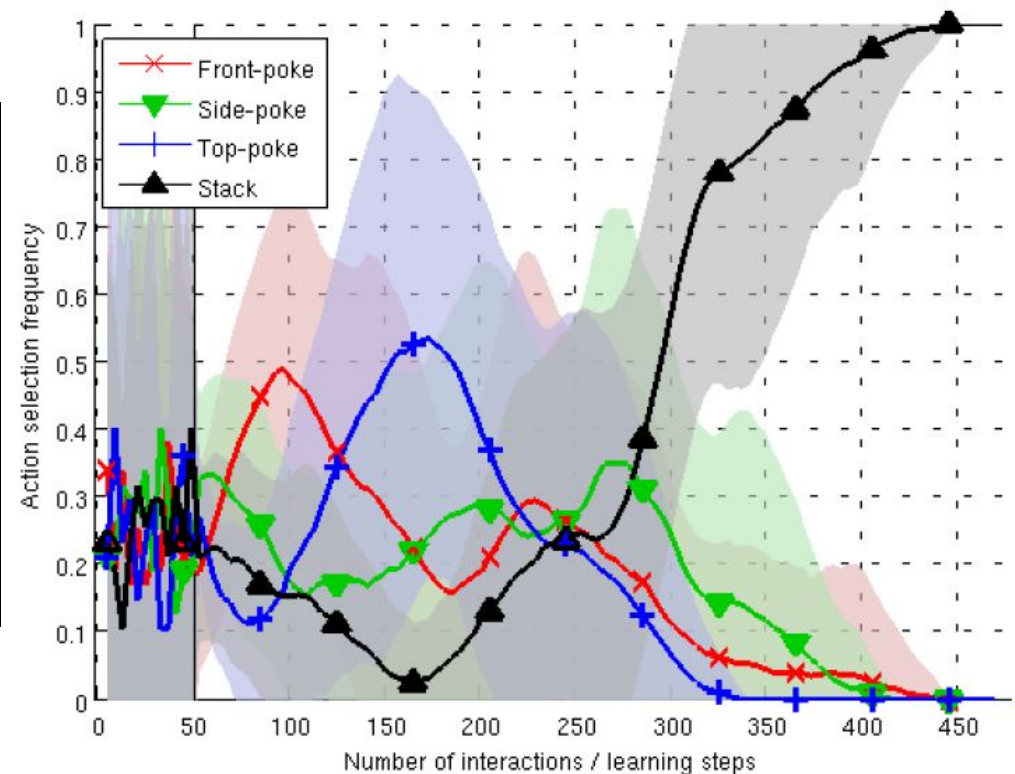
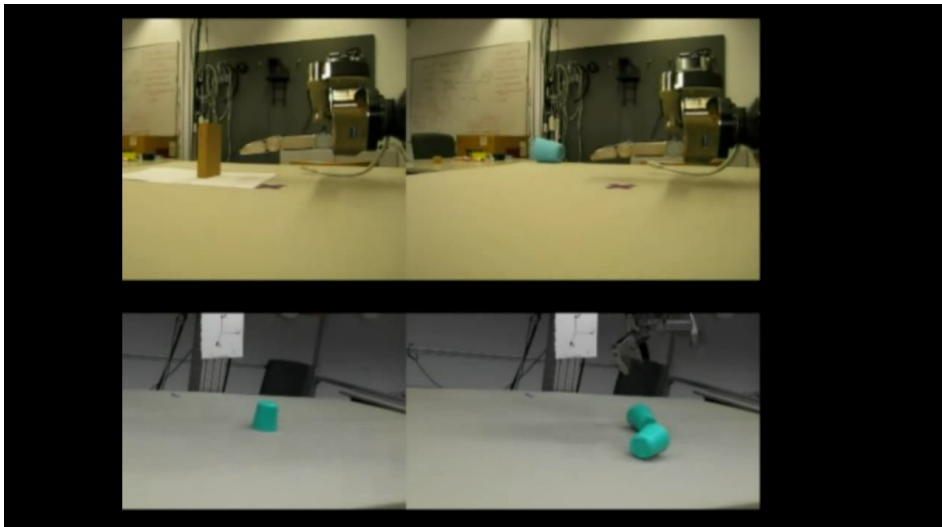
Emergence of communication



Oudeyer, Pierre-Yves, et al. "The playground experiment: Task-independent development of a curious robot." Proceedings of the AAAI Spring Symposium on Developmental Robotics. Stanford, California, 2005.

Robots with prediction driven IM

- ▶ Robot exploring with push and stack actions
- ▶ Learns the effects of actions on different objects
- ▶ Select actions based on learning progress



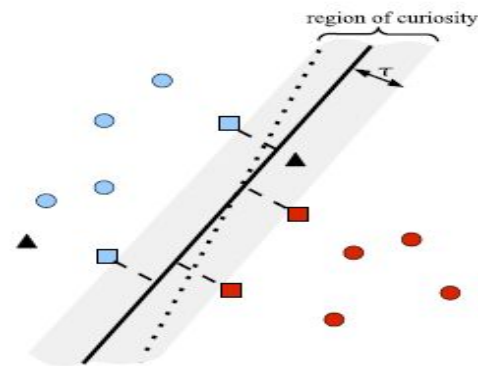
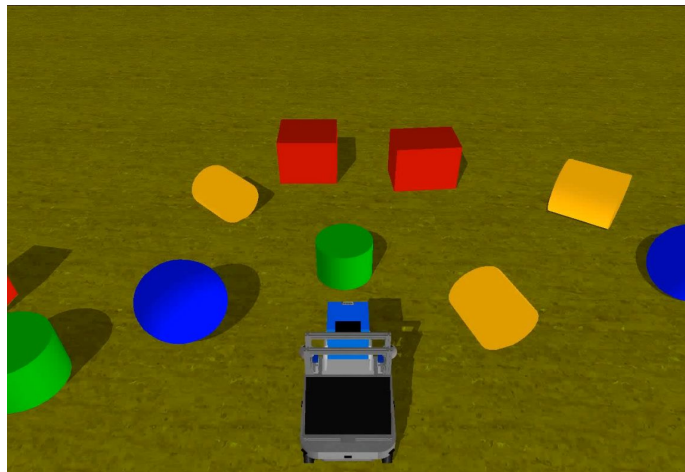
Rewarding IM-based robots based on Competence

► Original formulation

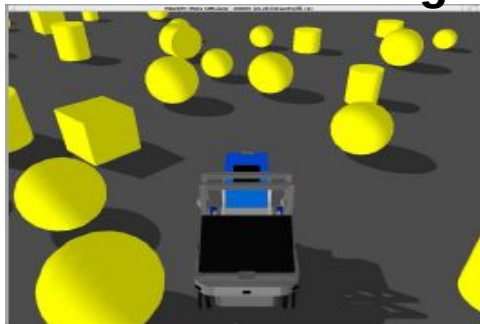
$$I_a(g_k, t_g) = ||\widetilde{g_k(t_g)} - g_k(t_g)||$$

- There are studies which learn contingencies based on surprise signals, but does not exactly follow the formulation above.

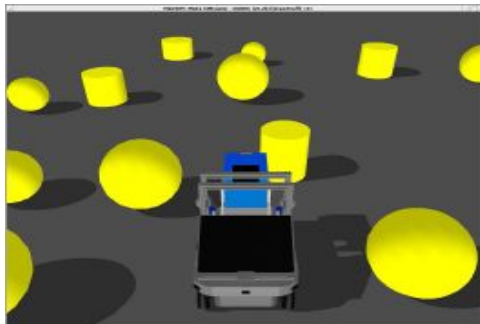
Knowledge (novelty or prediction) or competence based?



Not interesting

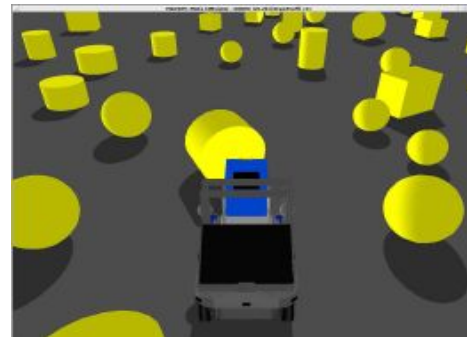


Probably traversable
No object in the vicinity

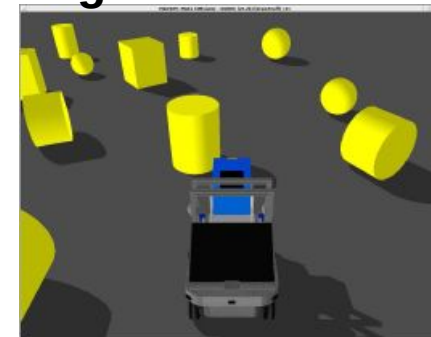


Probably non-traversable
Cylinder object is very close.

Interesting



Cylinder's surface is
similar to sphere's



Object is located
at the boundaries for
Go-forward action

Uğur, Emre, et al. "Curiosity-driven learning of traversability affordance on a mobile robot." Development and Learning, 2007. ICDL 2007