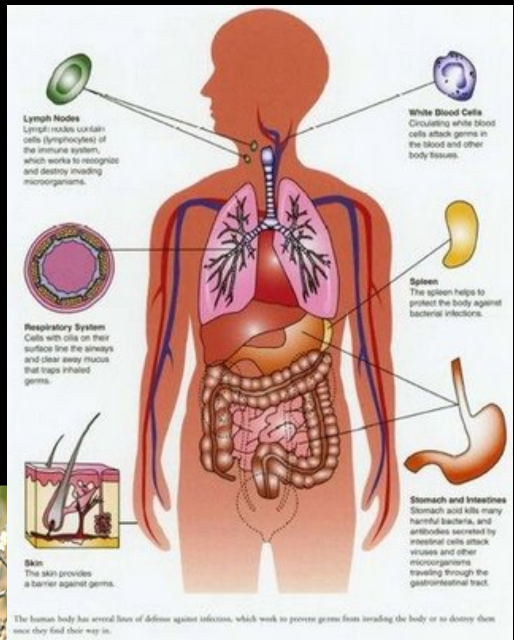
On Self-Organization and Developmental Mechanisms in the Origins of Novel Behavioral and Cognitive Structures: Computational Robotic Models and Experiments

Pierre-Yves Oudeyer
INRIA
http://www.pyoudeyer.com

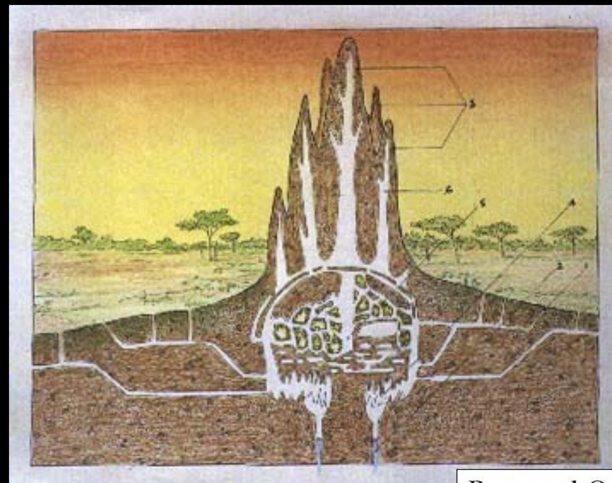
IHES, june 2010

Fascinating spatio-temporal structures in the biological world

Body shapes

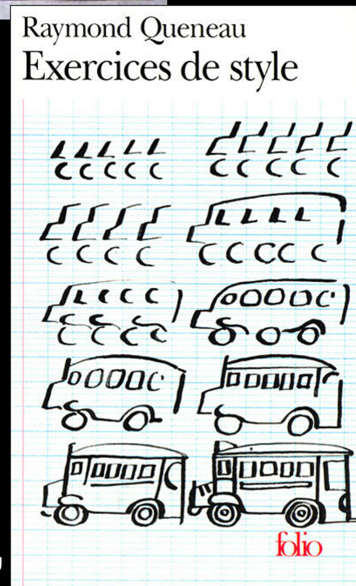


Internal body structural modularity



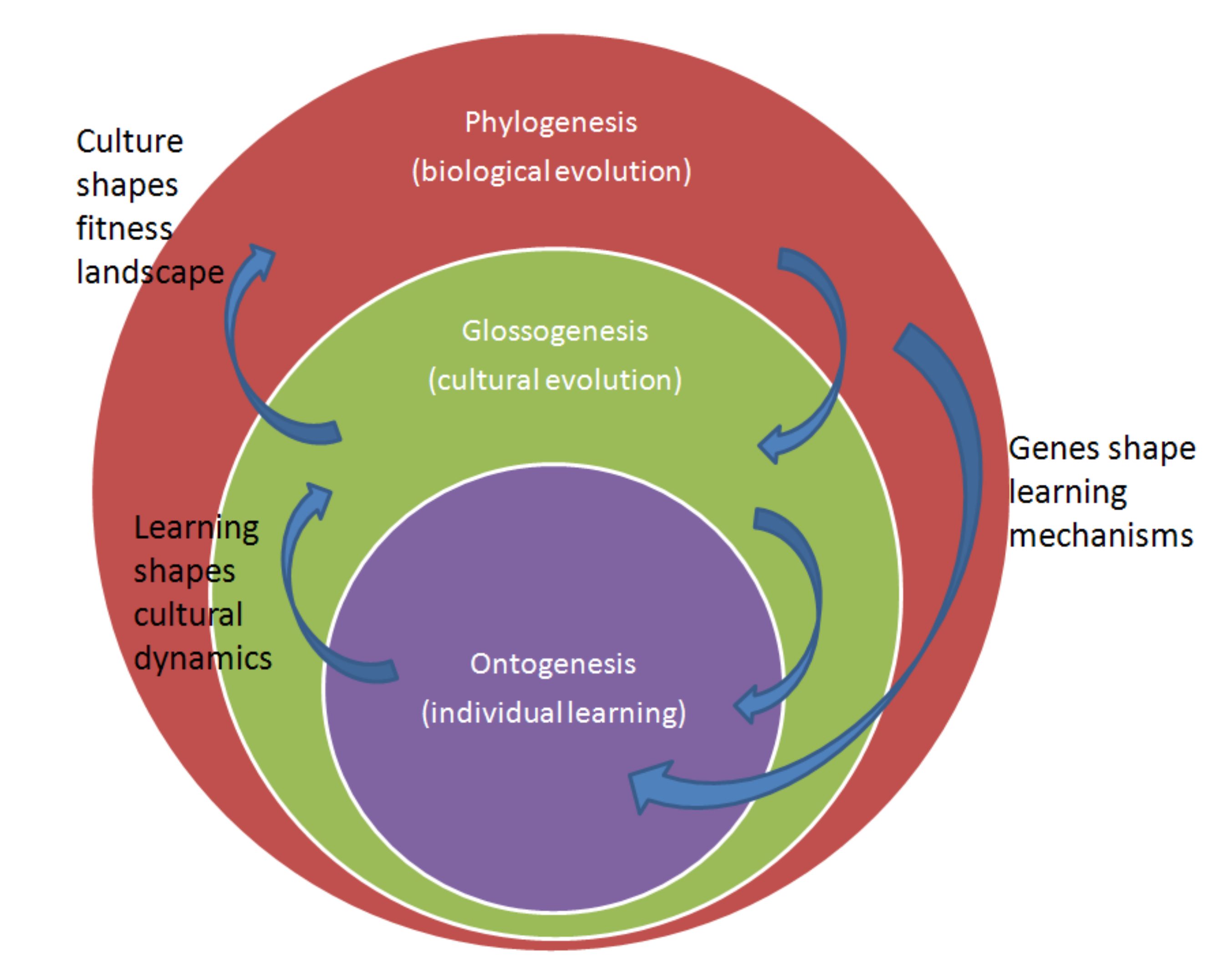
Collective constructions

Motor skills



Cognitive representations,
Abstract social cognitive constructions

Three interacting time scales for the generation and selection/learning of new structures



Exploration, selection and learning

Three commonly thought sources of structure ***innovation***:

1) Iterated random variation and selection at the phylogenetic level: Neo-Darwinian theory explains evolution of structure based on two pillars:
   1) Random variations (at the gene level);
   2) Differential selection of the fittest (to reproduce, yet debates on the level of selection);

   ➔ In practice most effort made on explaining why certain structures were <u>*selected*</u>, but less effort on explaining how they may have been <u>*generated*</u> through iterated "random" variations;
2) Iterated stochastic variations and selection for morpho- and neuro- genesis during ontogeny (e.g. stochastic cell differentiation, neural Darwinism);
3) Learning ontogenetic level: individual organisms acquire novel behavioral and cognitive skills by processing/generalizing "training measures/data" collected through exploration of the world

➔ Most theories of learning, both in humans and artificial systems, focus on <u>*learning/inference mechanisms*</u> but not on <u>*how observations are collected*</u>.

➔ Exploration is vastly understudied

Random exploration is not enough: spaces are very large

Random variations in the space of structures in phenotypic space not enough, especially for explaining relatively sudden formations of novel biological structures

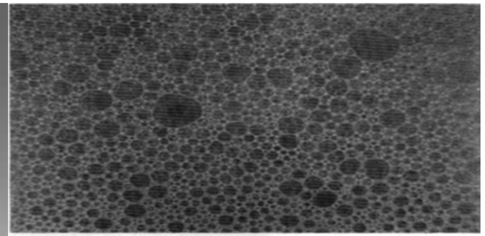
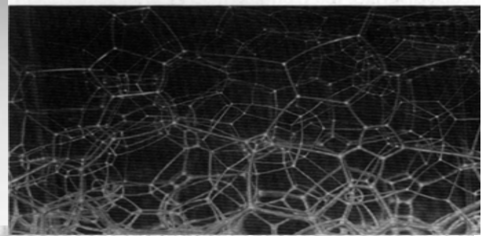
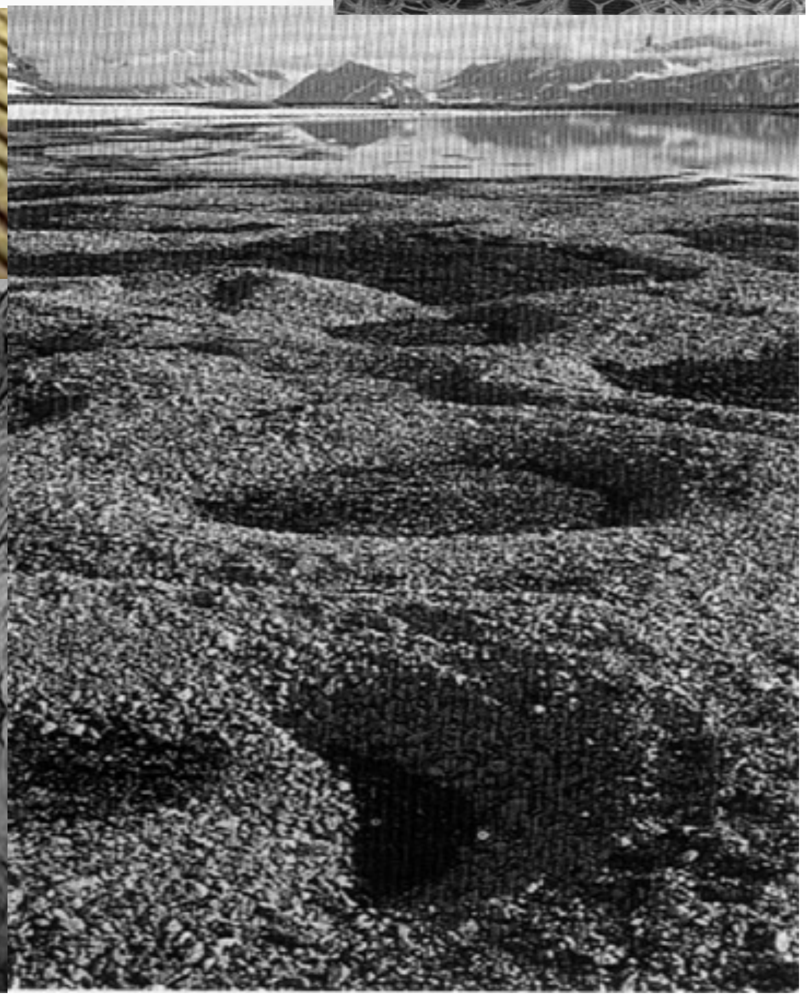
➔ Hopefully, random uniform gene variation do not produce random uniform exploration of the structures in phenotypic space;

Random exploration for learning new skills in ontogeny is hopeless given the ration lifetime/(size of sensorimotor skills that could be learnt) (e.g. just learning the dynamics of one owns body involves learning a manifold with thousands of dimensions, and even worse when interaction with objects and others for which no prior specific models can be imagined, e.g. learning to ride a bicycle or tennis which by the way cannot only be done by just observation of others);

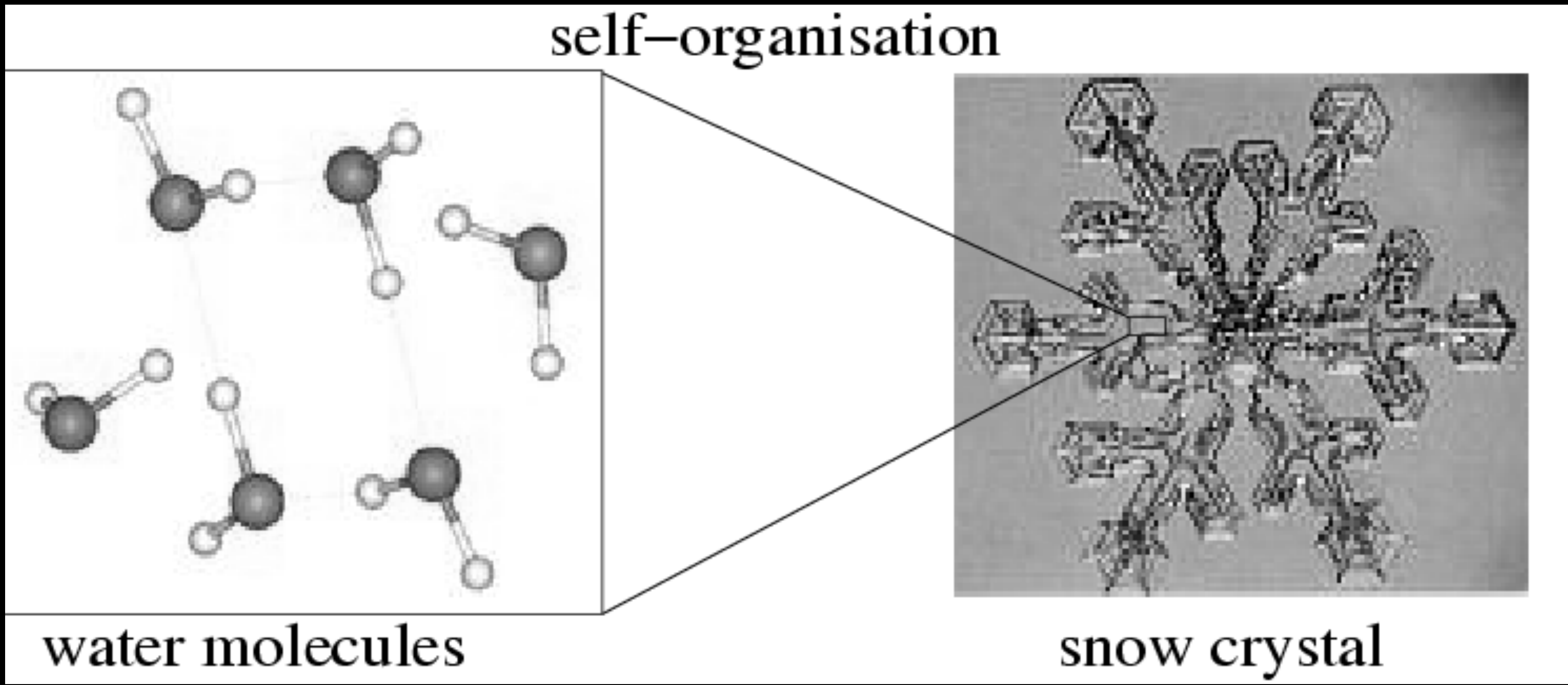
➔ There must be additional ***constraints*** and ***mechanisms that guide exploration*** at all levels

The origins of structures in the non-biological world



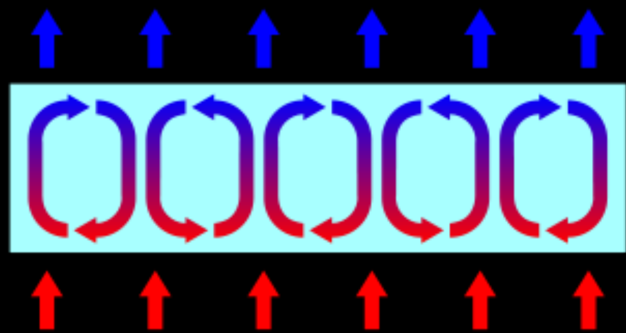

Self-organization in complex systems

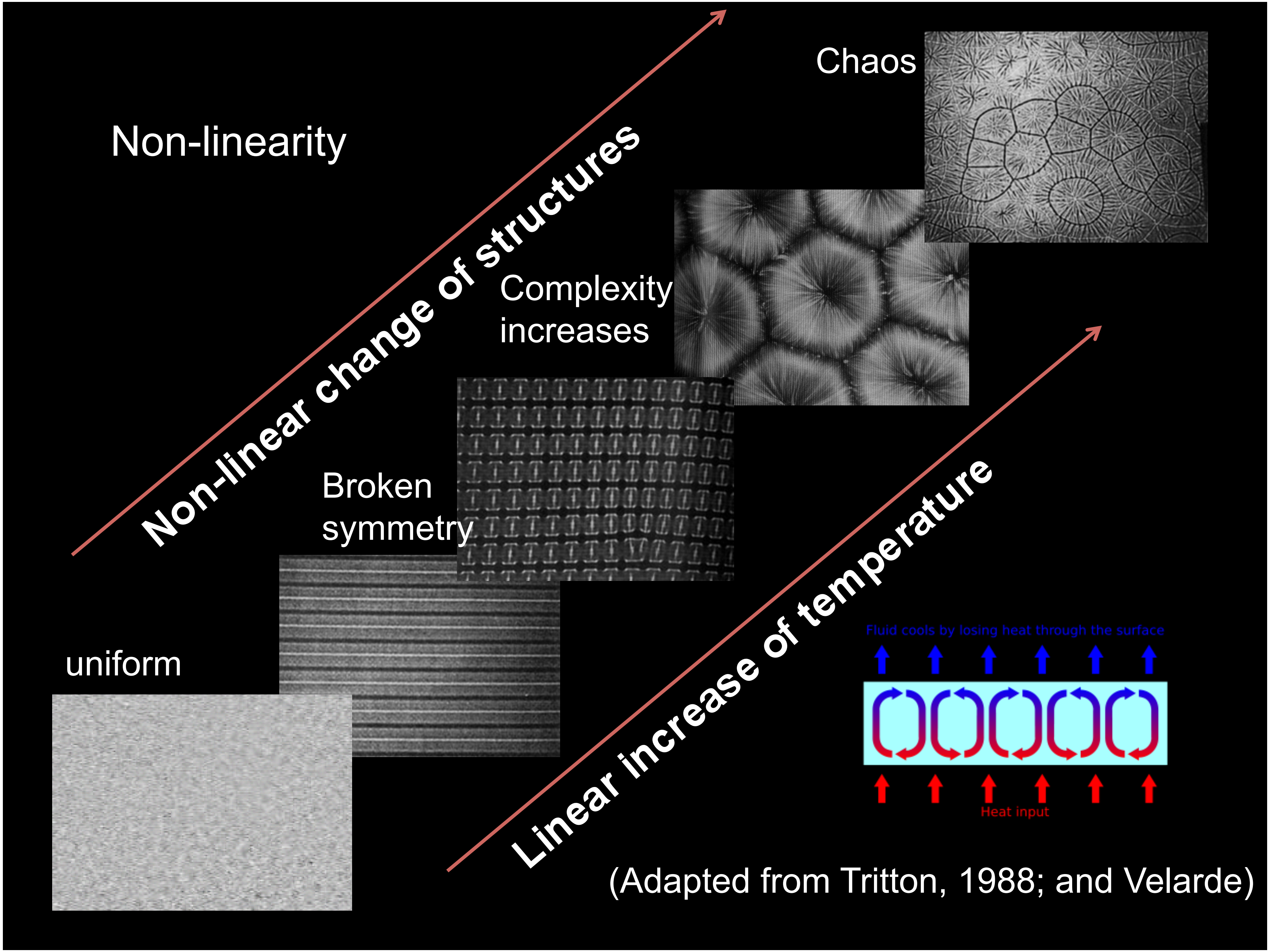


Self-organization: formation at a macro-level of shapes/structures/(as-) symmetries based on low-level physical laws which do not encode explicitly a map of these structures

➔ Typical of complex dynamical systems

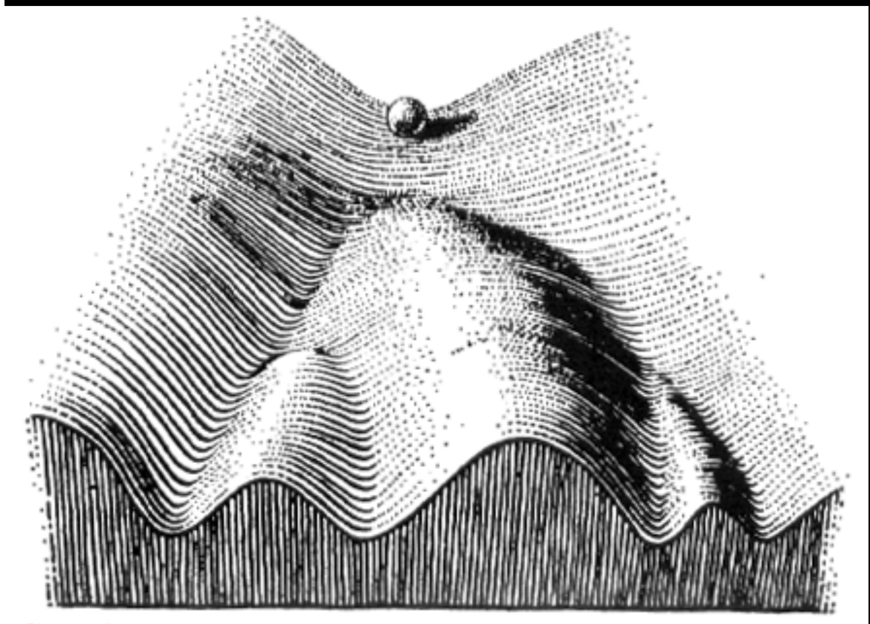
Rayleigh-Bénard cells




Non-linearity
Chaos
Non-linear change of structures
Complexity
increases
Broken
symmetry
uniform
Linear increase of temperature
Fluid cools by losing heat through the surface
Heat input
(Adapted from Tritton, 1988; and Velarde)

Biology is full of interconnected complex systems

→ Similar structure formation mechanisms also at play also

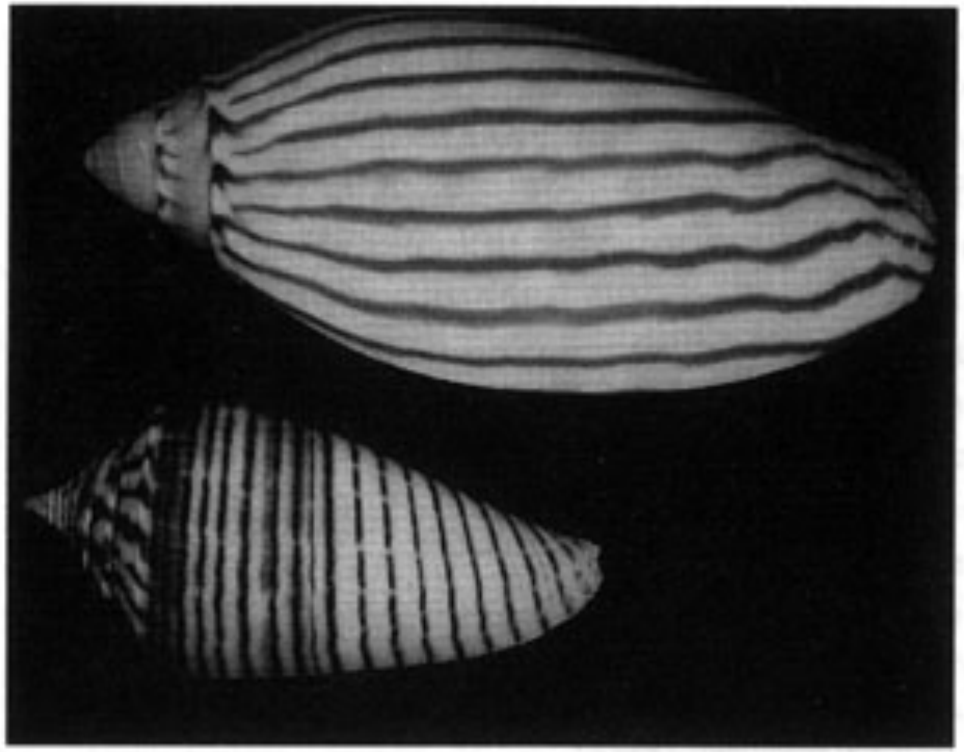


Epigenetic landscape of Waddington
(Waddington, 1956)

Self-organization is at all scales in developmental systems, in particular in embryogenesis and epigenesis

➔ Constraints on the space of forms/ structures, which are not all equally easy to generate given their bio-physical substrate
(D'arcy Thompson, 1917)

Example of structures that highlight the role of self-organization as a complement to Darwinian explanations



(Photo: H. Meinhardt)

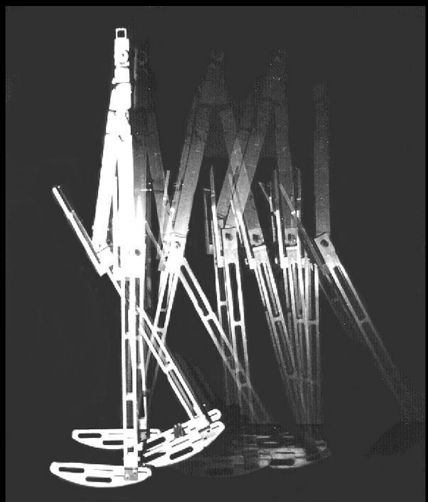
From structure to function: natural selection of self-organized structures (exaptation)

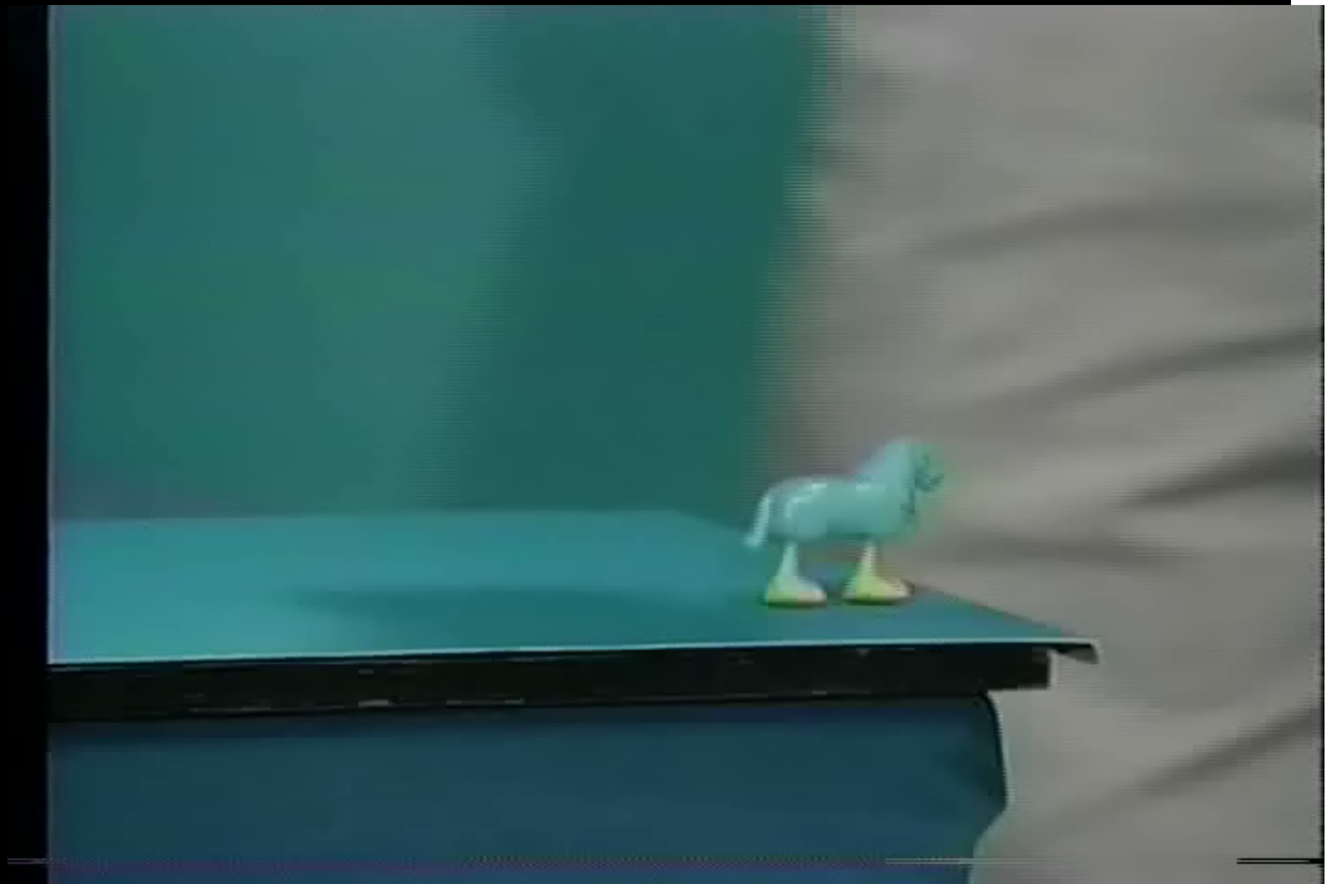
Spontaneous dynamical order in ontogeny

Passive dynamic walkers:

Morphology and physics provide constraints on possible movements

➔ Learning to walk consists in exploring how to modify and control this complex intrinsic dynamics, not in exploring all mathematically possible trajectories of the body parts.





Tad McGeer (McGeer, 1990)

Recursive self-organization

- ➔ ***Self-organization*** adds upon random variations to foster the non-linear generation of complex organized structures which are then selected;
- ➔ These emergent structures in turn introduce new constraints and potentialities for self-organization that guide further exploration and selection of novel structures;
- ➔ In particular, emergence of mechanisms for non-random genetic variations, as well as of ***mechanisms for explicitly organized spontaneous exploration in ontogeny ➔ see self-motivated curiosity driven learning later in the talk***;
- ➔ Recursive self-organization and constraints on exploration

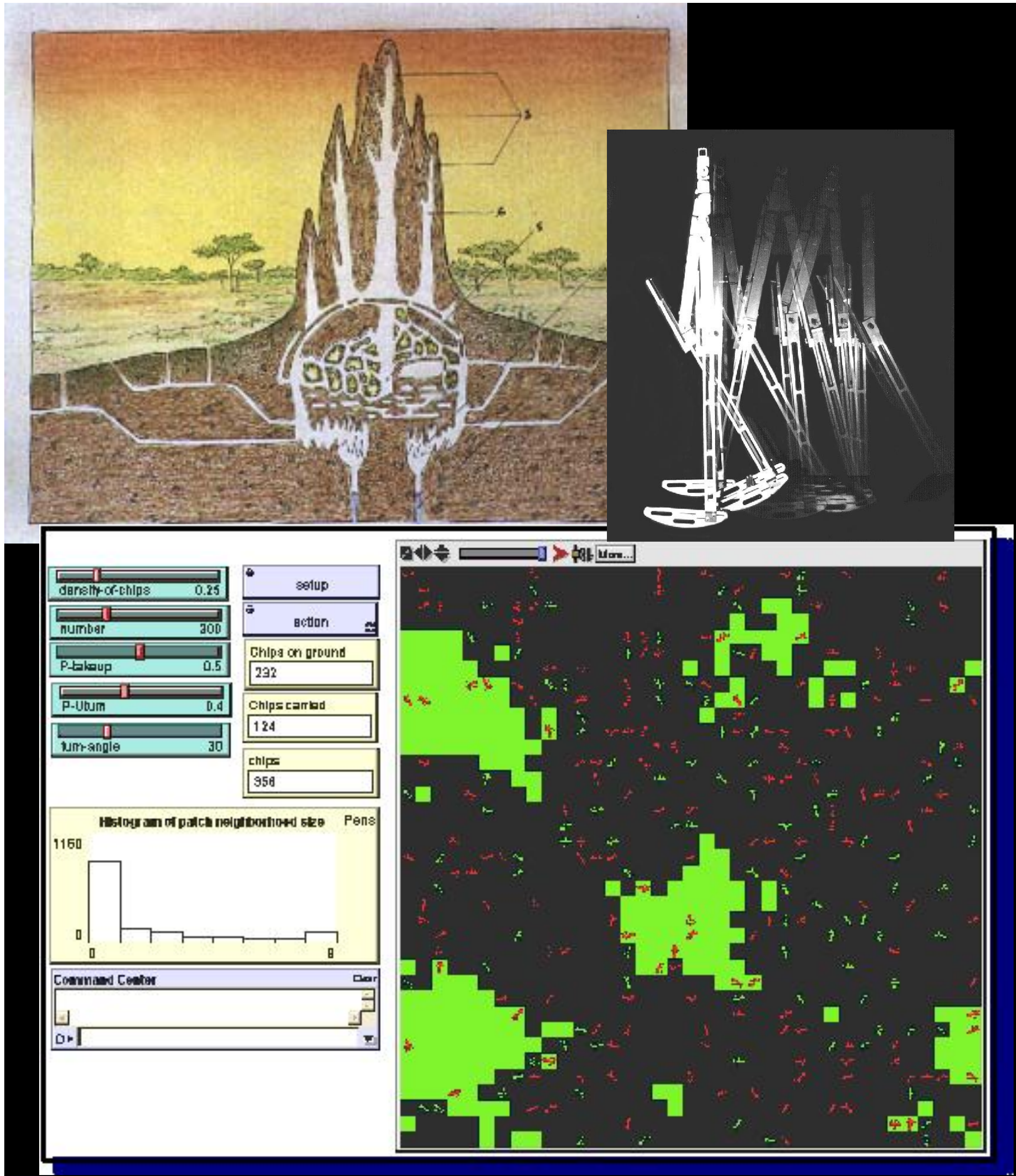
Studying those families of constraints for understanding the exploration and formation of novel behavioural and cognitive structures

- As an original complement to many studies done so far on models of the growth of the body (models of morphogenesis, in particular during embryogenesis)

- At both the individual and social levels, e.g.

→ Origins of language (coupled phylogenetic, glossogenetic and ontogenetic levels)

→ Learning of new sensorimotor skills (ontogenetic level)

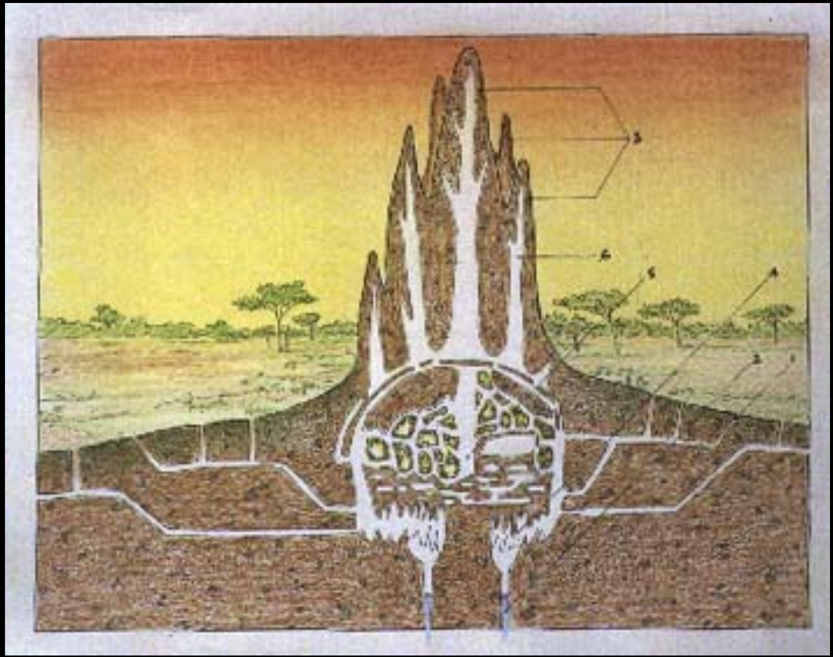


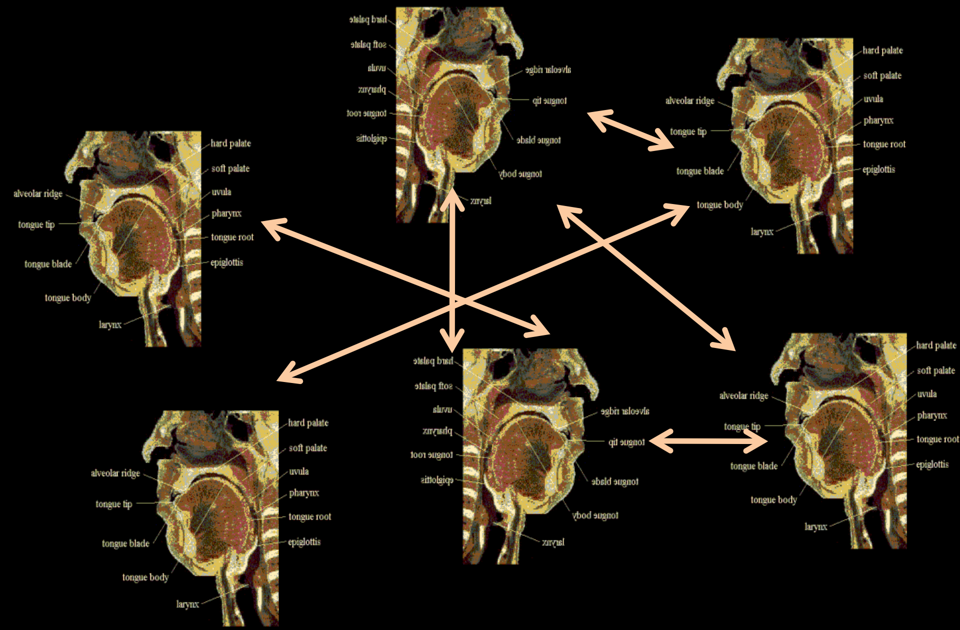
Tools: computational and robotic models and experiments

Explore the landscape of complex-system mechanisms, self-organization mechanisms in particular, to ***enhance our intuitions***

➔Stimulate reflexion by exploring new hypothesis spaces and verify the coherence of existing hypothesis
➔Organize the scientific debate: a meta-scientific activity

➔ Embryology (ex. Turing and morphogenesis, (Turing, 1952), Complex evolutionary dynamics (e.g. Maynard Smith and theoretical biology, Maynard-Smith, 1968), ethology of insect societies (e.g. Camazine et al., 2001);





Self-organization in the evolution of language and speech

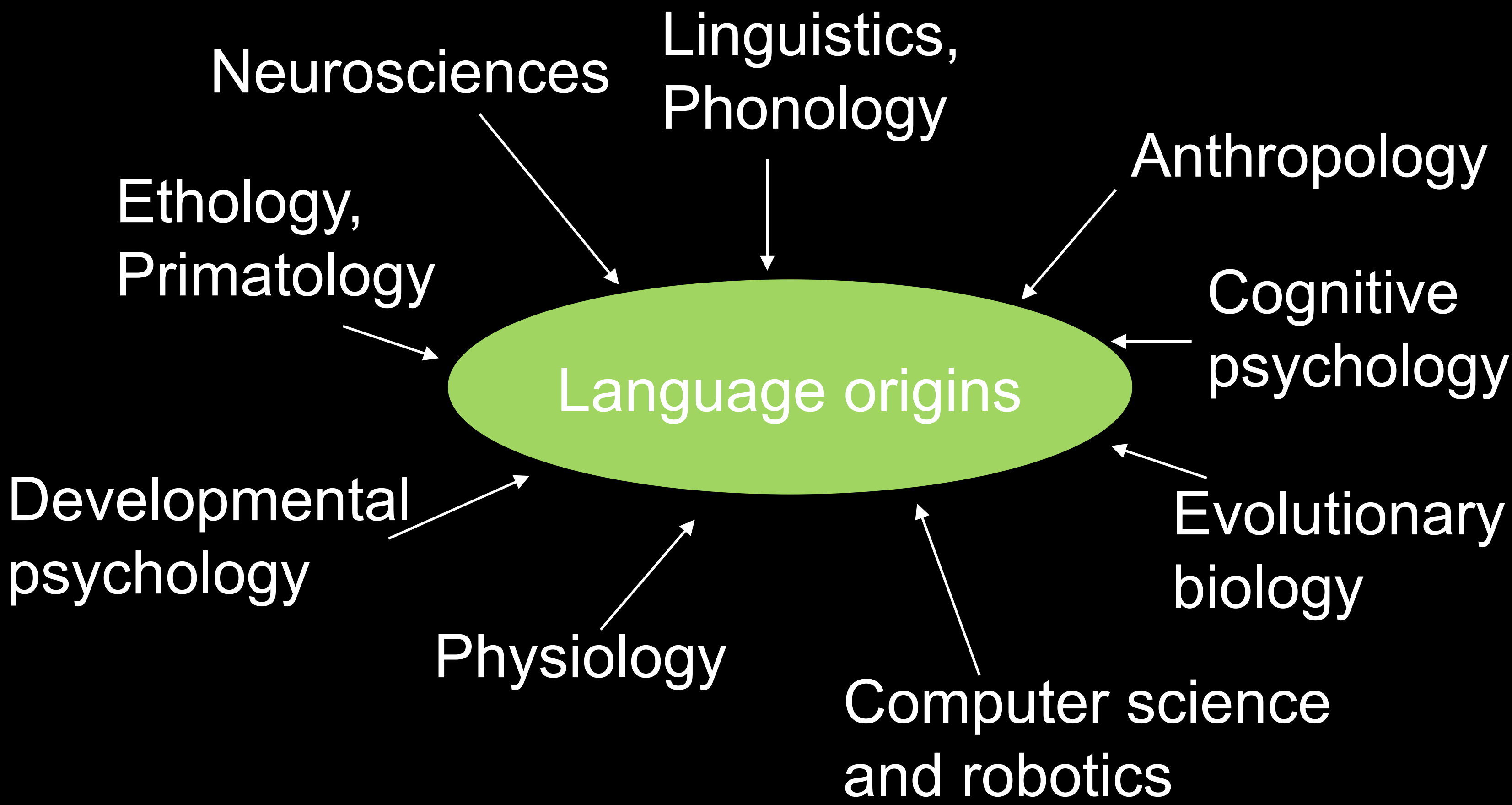
Michael Studdert-Kennedy, Peter McNeilage, Björn Lindblöm, James Hurford, …

Different kinds of questions

On the origins and evolution of language and languages:

- **Why** did language evolve? What were the selection pressures and the ecology and social context?
- **How** did language evolve? What are the biological and/or cultural innovations that were necessary for this? How were they generated?
- **Why** does language has the structure/forms it has and not others? **How** were they generated/found?

Multi-disciplinarity



Computational approaches to the origins of language

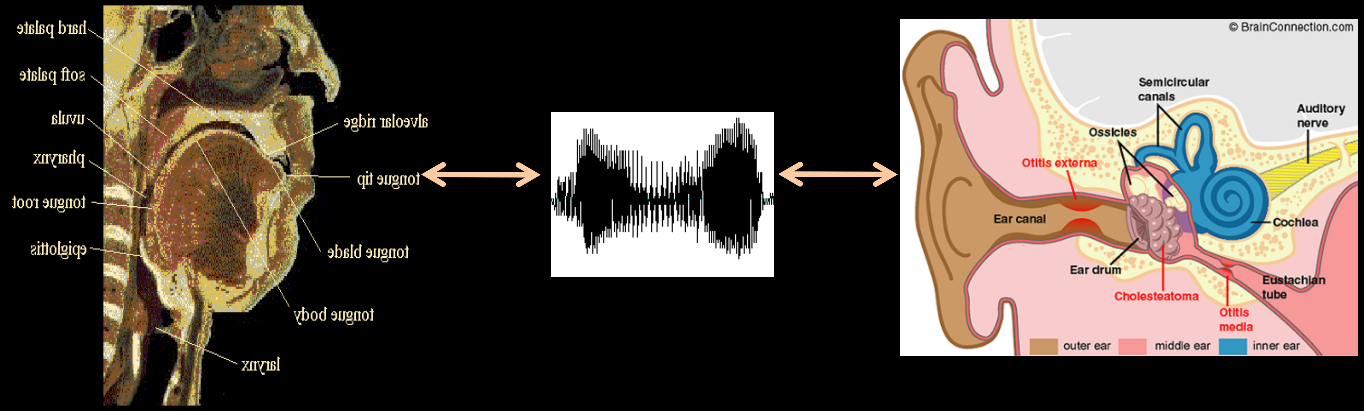
Formation of **lexical conventions** : Steels, Kaplan, Cangelosi, Parisi, Hurford, Smith, Vogt, Nowak, Niyogi, Komarova, Brighton,…

Formation of **shared categorization systems** (meanings): Kaplan, Steels, Loeveren, Brighton, Harnad, Cangelosi, Elman, …

Origins of **syntax**: Kirby, Batali, Steels, Nowak, Zuidema, Hurford, Komarova, Niyogi, Cangelosi, …

Origins of **grammar** : Steels, Chang, Bergen, …

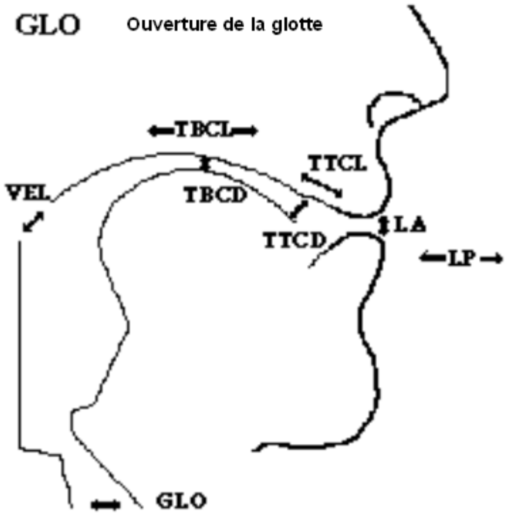
Origins of **speech**: Glotin, Berrah, de Boer, Oudeyer, Goldstein, …

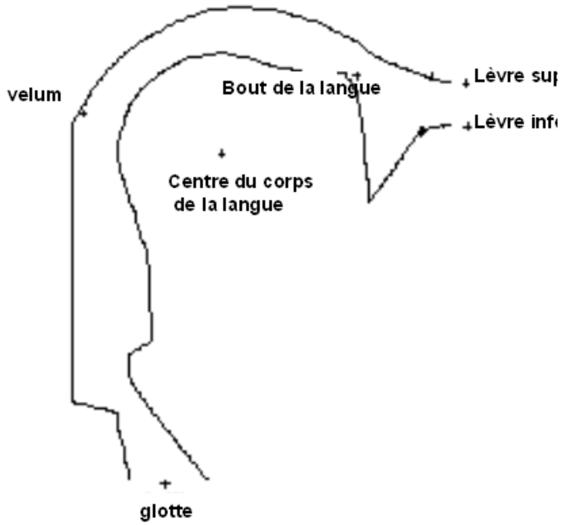

hard palate
soft palate
uvula
pharynx
tongue root
epiglottis
alveolar ridge
tongue tip
tongue blade
tongue body
larynx
© BrainConnection.com
Semicircular canals
Auditory nerve
Ossicles
Otitis externa
Ear canal
Cochlea
Ear drum
Eustachian tube
Cholesteatoma
Otitis media
outer ear
middle ear
inner ear

Speech

Physiologic grounding of speech

Variables du conduit vocal		Organes impliqués
LP	Protrusion des lèvres	lèvres inférieure et supérieure, machoire
LA	Ouverture des lèvres	lèvres inférieures et supérieure, machoire
TTCL	Lieu de constriction du bout de la langue	bout et corps de la langue, machoire
TTCD	Degré de constriction du bout de la langue	bout et corps de la langue, machoire
TBCL	Lieu de constriction du corps de la langue	corps de la langue, machoire
TBCD	Degré de constriction du corps de la langue	corps de la langue, machoire
VEL	Ouverture du velum	velum
GLO	Ouverture de la glotte	glotte





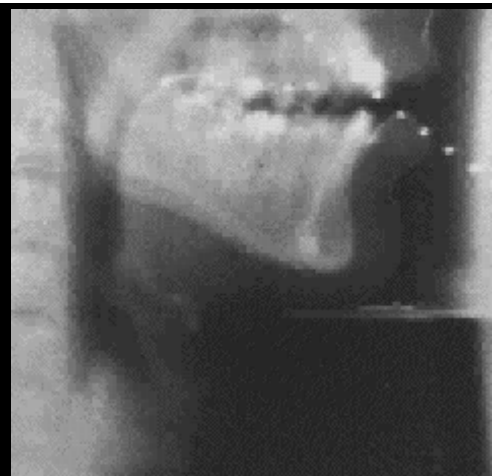
Activations de la cochlée

↕

Représentation gestuelle = lieux et manières des constrictions

↕

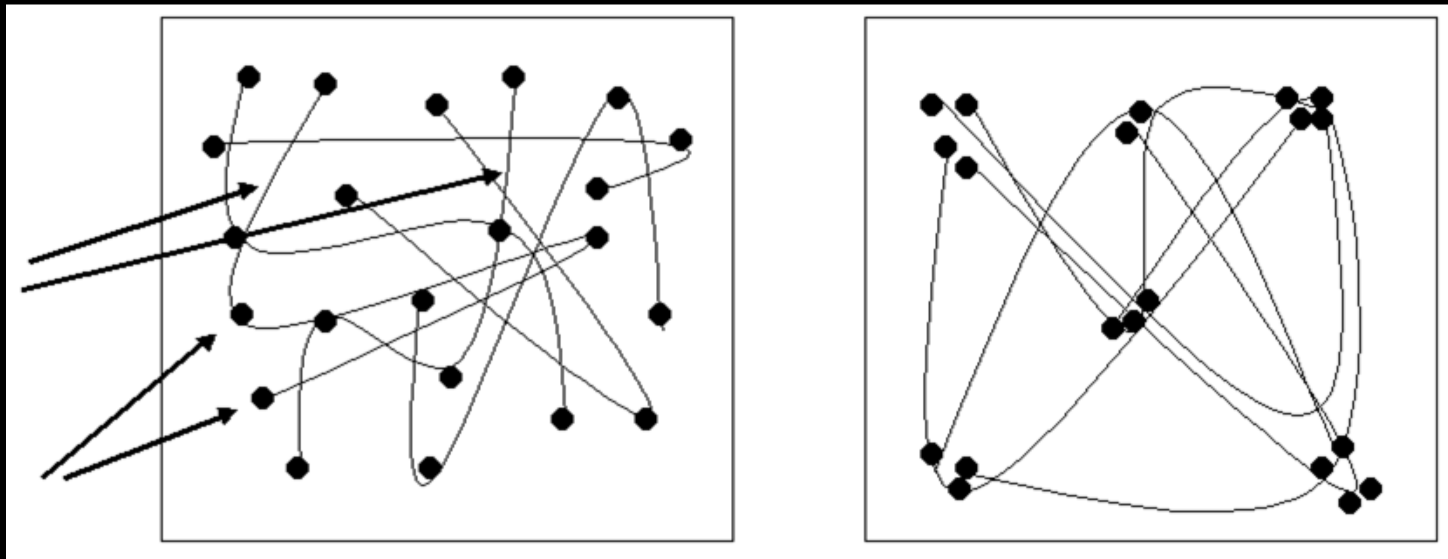
Activations musculaires



1) Speech is a shared conventional discrete combinatorial system

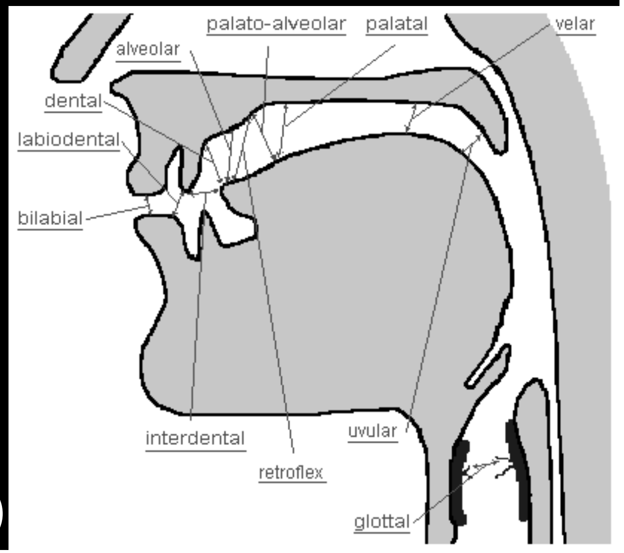
Vocalisations

Articulatory targets



Non combinatorial system   Combinatorial system

The repertoire of vocal gestures is shared in a given linguistic community, but different in different communities



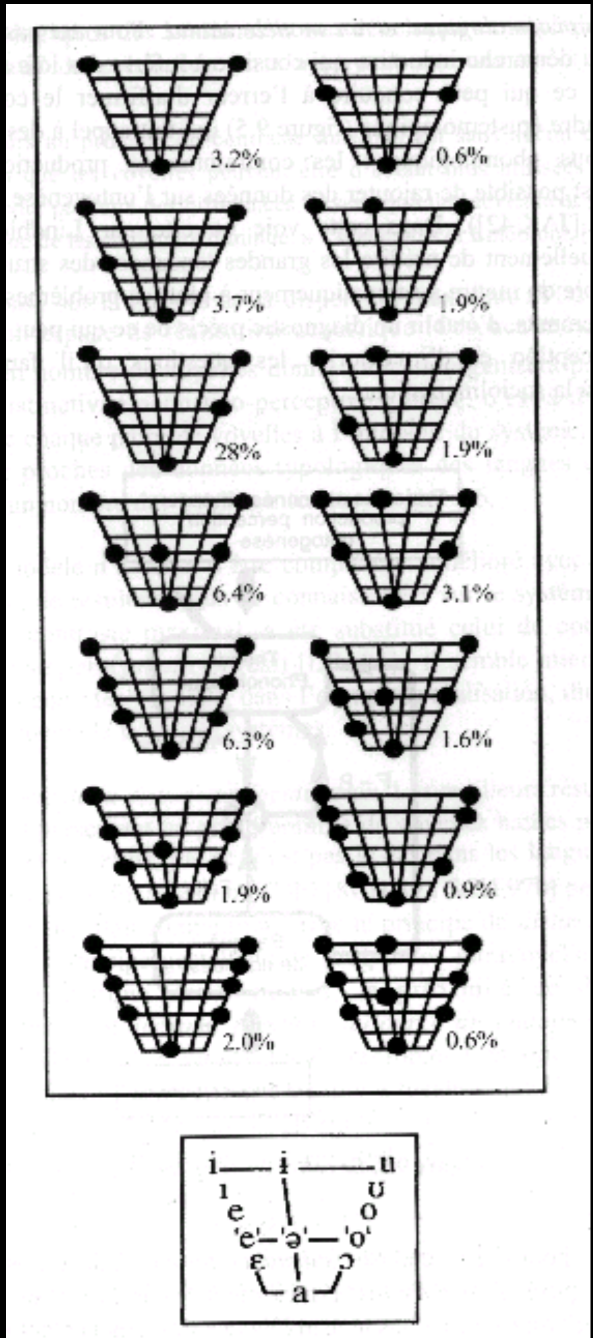
(Adapted from Bickford and Tuggy, 2002)

2) Speech perception is also language specific

→ Japanes don't hear the different between the [l] in « lead » and the [r] in « read »

→ French do not hear the difference among cantonese « ma » with different tone and meaning: ma ma ma ma (mère, cheval, jurer, haschish)

3) Universals and diversity



C	%	C	%
t	97.5	g	56.2
m	94.4	ŋ	52.7
n	90.4	ʔ	48.0
k	89.5	tʃ	41.8
j	84.0	ʃ	41.6
p	83.3	f	40.0
w	76.8	dz	34.9
s	73.5	ɲ	31.3
d	64.7	ts	29.3
b	63.8	k^h	22.9
h	62.0	p^h	22.4
l	56.9	v r	21.1

+ phonotactics

(Schwartz et al., 1997)

Questions on the origins of speech

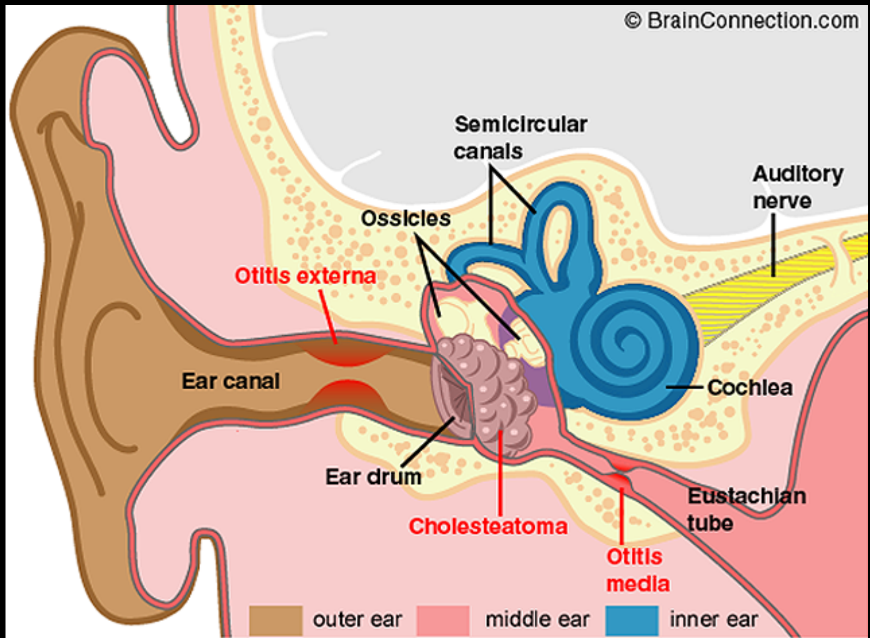
1) What are the biological pre-requisites allowing the formation of speech codes? Do they correspond to major or minor biological changes?

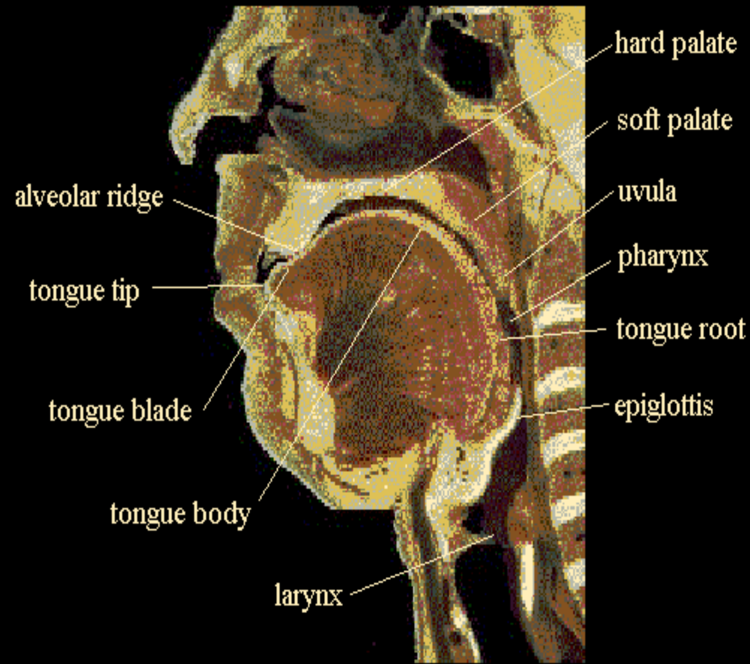
2) How can a community of individuals can come to share one single code among many possible codes?

3) Why does speech have such a structure? How was this structure generated?

The morpho-functional approach

Constrained functionalism:





Communicative function + morphological constraints

Macroscopic models
(Lindblom and Liljencrants, 1972)

Energy of a vowel system =

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{d^2_{S_i S_j}} + \beta \sum_{i=1}^{n} \left(\frac{1}{Glob\,Eff_{S_j}} \right)$$

Inter-syllabic acoustic distinctivity

$= \dfrac{\text{Acoustic salience}}{\text{Articulatory cost}}$

Minimisation on the set of possible vowel systems:
➔ we find the most frequent vowel systems in humans

Limits

1) This kind of model does not explain how this optimiztion might be achieved in nature (or culture), and whether « good » solutions were « easy » or difficult to find;

➔ Classical neo-Darwinian explanation with no reference to the problem of search and exploration

2) This does not explain how a community can « choose » a speech code rather than another one;

➔ Search/exploration (and convergence) mechanisms are lacking from explanation!

2) This does not explain the universals/diversity duality;

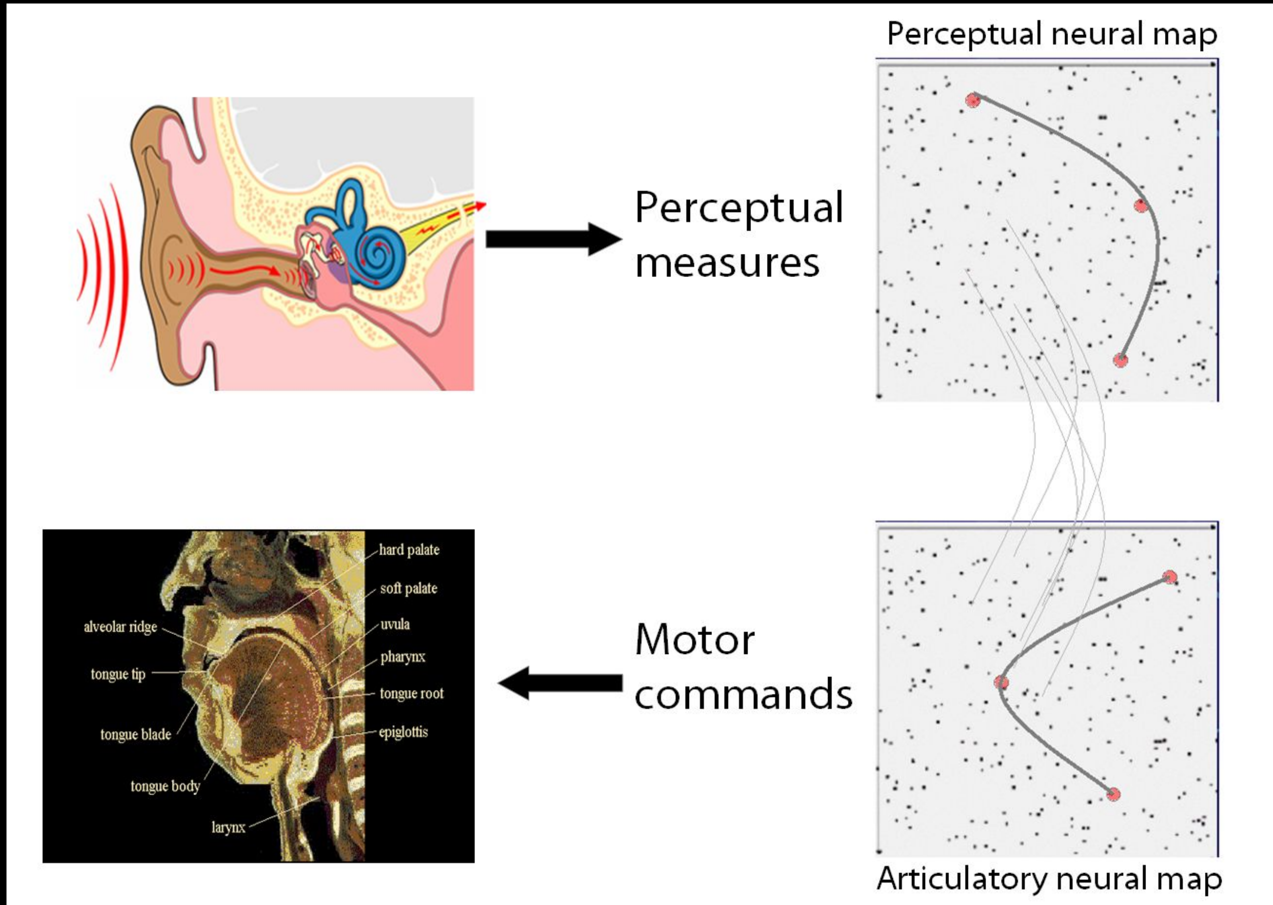
An experiment to stimulate our thinking of the origins of speech in a pre-linguistic context (and not their evolution today)

Oudeyer, P-Y. (2006) Self-Organization in the Evolution of Speech, Studies in the Evolution of Language, Oxford University Press. (Translation by James R. Hurford)

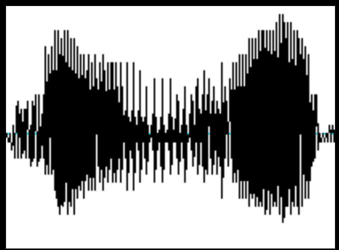
Oudeyer, P-Y. (2005) The Self-Organization of Speech Sounds, Journal of Theoretical Biology, 233(3), pp. 435--449.

The basic neural kit for basic holistic/analogic vocal imitation

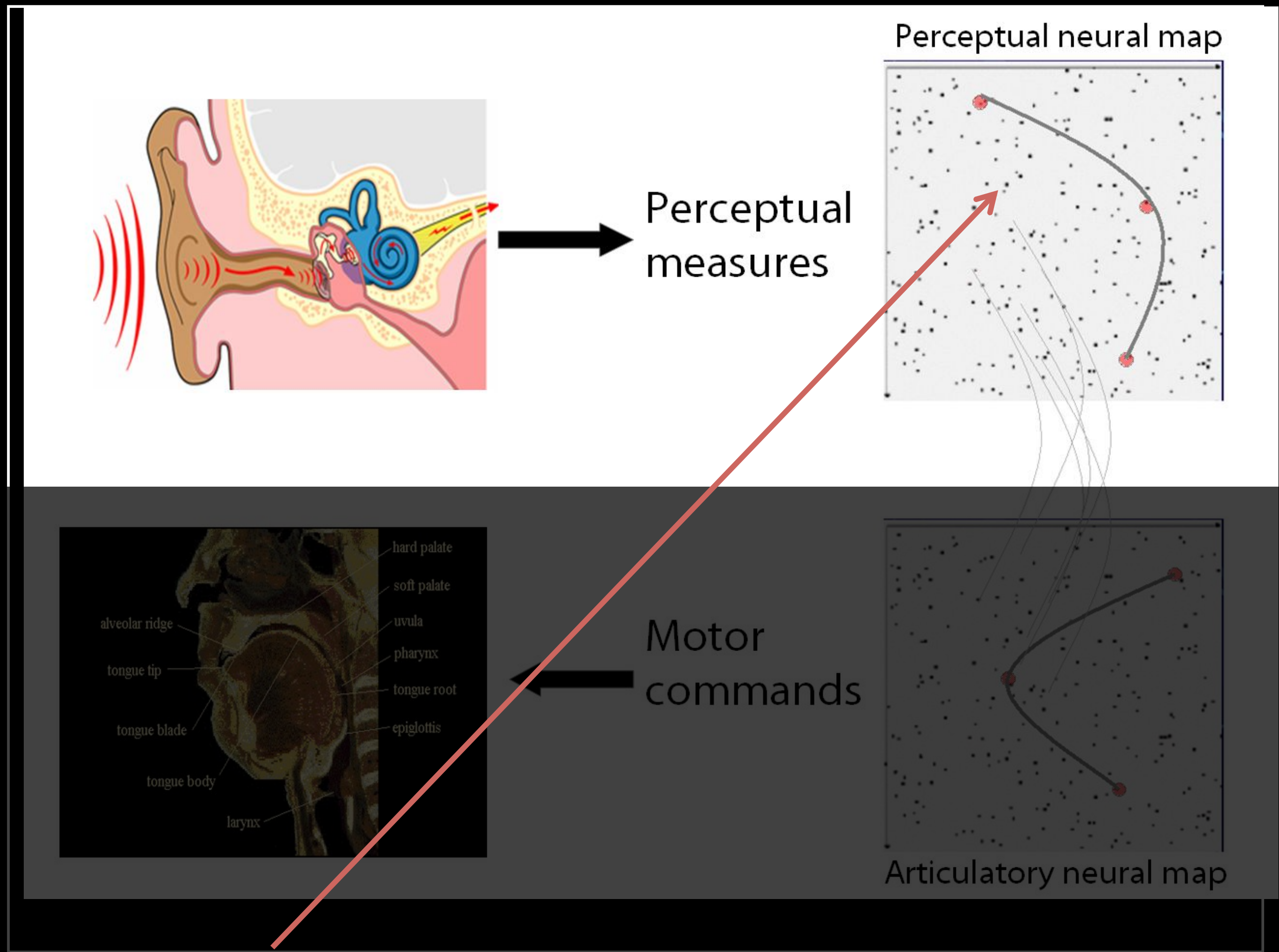
agent's sensori-motor architecture



(Passive) Plastic learning of sound categories

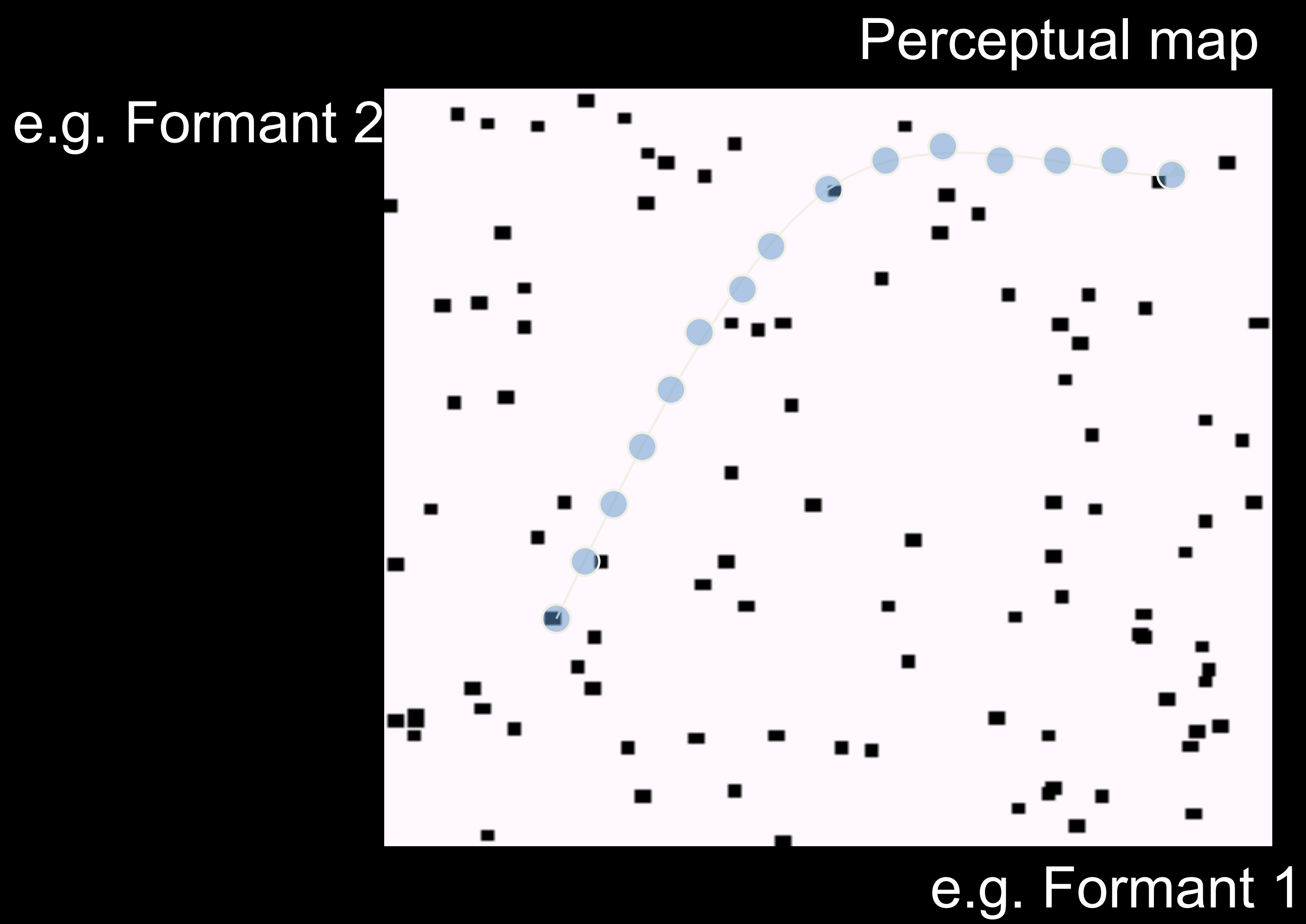


Learning the distribution of sounds in the environment (classical kohonen maps)

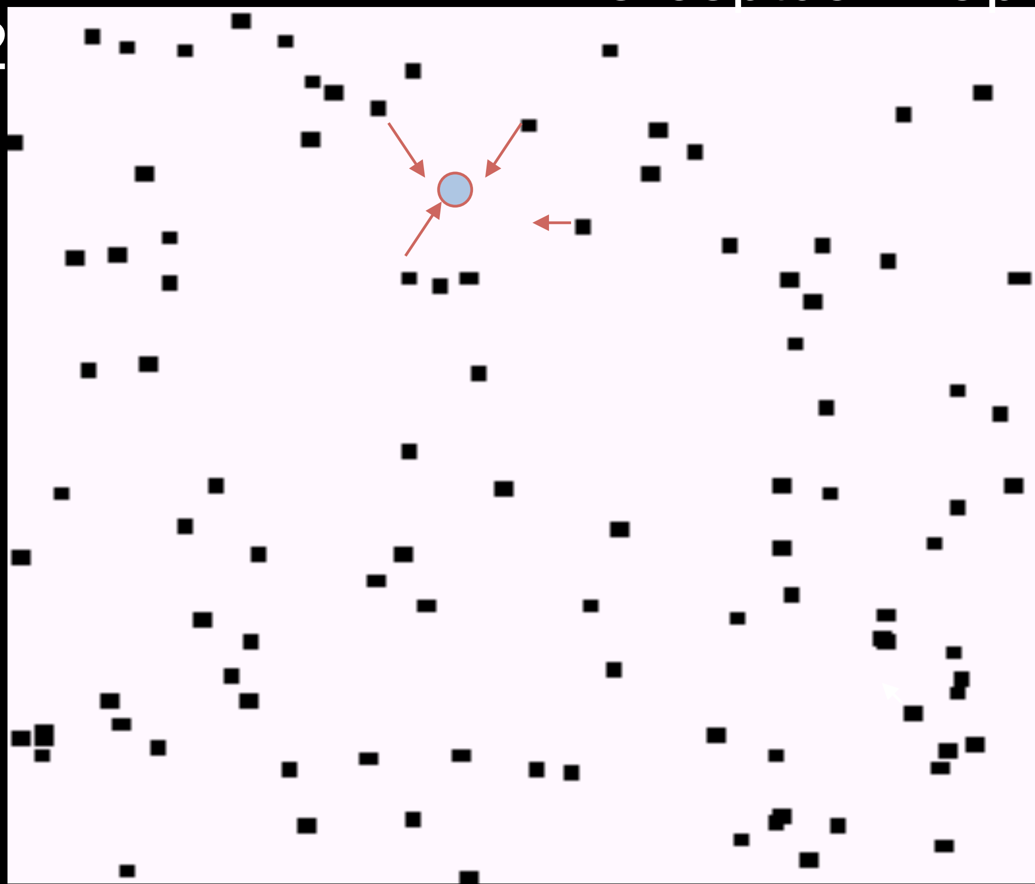


Plasticity of auditory/perceptual map

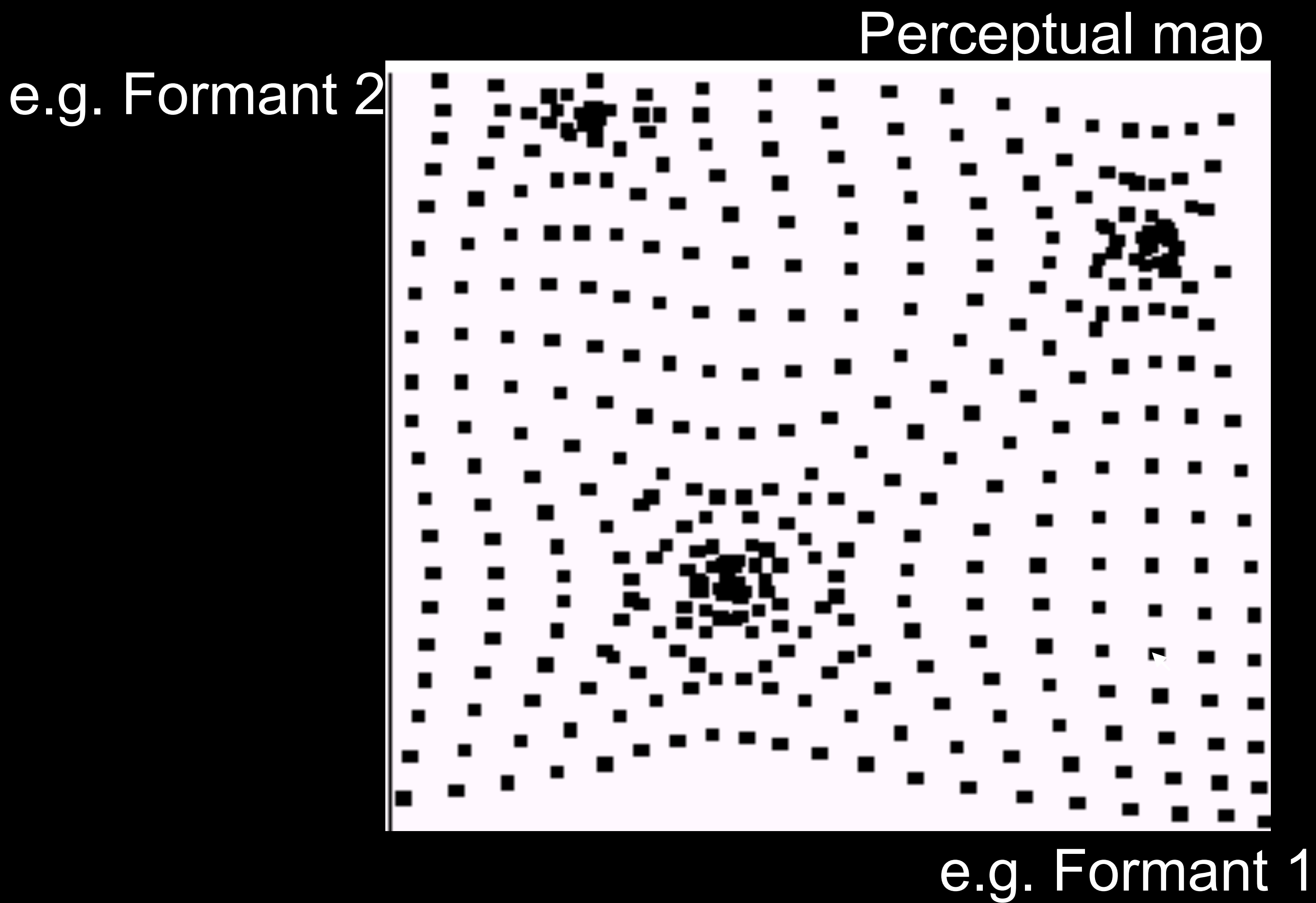
Perceiving a vocalization



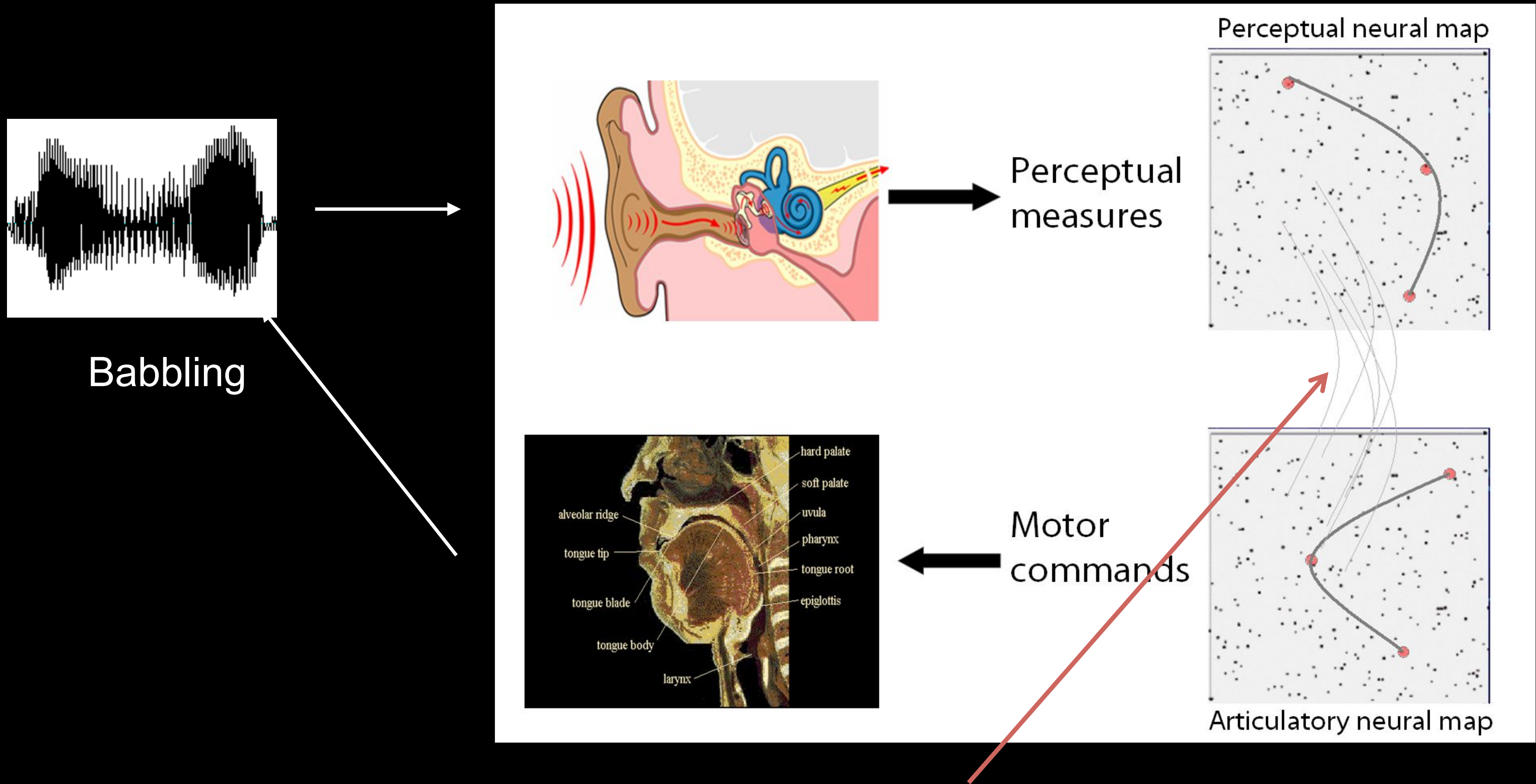
Plasticity inside neural maps



Examples of learnt « modes »



Plastic learning of articulatory-motor correspondences through babbling



Babbling

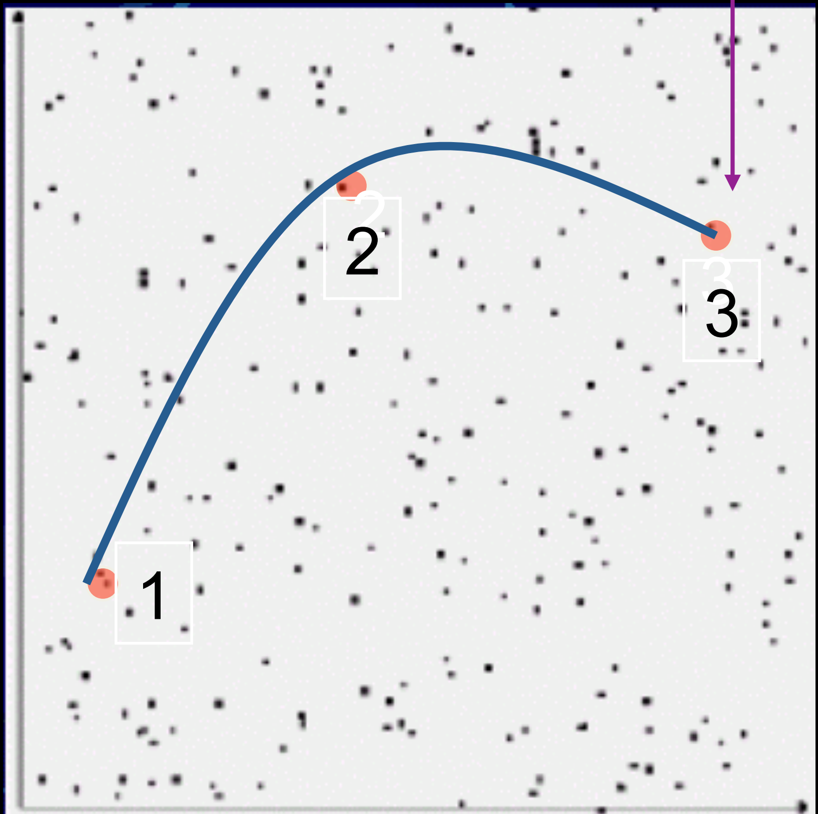
Hebbian plasticity of intermodal connections

Motor babbling

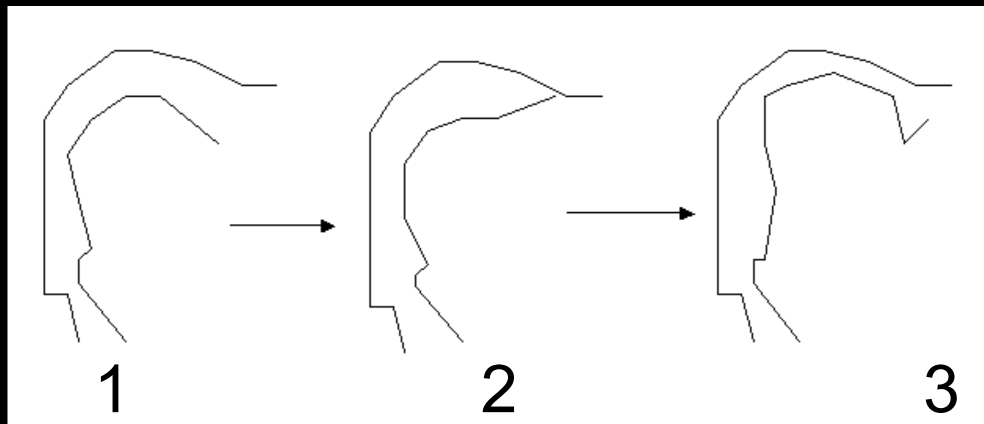
Dimension 2
e.g. : lip
rounding

Neural map

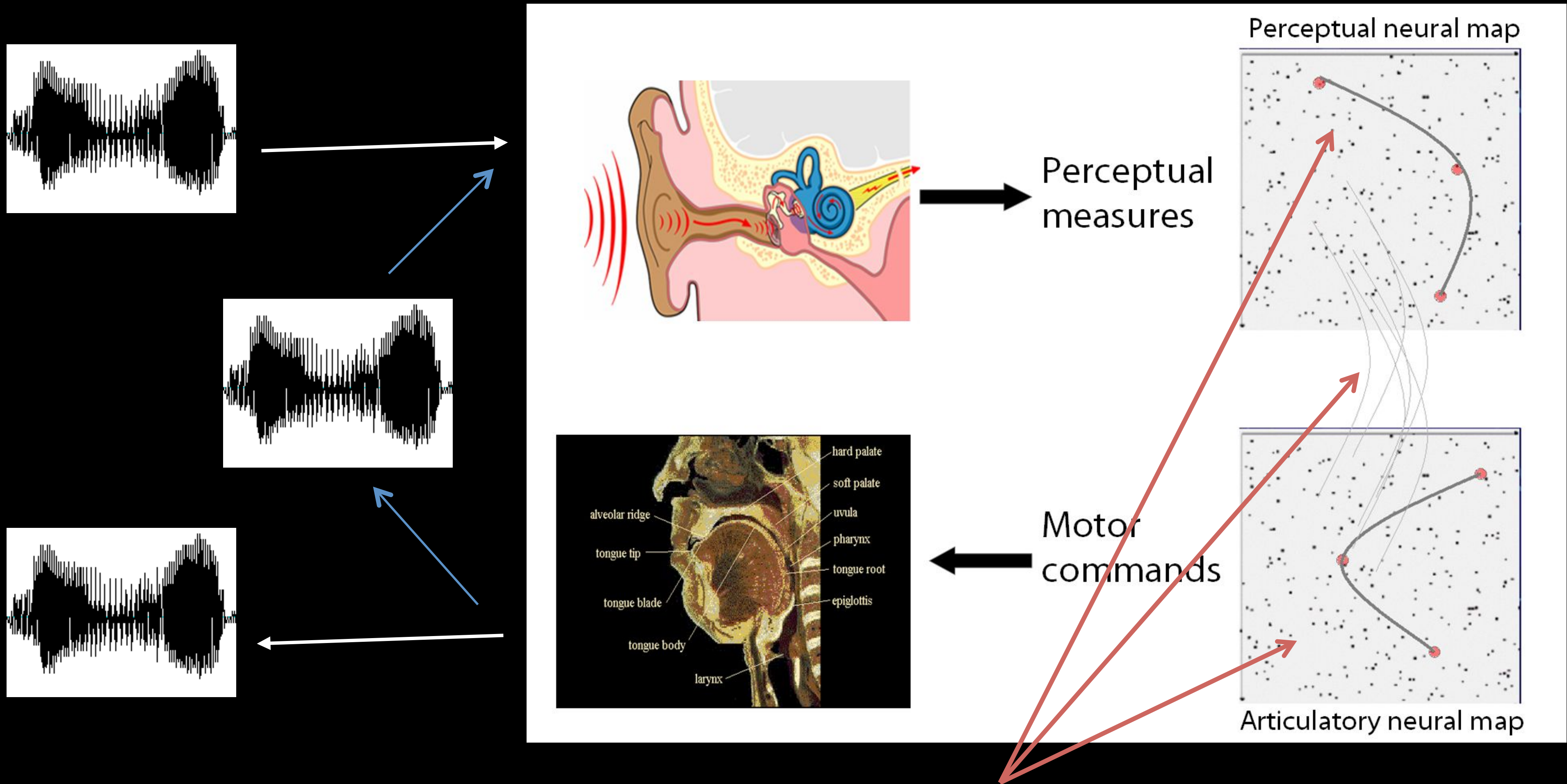
Articulatory target
(organ relation)



Dimension 1 : e.g. : tongue body position



Learning inside AND across maps: adaptive babbling

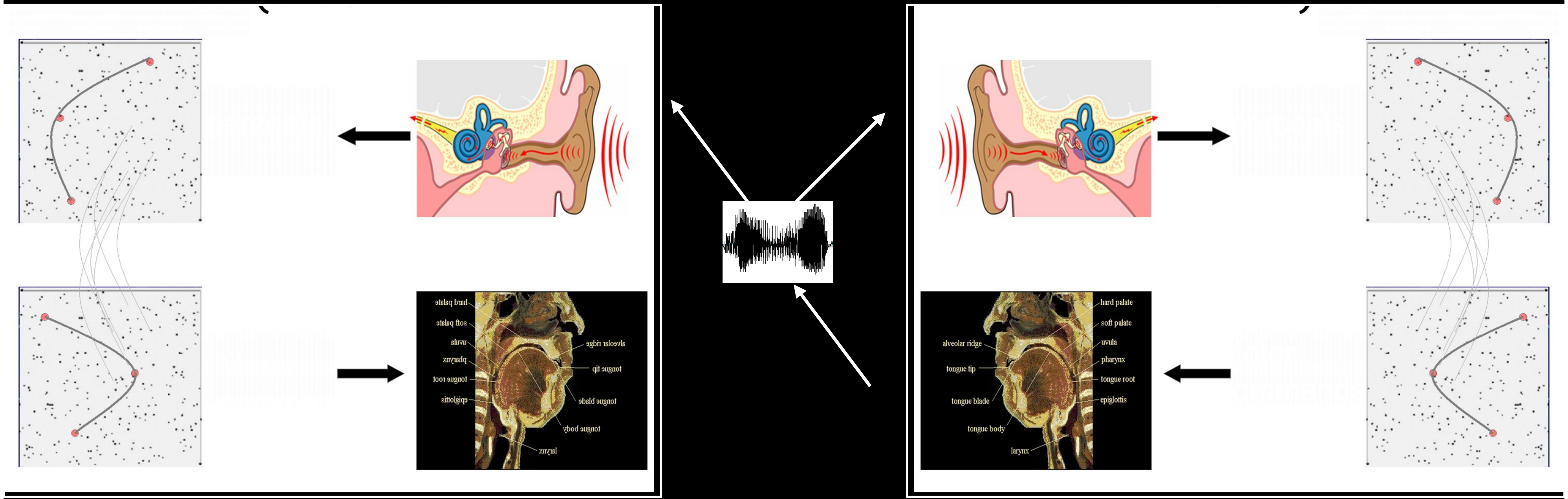


Plasticity at both levels

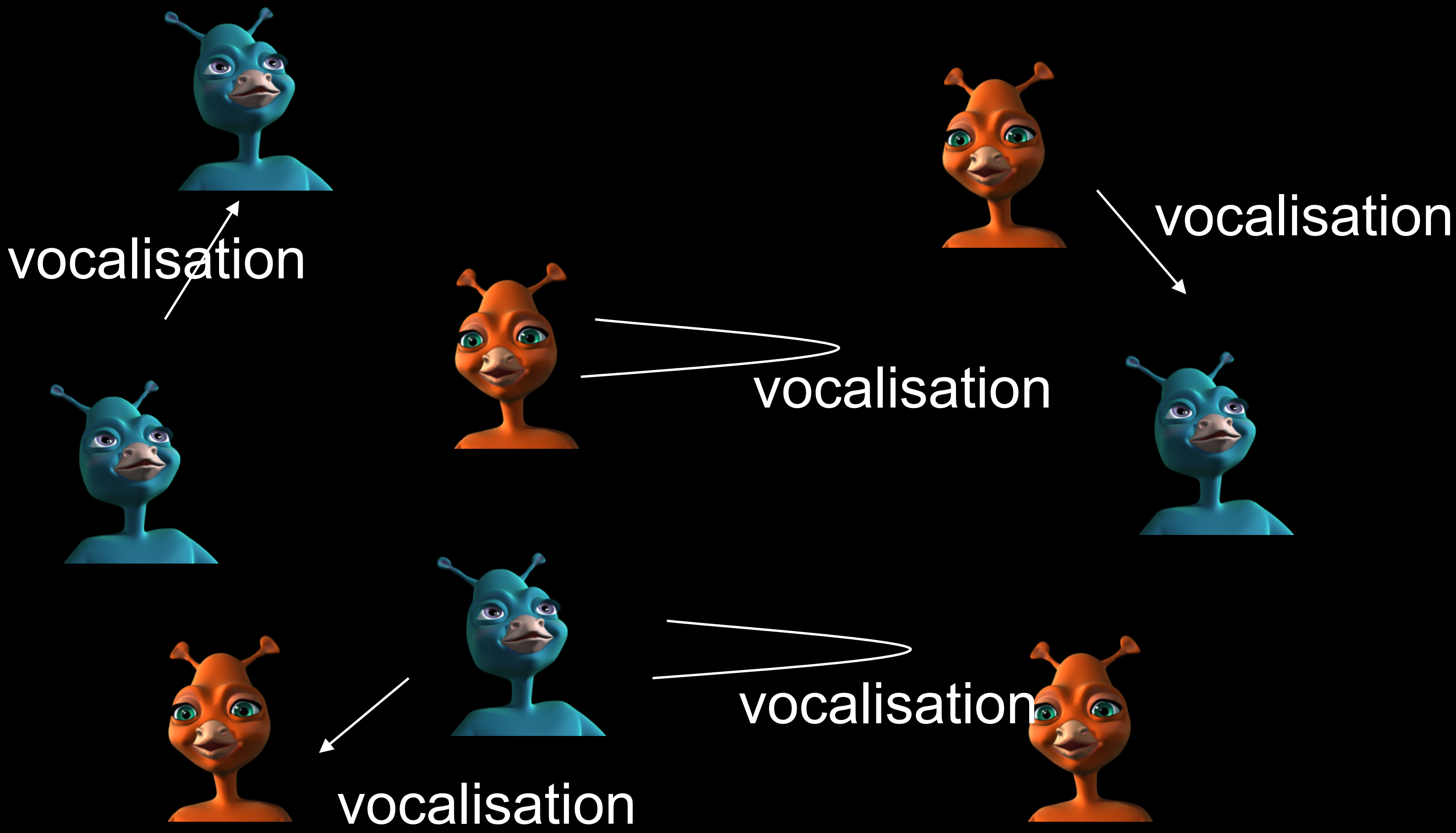
Coupling the perceptual and motor map
+
basic neural plasticity
=
the distribution of sound produced by an agent tends
to
approximate the distribution of sounds that it hears;

Also, if an agent perceives certain sound combinations
more often than others, this will favor its own
production
of these combinations;
➔Quite compatible with the phonological attunement
observed in babies (see Vihman, 1996) and adults
(Goldinger, 2000)

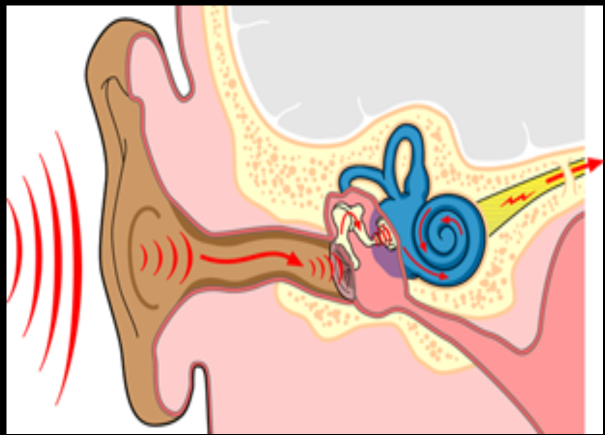
What happens if babbling agents interact together?

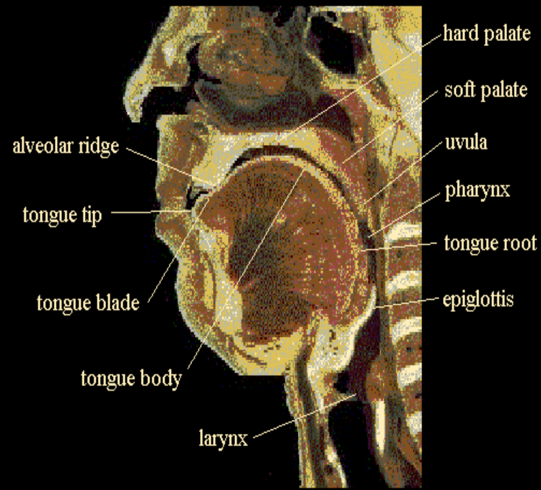


A community of babbling agents



Abstract and linear vocal tract model



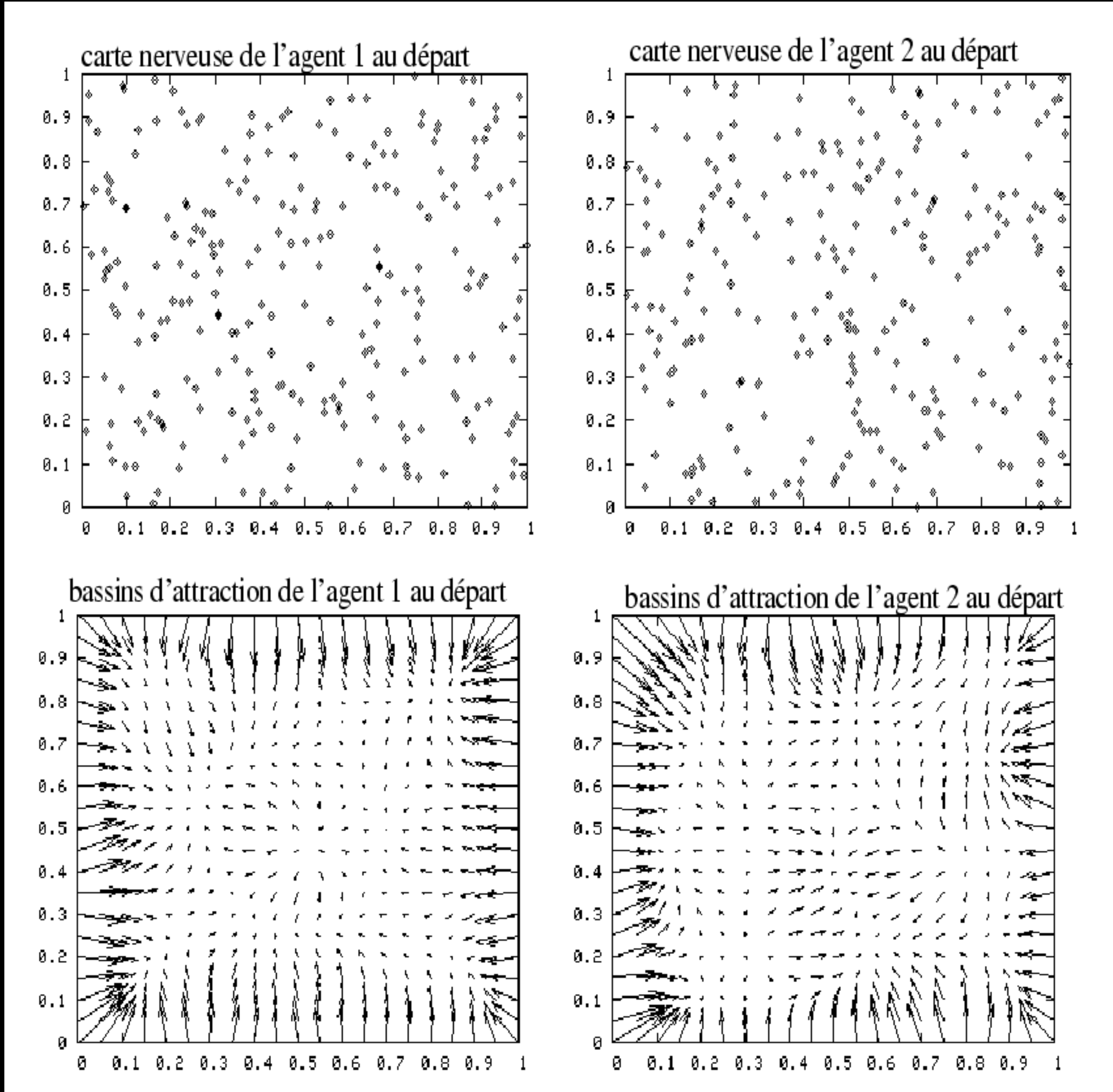


(Xp, Yp) ←→ (Xm, Ym) (2 dimensions)

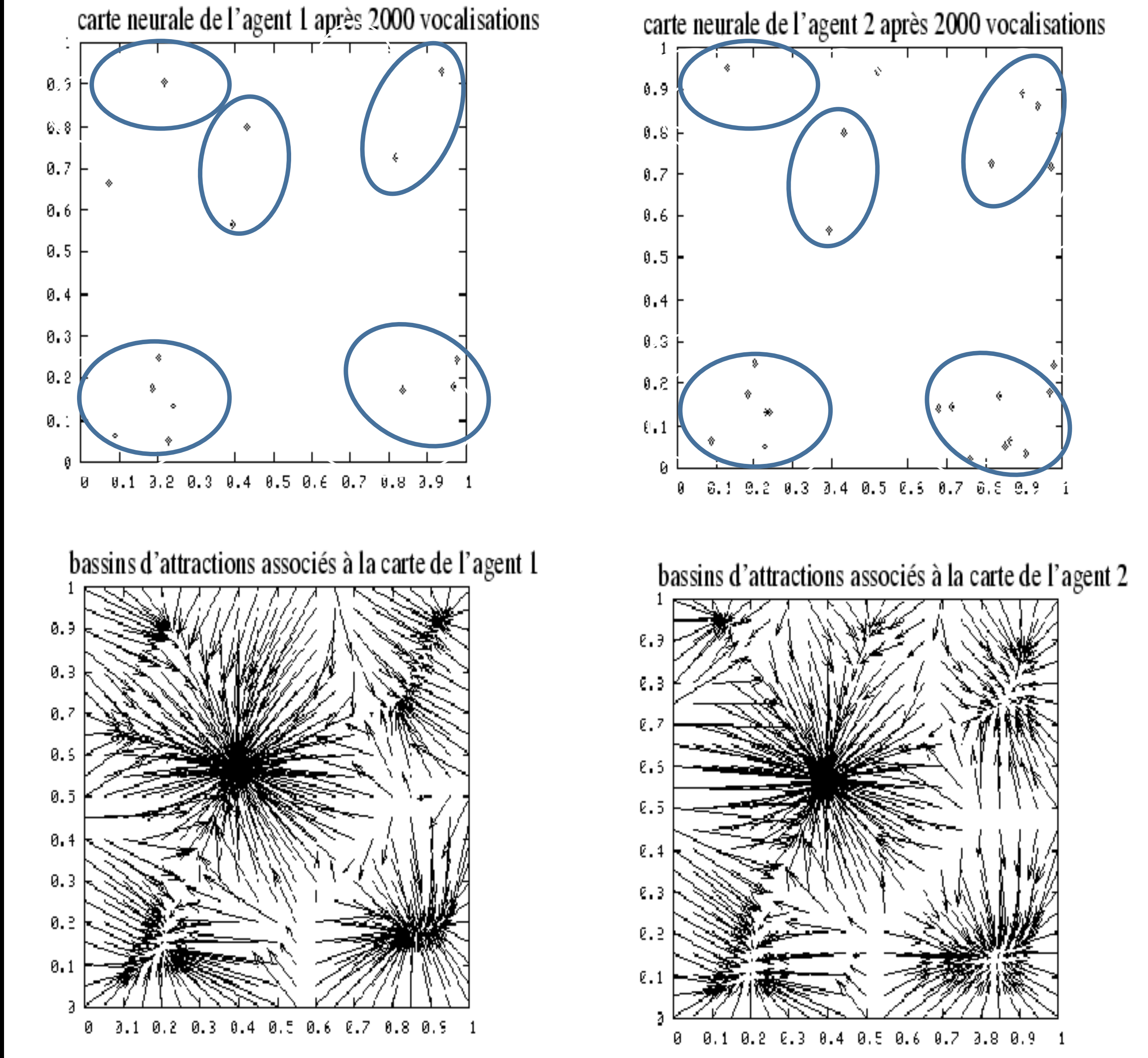
Xp = a Xm + b Ym
Yp = c Xm + d Ym

a et b random and fixed in a given simulation

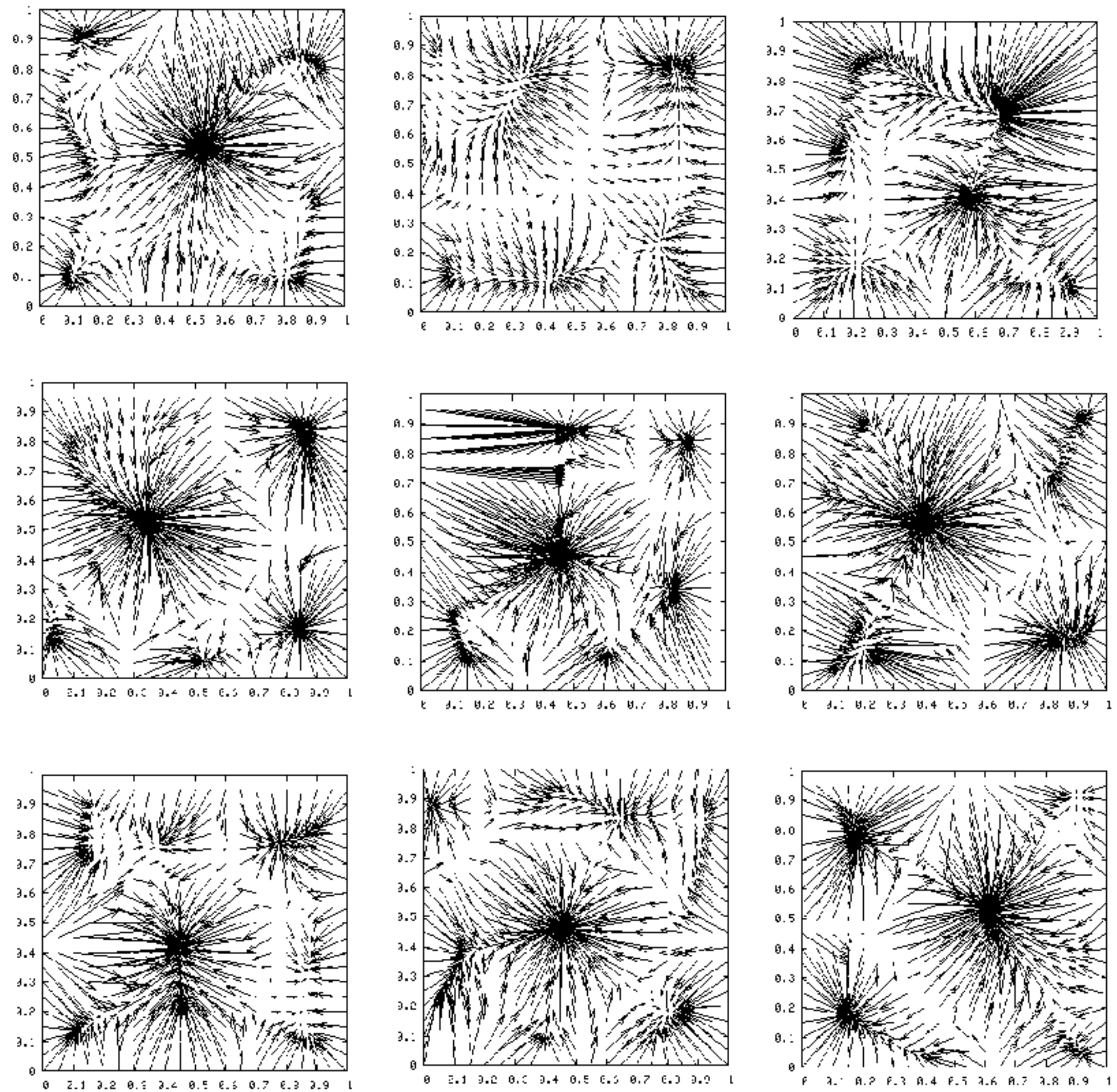
Results: initial state



Results: final state



Diversity



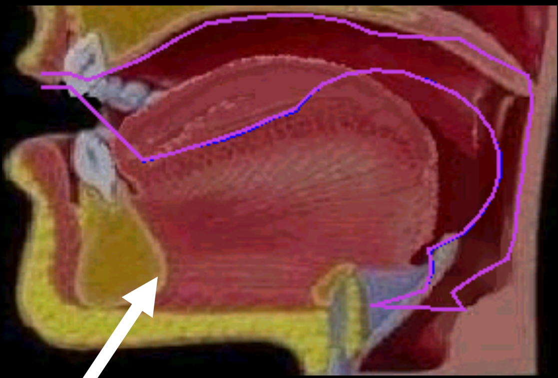
Two independant results

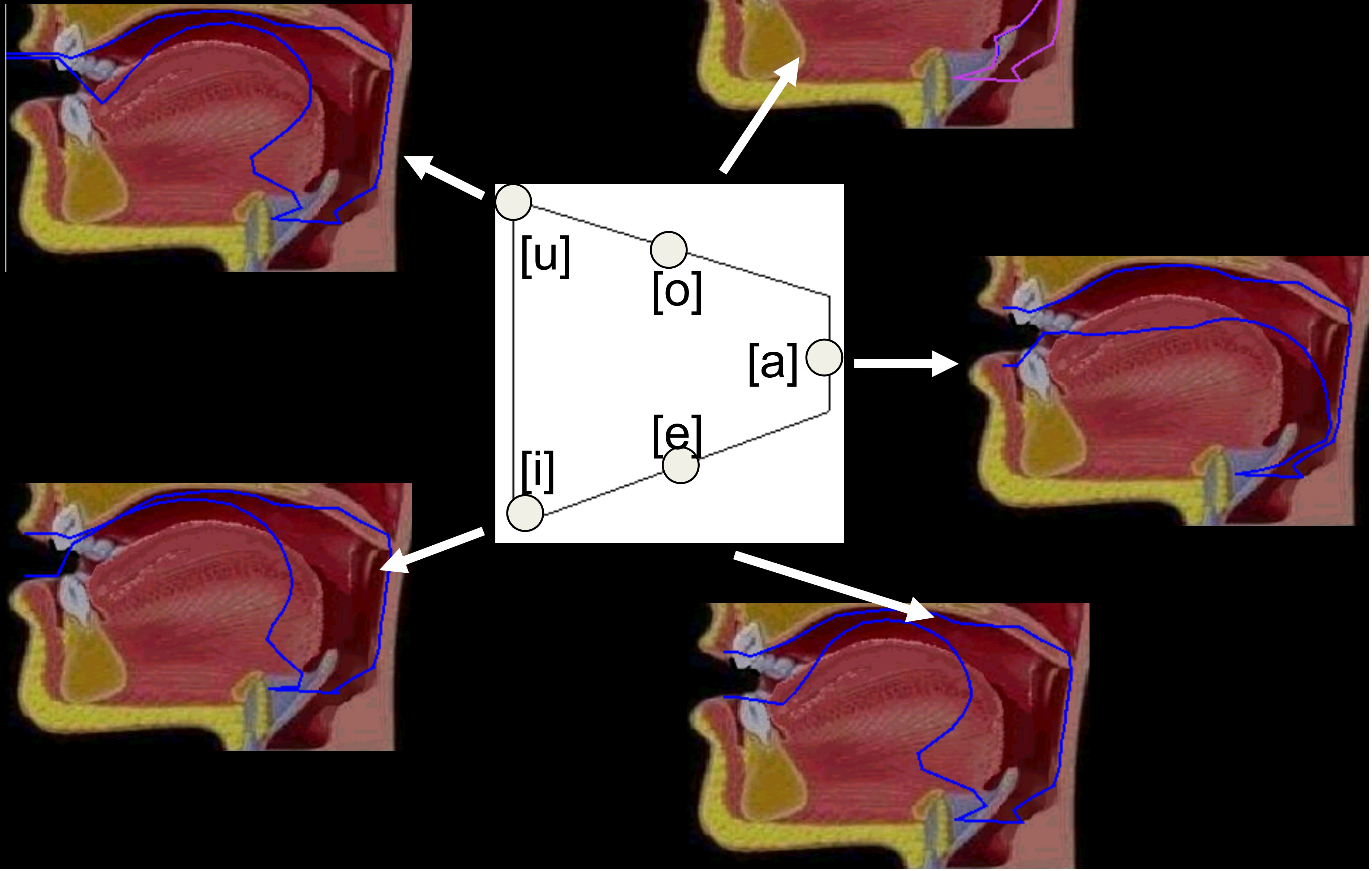
→If an agent is alone, and can only hear its own vocalizations, they still self-organize into a combinatorial system! So, combinatoriality is here a result of the internal coupling between production and perception;

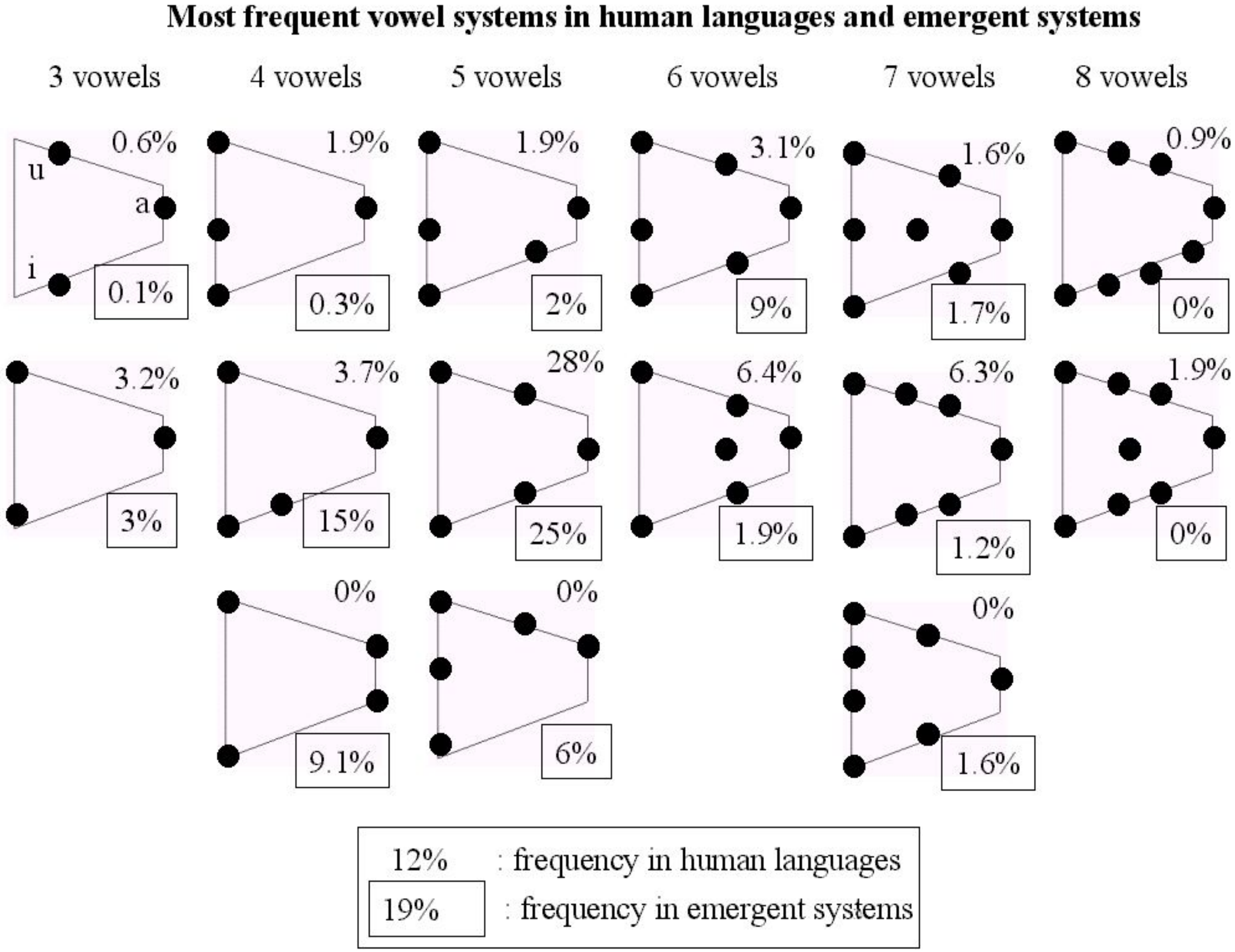
→Several agents left alone develop different combinatorial vocalization systems;

BUT If one couples agents in a shared environment, their vocalization systems spontaneously synchronize;

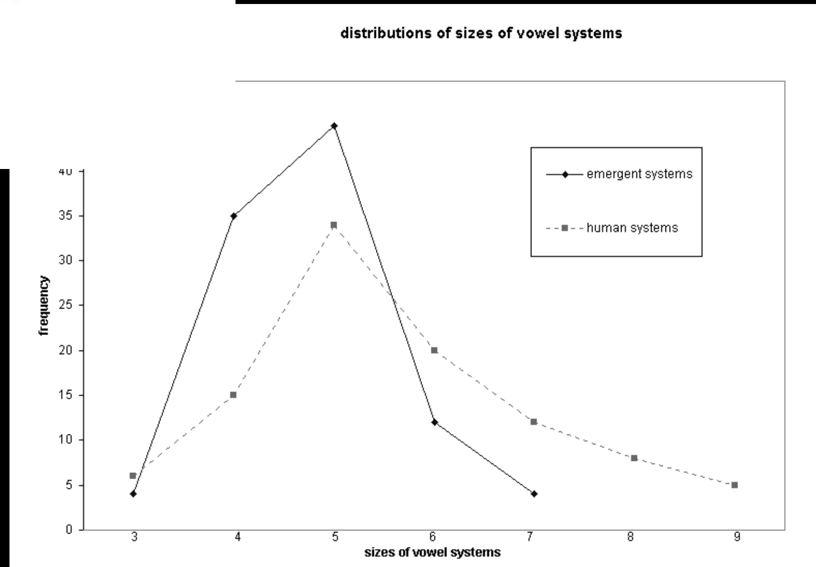
Realistic model of vowel production



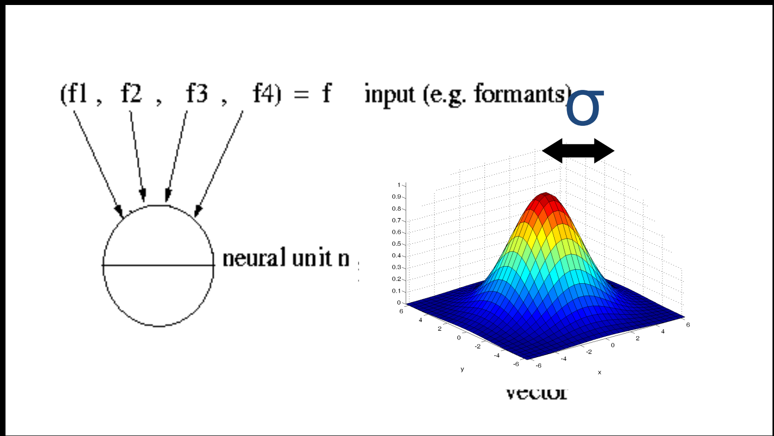




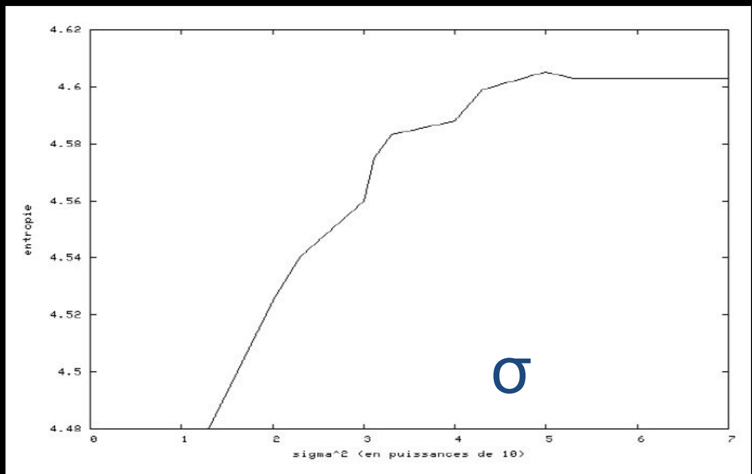
Predicts the most frequent vowel systems in human languages

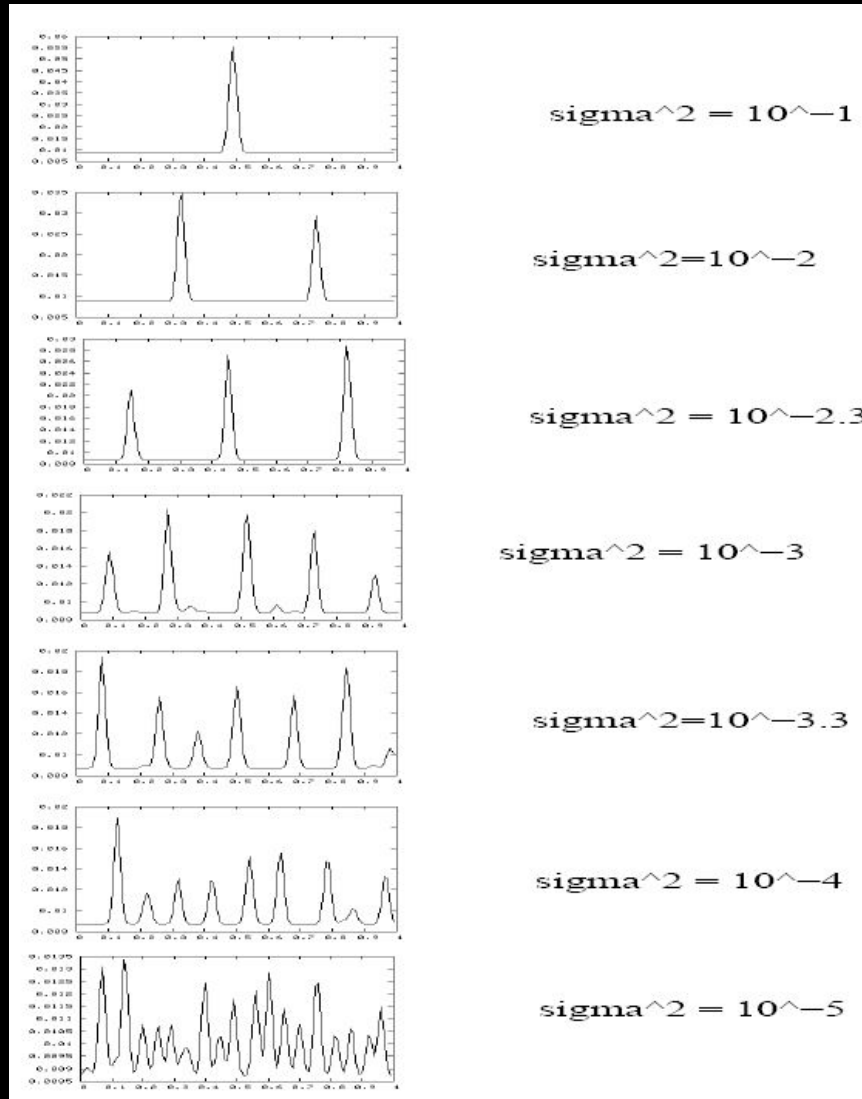


Robustness? The parameter σ



Degré de crystali-sation





Conclusions of the experiment (1)

- **Sharing:** Shows an example of decentralized mechanism that allows a population to « choose » a given speech code;

- **Universals and diversity**: Gives an elaborated insight for understanding why and how there can be phonological universals AND a wide diversity;

- **Combinatoriality:** Shows that in principle discreteness/ combinatoriality/ phonemic coding does not require non-linearities in the articulatory-acoustic-perceptual mapping (as opposed to e.g. Stevens or Mrayati, Carré and Guérin)

Conclusions of the experiment (2)

Basic neural kit for analog/holistic imitation

(There are many possible reasons for its prior evolution, e.g. no absolute necessity of a linguistic context)

small change in a small parameter space

➔

evolution + self-organization

Combinatorial and **shared** speech codes with **universals** and **diversity**

- **Evolutionary scenarios**:
  A small step for evolution, a big step for language
  ➔ opens roads to understanding language bootstrapping,

Exploration and learning of new skills during ontogeny

Developmental and social robotics: processes of extension of the repertoire of skills

➔**The central target of developmental robotics** is to understand the mechanisms that allow animals and machines to acquire novel skills (i.e. not pre-specified in the genes or by the engnieer/designer) by themselves or through interaction with humans;

➔Import, formalize, extend, implement and experiment concepts and theories of developmental and social psychology, developmental neuroscience, and linguistics into robotic model and confront them to reality;

An innate cerebral and morphological equipment …

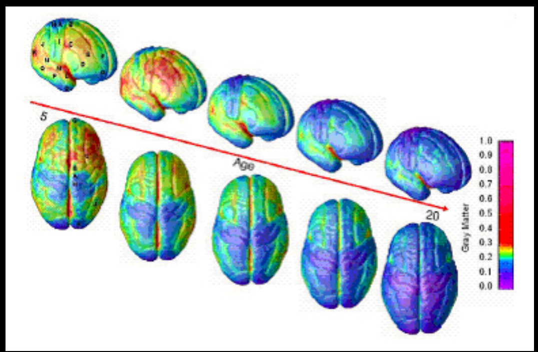
Motor primitives that constrain the space of motor commands and gestures: e.g. muscles are not controlled individually and independently, oscillators, …

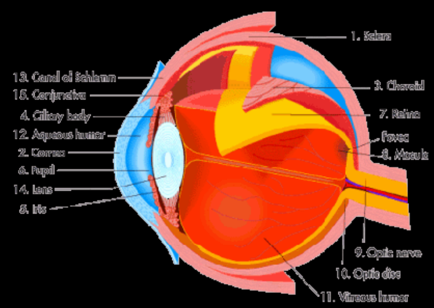
Sensori detectors and trackers that allow the baby to bootstrap its attentional and emotional systems: e.g. movement, high pitch, faces, …

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 continuously extended thanks to a generic learning and developmental system

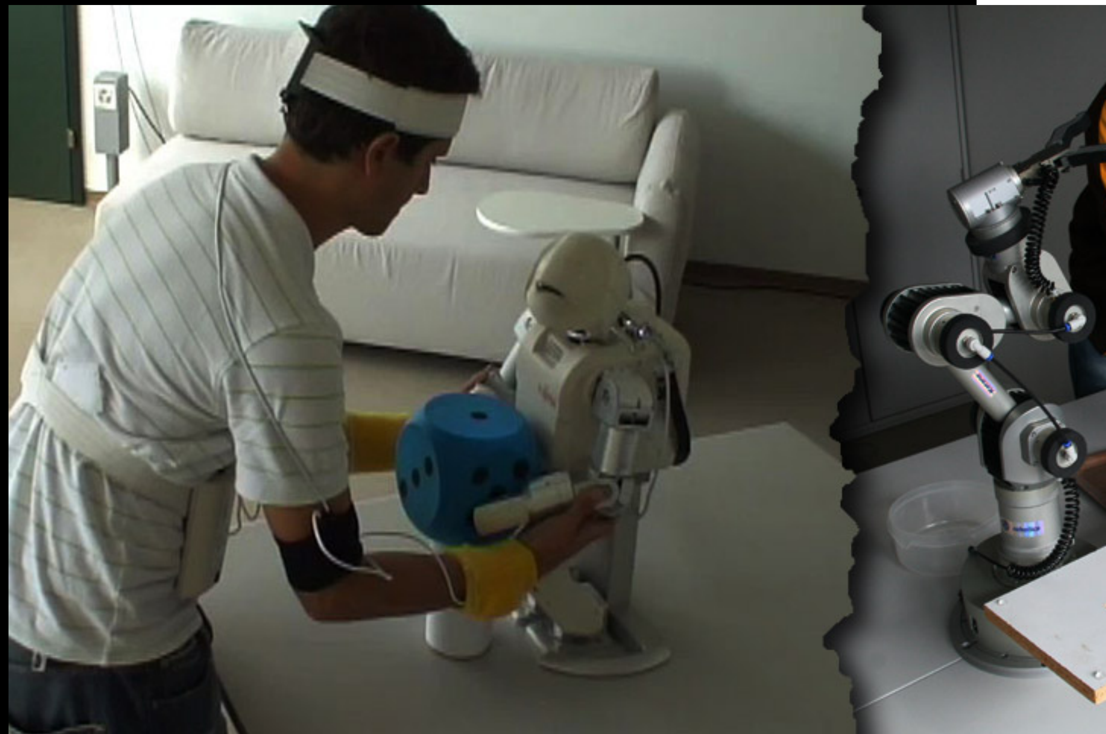
Central challenge: finding the right balance between constraints and plasticity

The sensorimotor spaces of real-world robots are typically very large, especially if one does not want to constrain them too much towards the acquisition of pre-defined specific tasks;

➔ The big problem is that in such spaces there are typically much more potential skills to be learnt than what is actually possible to learn in the life time: indeed, learning requires experimentation and/or observation, and this takes time!

➔ So we still need some constraints/biases, but constraints that are not too specific of pre-defined tasks, i.e. generic task-independent constraints or scaffoldings, which corresponds to a number of properties of the innate equipment and social embedding of human infants;

Social guidance: Learning by imitation/ observation



Supervised learning of new skills

- A (discrete or continuous) state space S (e.g. sensori state and memory of a robot)
- A (discrete or continuous) action space A (e.g. motor commands of a robot)
- A transition function $W : S(t) \times A(t) \rightarrow S(t+1)$
- A parameterized action policy $\pi_\theta : S \rightarrow A$
- Demonstrations providing supervised training data $\{(S_i, A_i)\}$
- An optimization procedure (typically a regression method) that find a policy such that

$$\theta = \arg\min_{\theta_i} \sum_i \left\| \pi_{\theta_i}(S_i) - A_i \right\|^2$$

➔ Mathematical and computational problems: how to do accurate and fast regression in high-dimensional spaces (dimensionality reduction) given potentially very noisy training data including irrelevant dimensions? How to do regression with hidden variables? How to generalize/extrapolate? How measures of similarities between states should be done? How to factorize and build abstractions? …

E.g. The Stanford Helicopter

(Abbeel et al., 2010)

Constrained learning by demonstration (combination of imitation and physical model)

Shortcomings

- Very tedious, learning not autonomous;
- No real creative solutions to problems can be found;
- Once demonstration finished, no new things are learnt;

Reinforcement learning of new skills

- A (discrete or continuous) state space S (e.g. sensori state and memory of a robot)
- A (discrete or continuous) action space A (e.g. motor commands of a robot)
- A transition function $W : S(t) \times A(t) \rightarrow S(t+1)$
- A parameterized action policy $\pi_\theta : S \rightarrow A$
- A reward/value/fitness function

$R : S \rightarrow \Re$

or

$R : S \times A \rightarrow \Re$

- An optimization procedure (e.g. model learning of W with approximate dynamic programming/stochastic optimal control to find theta) that find a policy such that

$$\theta = \arg\max_{\theta_i} \sum_{n=t+1}^{\infty} \gamma^t R(S(n))$$

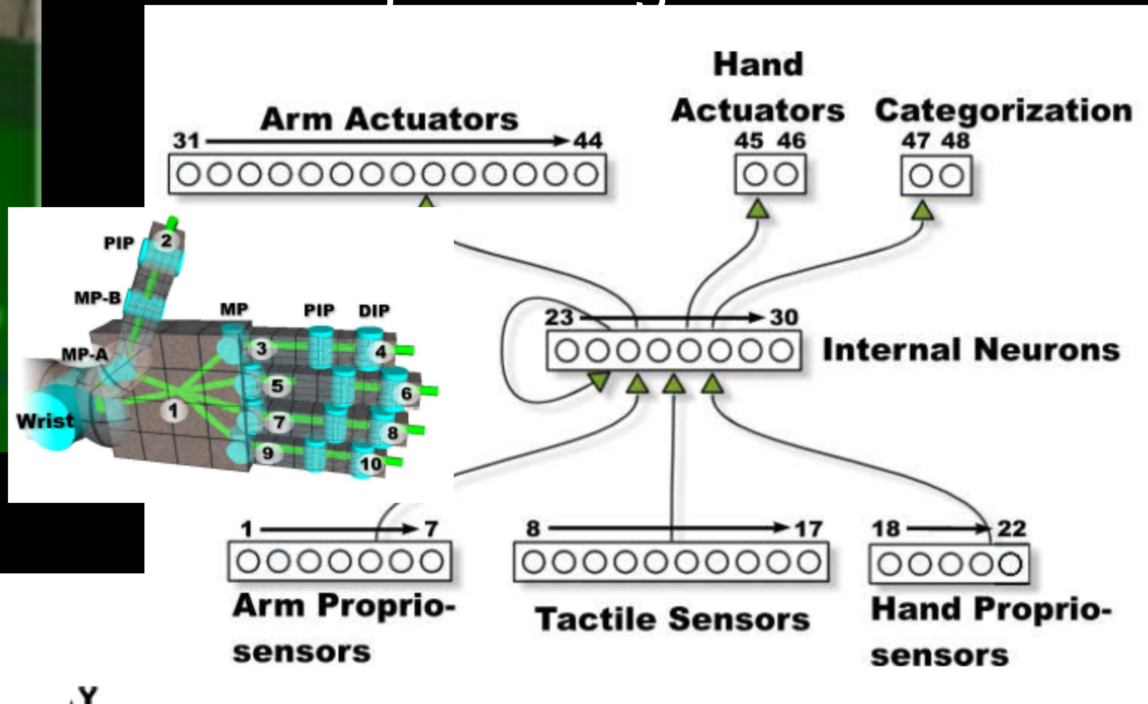
(implies exploration to learn W as well as to optimize theta, but most often random !)

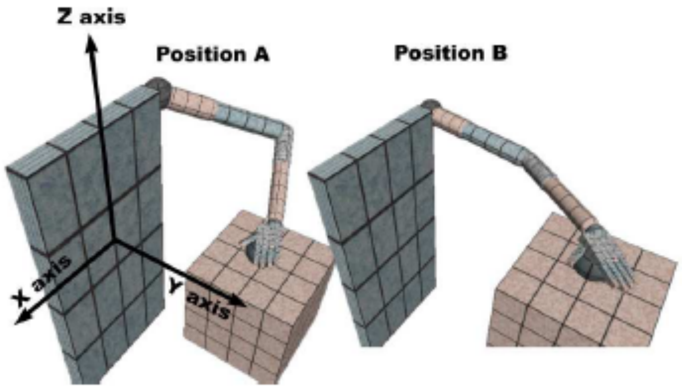
Mathematical and computational problems: same as for supervised learning + how to represent policies and approximate search of optimal sequences of actions (how to plan? How to encode policies? …)

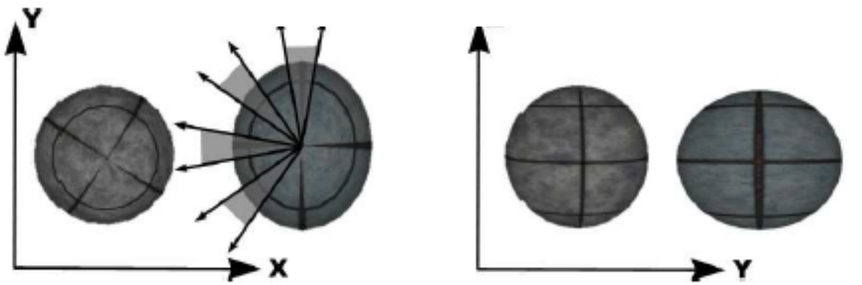
Examples

$R(S,A) =$ performance in tactile object recognition

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



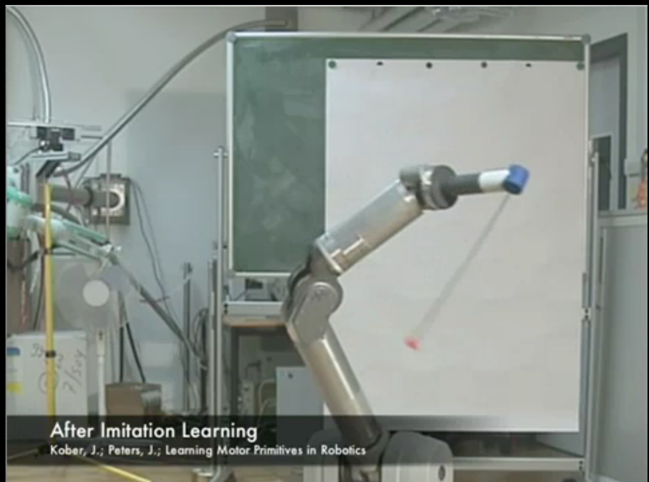




Tuci E., Massera G., Nolfi S. (2009). Active categorical perception in an evolved anthropomorphic robotic arm. IEEE International Conference on Evolutionary Computation (CEC), special session on Evolutionary Robotics.

Examples

$R(S(t), A(t)) =$ is the ball in the cup?



Kober, J.; Peters, J. (2009). Policy Search for Motor Primitives in Robotics, *Advances in Neural Information Processing Systems 22 (NIPS 2008), Cambridge, MA: MIT Press*.

Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion. Nate Kohl and Peter Stone.
In Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2619–2624, May 2004.

Shortcomings

• Most work so far in rather small sensorimotor spaces because algorithms inefficient in high-dimensional spaces → exploration has mostly been left a vastly underexplored topic here also (i.e. elaborated techniques try to avoid exploration as much as possible);

•Less tedious, but requires careful design of the reward function + a new design for each novel skill ➔ a lot of work for the engineer for each novel skill !

• Once a skill is mastered, no more skills are learnt if engineer does nothing;

Human infants do not only learn/ discover new skills by observation/ imitation or task-specific reinforcement learning

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

→ **INTRINSIC MOTIVATION**

Intrinsic motivation

Hull (1943), White (1959): Basic forms of motivations (e.g. motivation for food and water, for sex, motivation for the maintainance of physical integrity, search for social bonding) can not account for the whole diversity of spontaneous exploratory behaviours of humans.

➔ Search for novelty, surprise, challenge, incongruences, …

What makes intrinsically motivating activities/situations motivating?

What are the features of interestingness?

Why are some activities fun to practice (alone) ?

Drive for novelty ?

(Hull, 1943; Montgomery, 1954) proposed a drive for novelty: the experience of novel situations is rewarding.

Reduction of cognitive dissonances ?

Festinger's theory of cognitive dissonance (Festinger, 1957) asserted that organisms are motivated to reduce dissonance, which is the incompatibility between internal cognitive structures and the situations currently perceived.

(Kagan, 1972) a primary motivation for humans is the reduction of uncertainty in the sense of the ”incompatibility between (two or more) cognitive structures, between cognitive structure and experience, or between structures and behavior.

However, these theories were criticized on the basis that much human behavior is also intended to increase uncertainty, and not only to reduce it.

Optimal incongruity ?

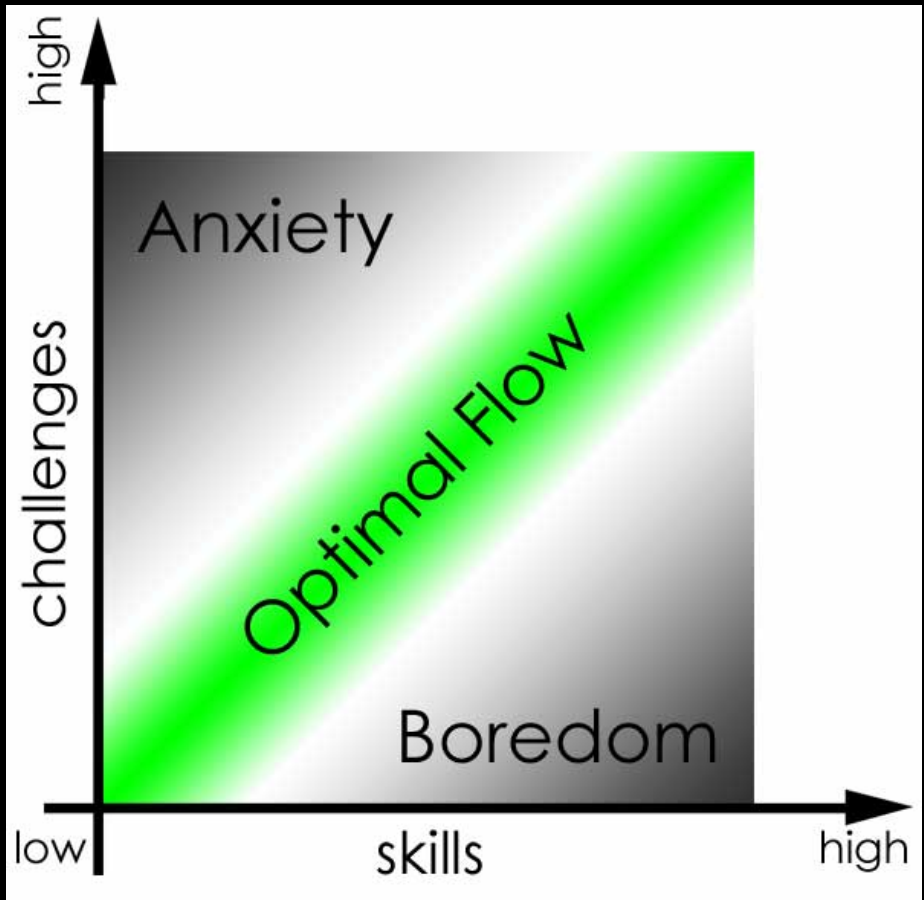
(Hunt, 1965) children and adult look for optimal incongruity.

(Berlyne, 1960) developed similar notions as he observed that the most rewarding situations were those with an intermediate level of novelty, between already familiar and completely new situations.

Effectance and personal causation ?

(White, 1959; De Charms, 1968) activities that we master and events that are caused by our own action are rewarding. The higher the degree of control, the higher the interest.

Optimal challenge ?

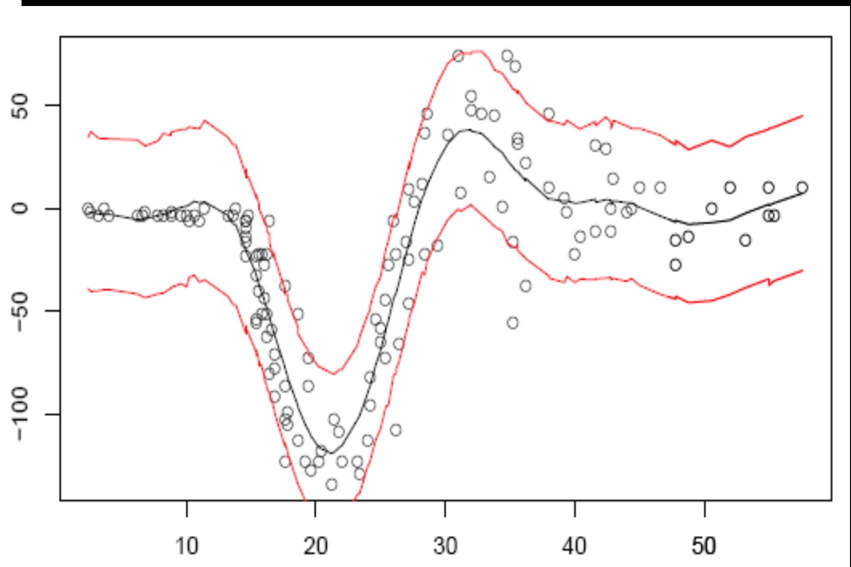


(Csikszentmihalyi, 1991) activities/goals that are not already mastered but within reach, i.e. of intermediate level of difficulty, are rewarding.

➔ Theory of « Flow ».

Intrinsically motivated reinforcement learning

Y



X

f

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

➔ 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)
- **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”

• 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 data is maximal;

• Places where the entropy of data is maximal;
…

• **in RL:** Counter-based, recency-based, novelty-based, « exploration bonuses »
(Sutton, 1990; Brafman and M. Tennenholtz, 2002; Strehl et Littman, 2006; Szita and Lorincz, 2008, …)

These measures are inoperant in real-world sensorimotor spaces

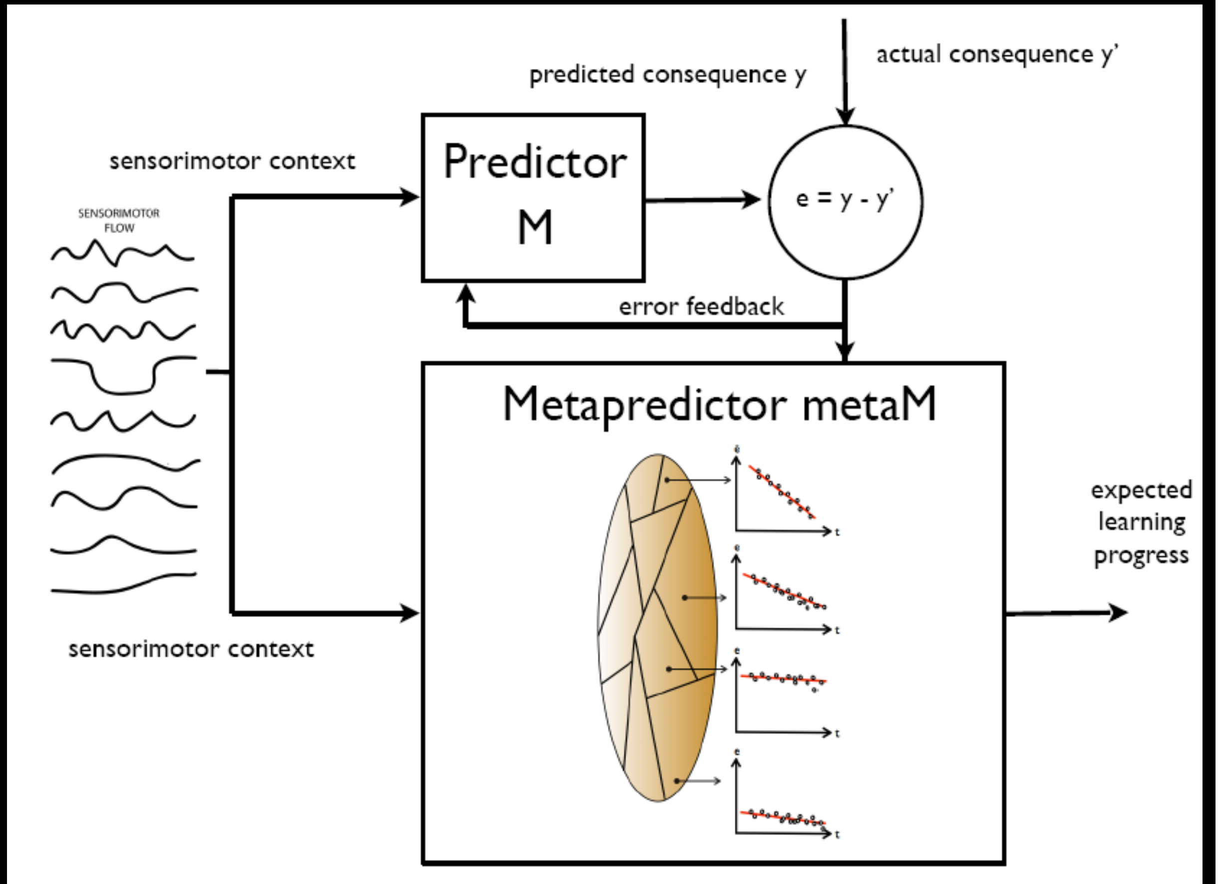
➔ interestingness as optimal intermediate complexity

➔ How to model and implement this?

Active regulation of the growth of complexity in exploration

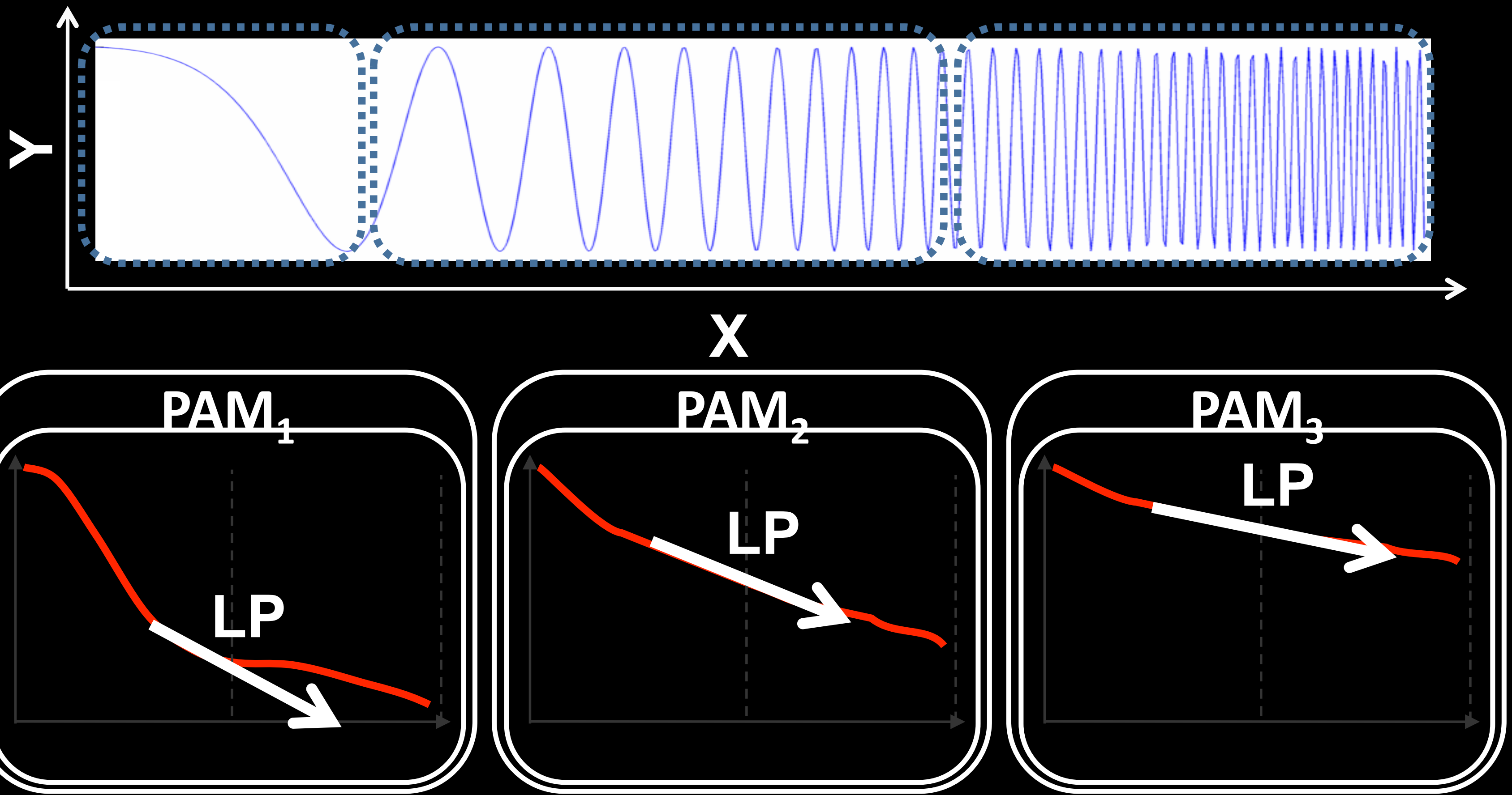
Optimizing learning ***progress***, i.e. the decrease of prediction errors (derivative)

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



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

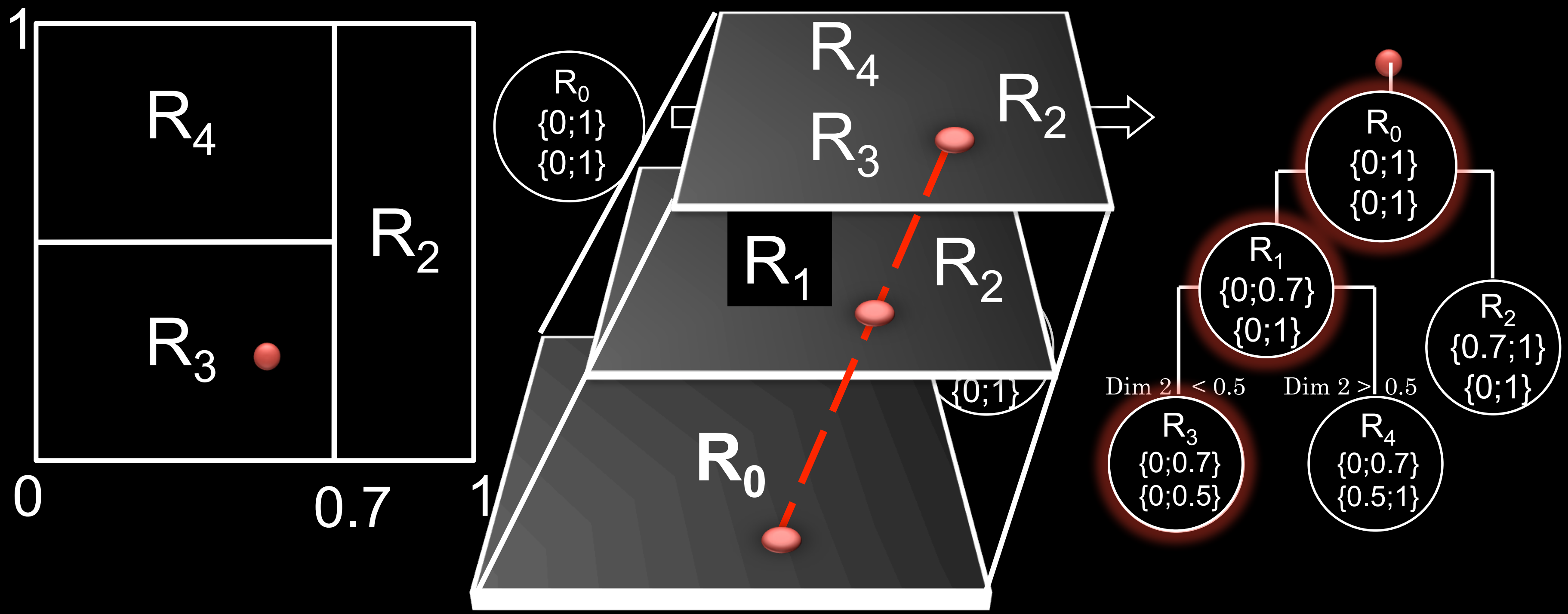
R-IAC: multi-resolution probabilistic region-based learning progress



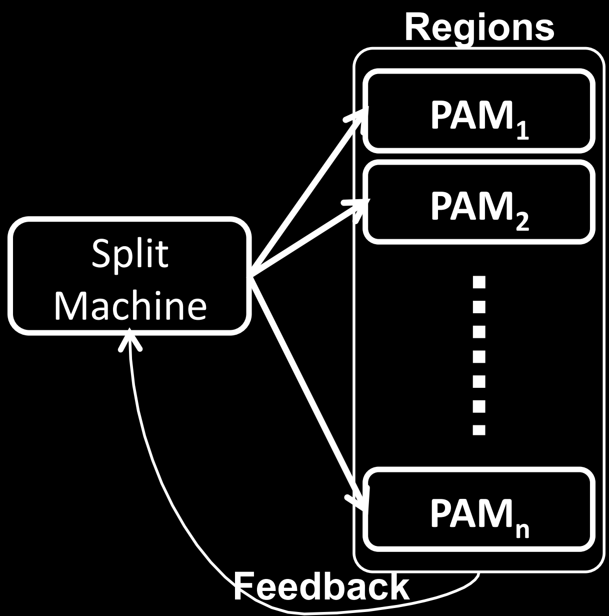
Learning Progress = decrease of mean prediction errors in a region

(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

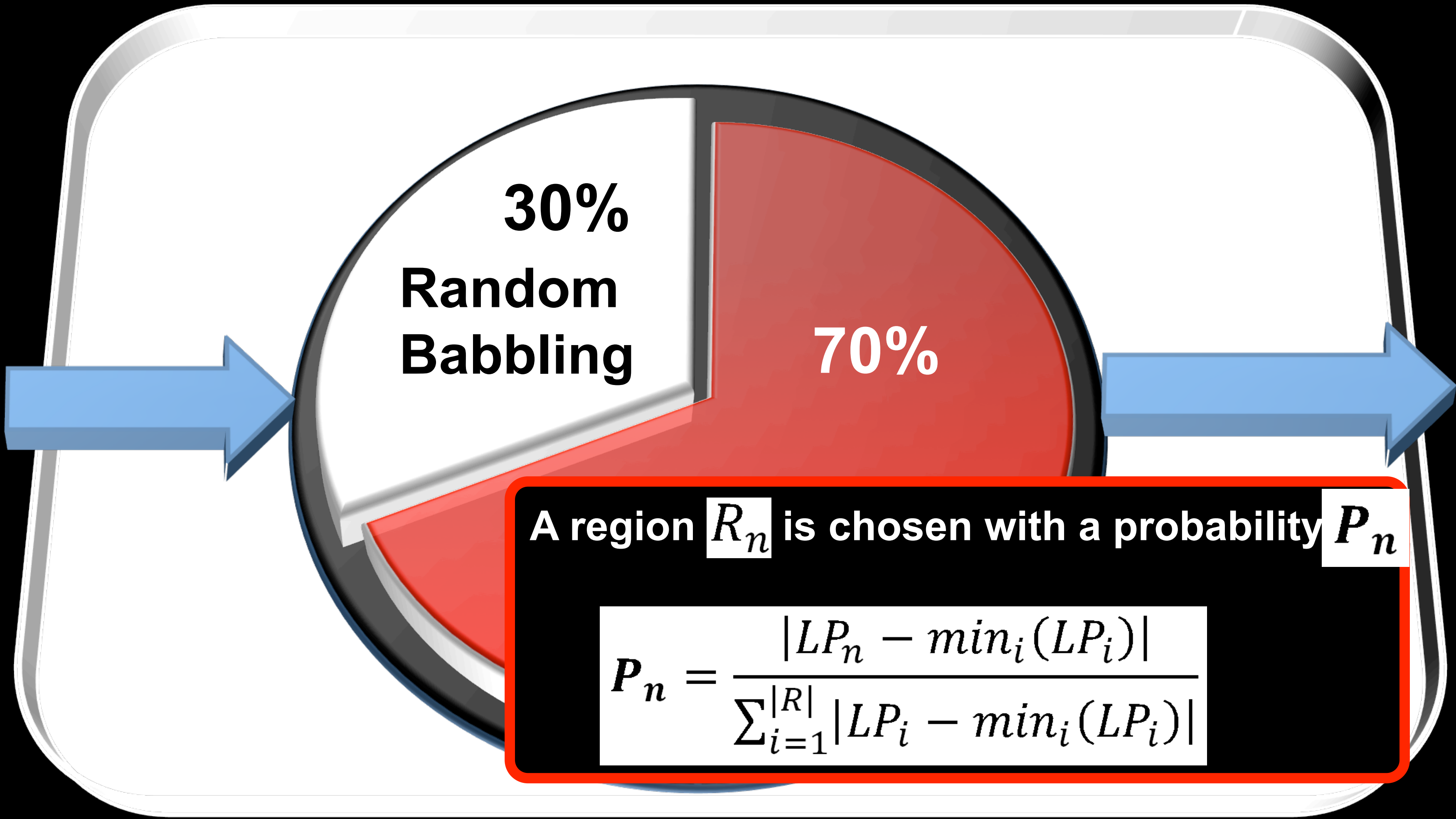
j cutting dimension

v_j associated cutting value

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

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

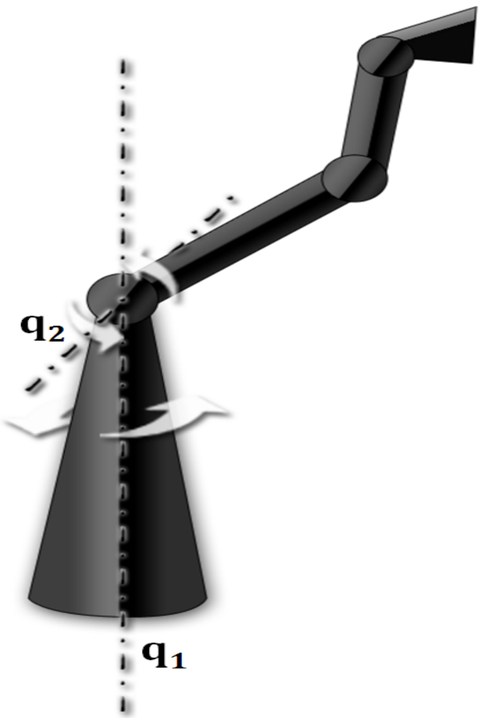
R-IAC: multi-mode probabilistic experiment selection



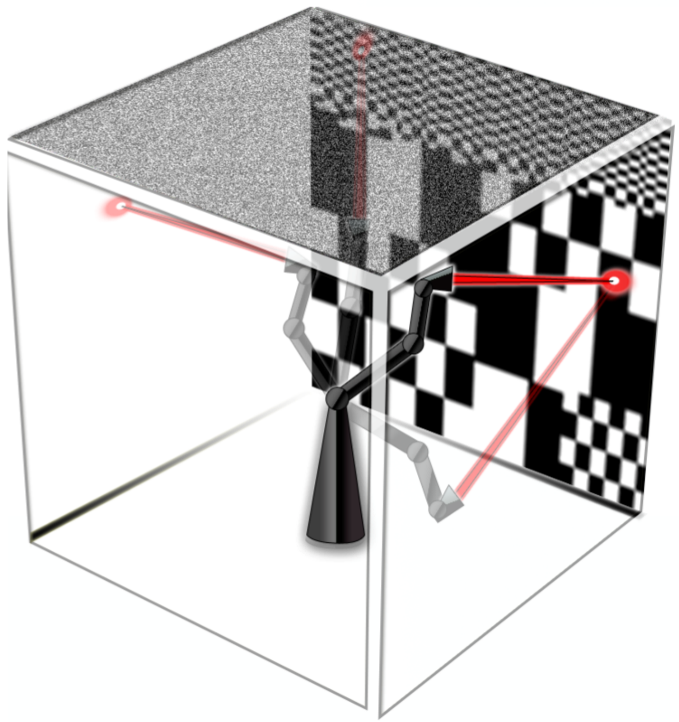
A region R_n is chosen with a probability $\boldsymbol{P_n}$

$$\boldsymbol{P_n} = \frac{|LP_n - min_i(LP_i)|}{\sum_{i=1}^{|R|} |LP_i - min_i(LP_i)|}$$

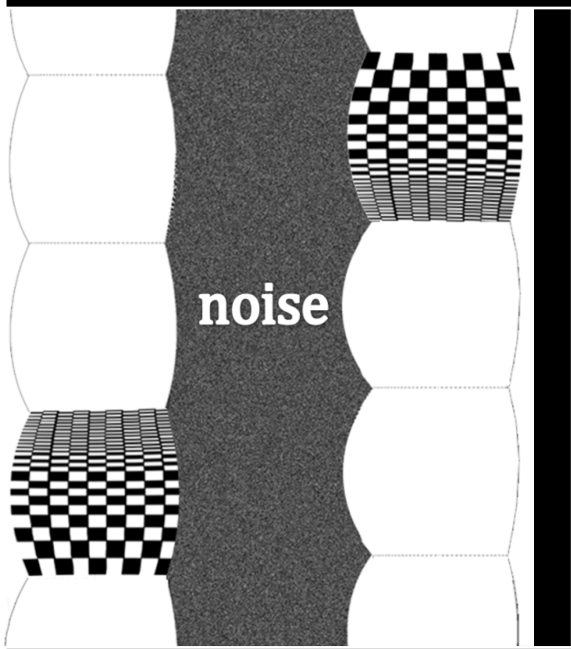
Example in a (not so) simple experiment



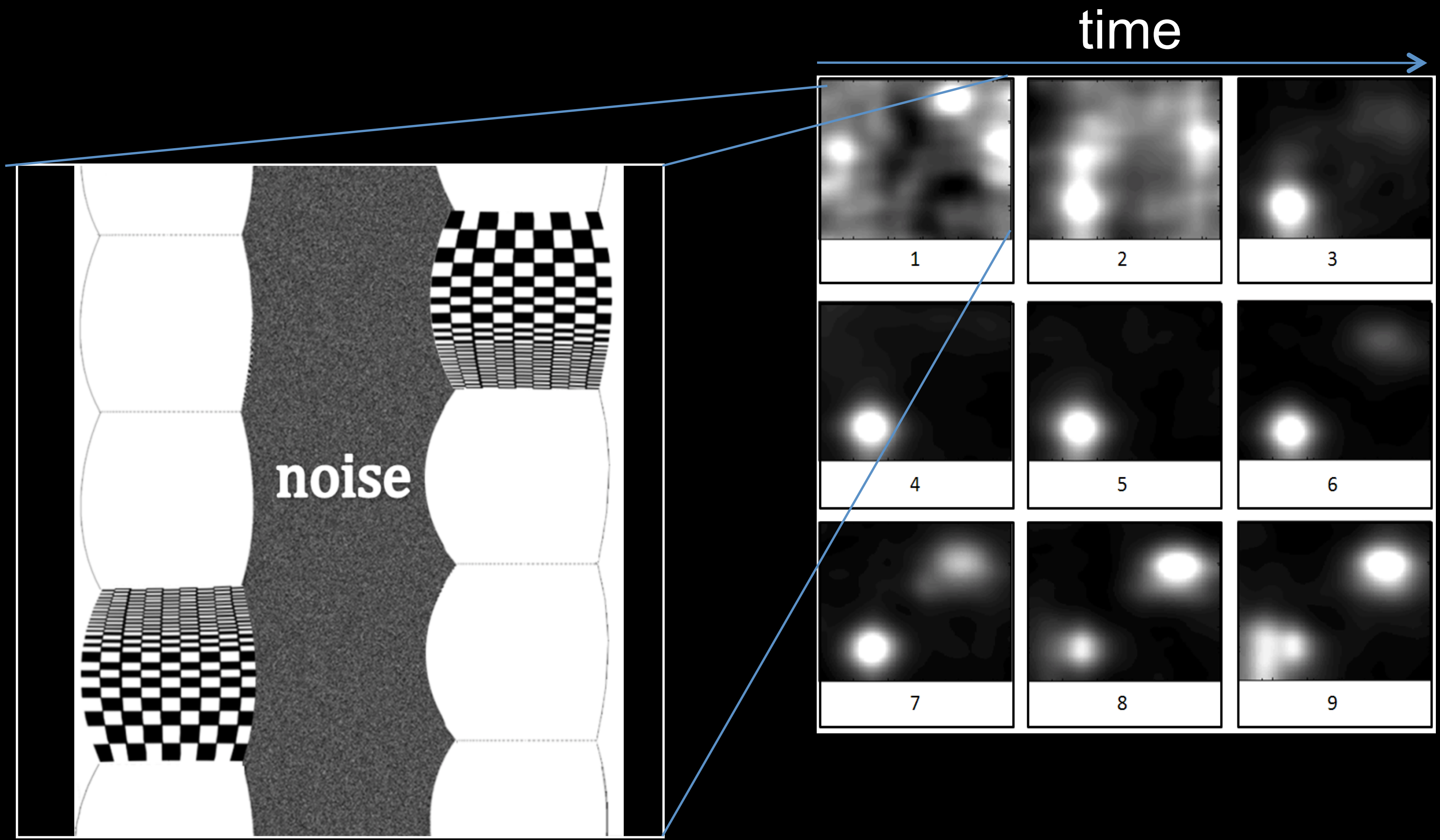
2 DOF redundant robotic arm, with a 1-pixel camera



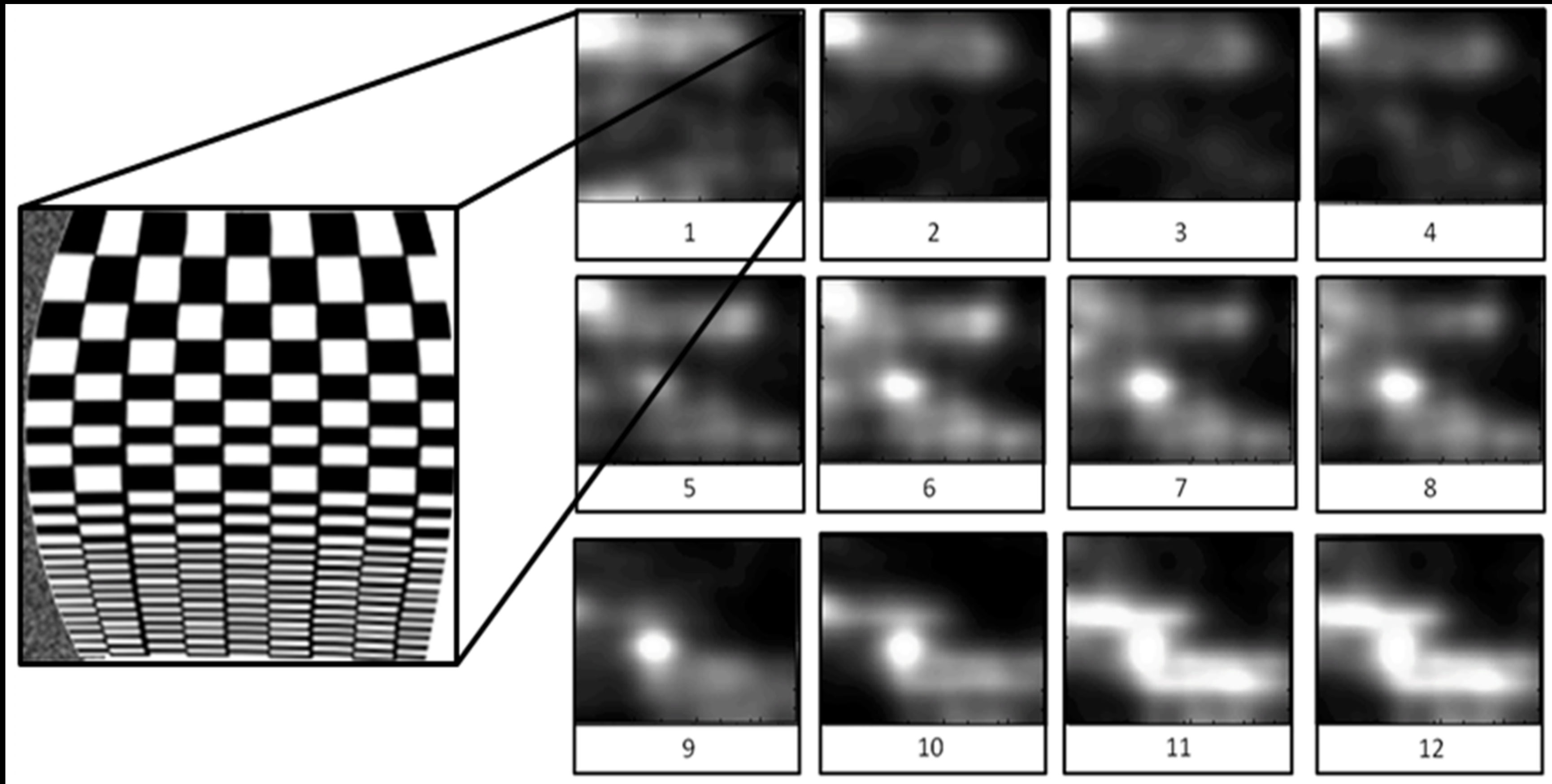
An inhomogeneous space to be explored



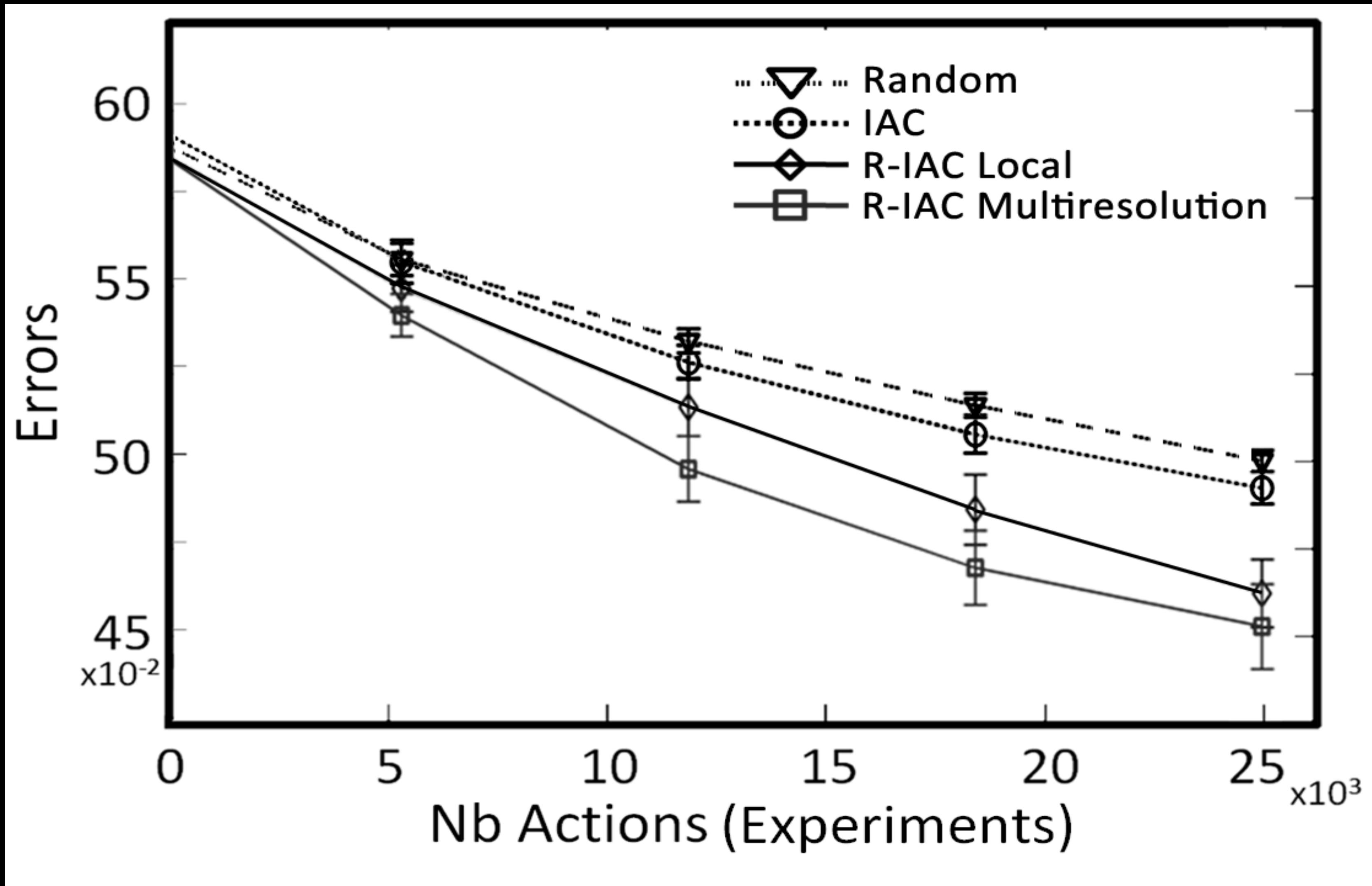
Visualization of the mapping to be learnt



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;

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;***

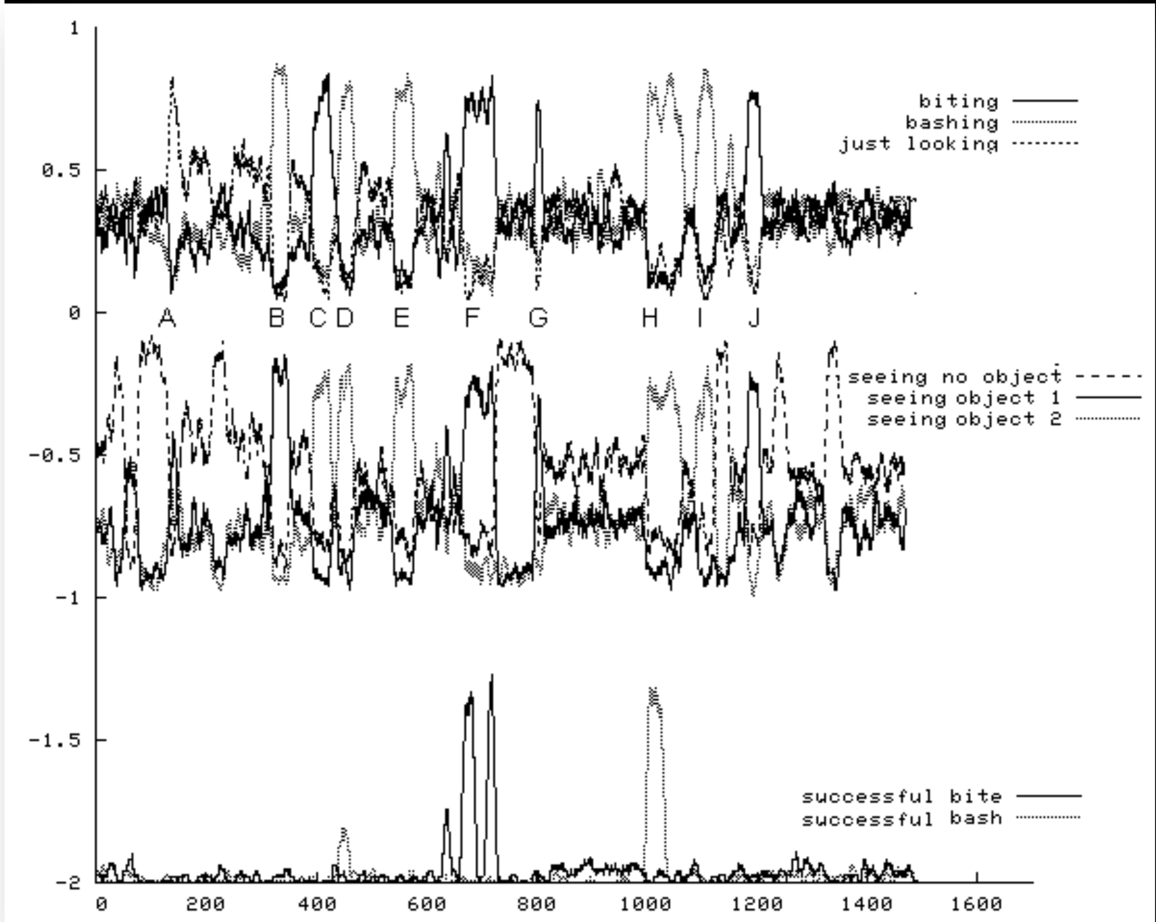
These primitives are often conceived as parameterized dynamical systems;

e.g. CPG, oscillators

The Playground experiment

http://playground.csl.sony.fr

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

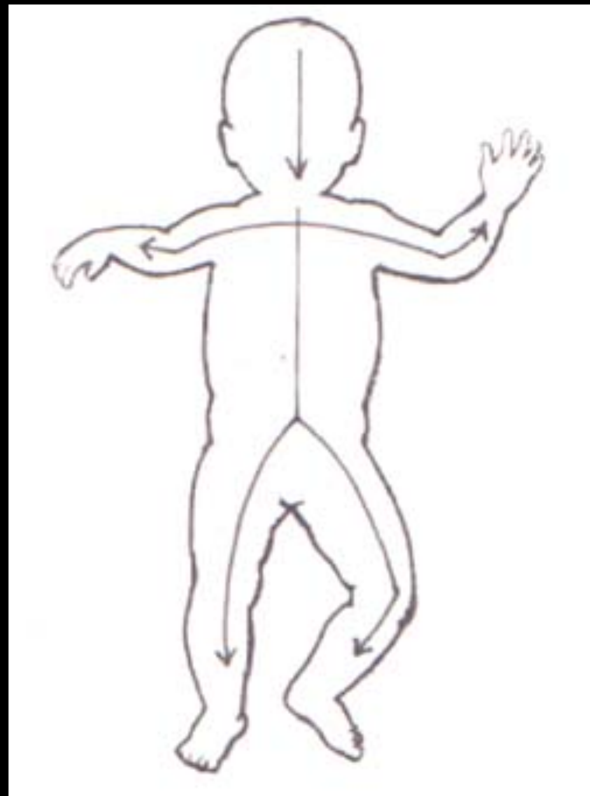


Self-organization of developmental patterns

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

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

Sensori resolution for state estimation

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

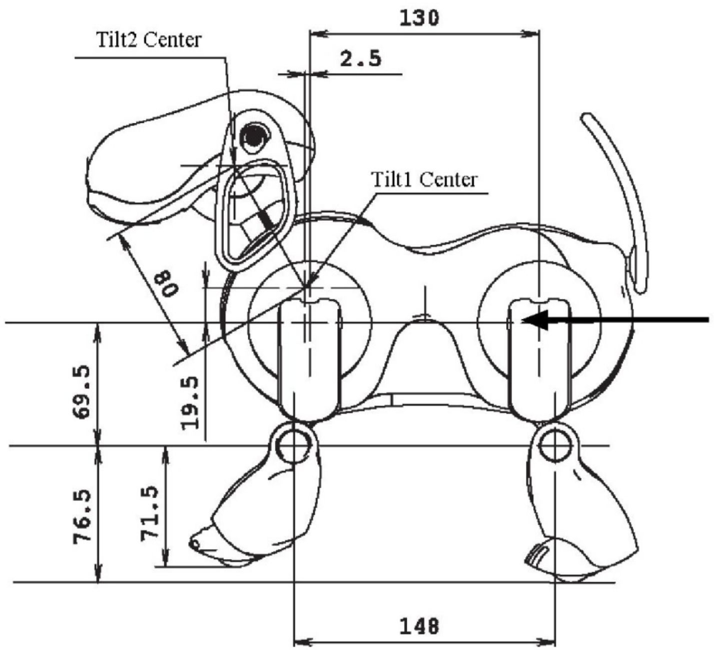
Volume/range of explorable values in motor channels, with proximo-distal law

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

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.*

2nd experiment: developmental learning of locomotion

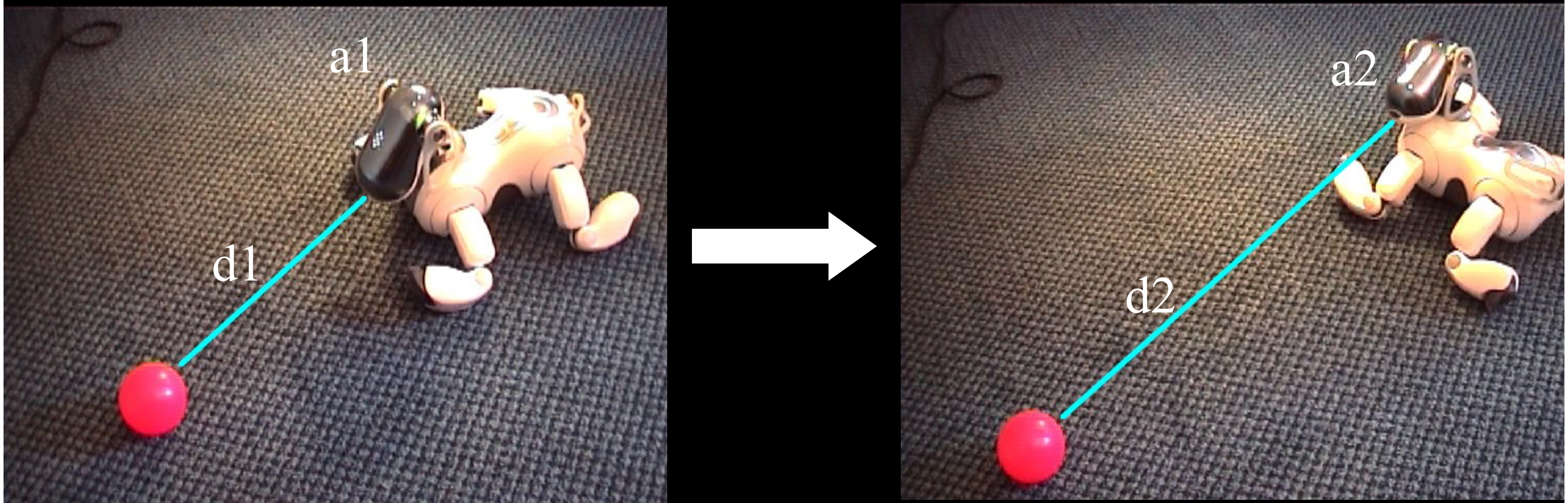


$j(t) = m + a.\sin(\omega.t + \phi)$

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 action repertoire to reach any particular location in its field of view

Developmental constraints on exploration: 3) Morphological computation

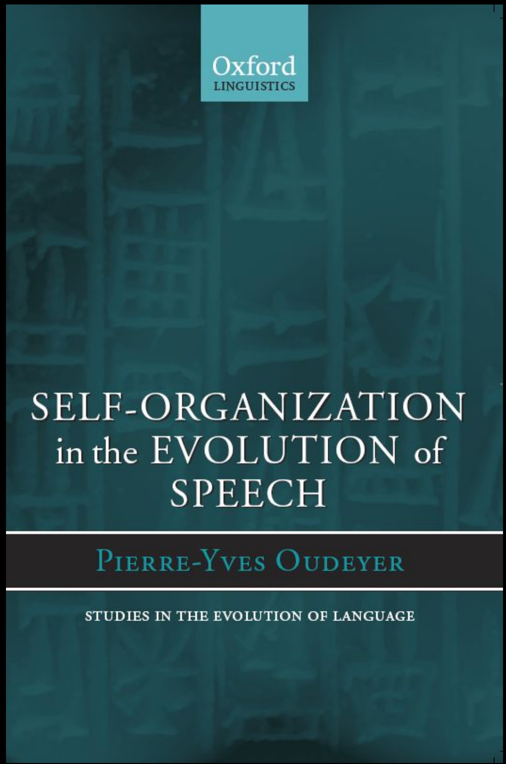
The example of passive dynamic walkers [Video]
Tad McGeer (McGeer, 1990)

Acroban and semi-passive morphological computation

• **Acroban:** the **first French humanoid robot, with a vertebral column**, dynamically equilibrated with advanced motor primitives and with 32 degrees of freedom and the possibility to interact physically softly with the robot (Olivier Ly)

➔ Versatility of the dynamical system (morphology + motor primitives): e.g. driving through physical HRI without any (specific) reprogramming !

Videos available at http://flowers.inria.fr/acroban.php



Thank you!

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

INRIA Flowers

References (1)

Waddington, C. H. (1956). *Principles of Embryology. London : George Allen & Unwin.*

Thompson, D W., 1992. *On Growth and Form. Dover reprint of 1942 2nd ed. (1st ed., 1917).*

McGeer, Tad. Passive Dynamic Walking. *Int. J. Robotics Research. 1990, Vol. 9, 2.*

Turing, A., The Chemical Basis of Morphogenesis, Phil. Trans. R. Soc. London B 237 pp 37-72 (1952)

Maynard Smith, J. (1968) *Mathematical Ideas in Biology. Cambridge University Press.*

Camazine, Scott, Jean-Louis Deneubourg, Nigel R. Franks, James Sneyd, Guy Theraulaz and Eric Bonabeau (2001). Self-Organization in Biological Systems. Princeton, New Jersey: Princeton University Press.

Schwartz, J., Bo, L., Valle, N., Abry, C., 1997. Major trends in vowel systems inventories. J. Phonetics 25, 255–286.

Liljencrants, J. & Lindblom, B.: Numerical simulation of vowel quality systems: The role of perceptual contrast. Language 48: 839-862 (1972).

Oudeyer, P-Y. (2006) Self-Organization in the Evolution of Speech, Studies in the Evolution of Language, Oxford University Press. (Translation by James R. Hurford)

Oudeyer, P-Y. (2005) The Self-Organization of Speech Sounds, Journal of Theoretical Biology, 233(3), pp. 435--449.

Autonomous Helicopter Aerobatics through Apprenticeship Learning Pieter Abbeel, Adam Coates and Andrew Y. Ng. Accepted for publication in the International Journal of Robotics Research (IJRR), 2010.

Tuci E., Massera G., Nolfi S. (2009). Active categorical perception in an evolved anthropomorphic robotic arm. IEEE International Conference on Evolutionary Computation (CEC), special session on Evolutionary Robotics.

Policy Gradient Reinforcement Learning for Fast Quadrupedal Locomotion. Nate Kohl and Peter Stone. In Proceedings of the IEEE International Conference on Robotics and Automation, pp. 2619–2624, May 2004.

Kober, J.; Peters, J. (2009). Policy Search for Motor Primitives in Robotics, *Advances in Neural Information Processing Systems 22 (NIPS 2008), Cambridge, MA: MIT Press*.

Hull, C. L. (1943). Principles of behavior: an introduction to behaviortheory. New-York: Appleton-Century-Croft.

Montgomery, K. (1954). The role of exploratory drive in learning. Journal of Comparative and Physiological Psychology, 47:60–64.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston,Row, Peterson.

Kagan, J. (1972). Motives and development. Journal of Personality andSocial Psychology, 22:51–66.

Hunt, J. M. (1965). Intrinsic motivation and its role in psychologicaldevelopment. Nebraska symposium on motivation, 13:189–282.

References (2)

De Charms, R. (1968). Personal causation: the internal affectivedeterminants of behavior. Academic Press, New York.

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

References (3)

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.*

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.*