Toward an Artificial Intelligence with Intrinsic Semantics

AI/ML Lecture

Fall 2022

Yoonsuck Choe, Ph.D.

Department of Computer Science & Engineering

Texas A&M University

With N. Smith, H.-F. Yang, and N. Misra.

Introduction

Brief history of AI

- Symbolic logic, expert systems
- Probabilistic models: Bayesian Networks, Graphical Models
- Machine learning

Issue:

- Meaning of symbols, random variables, learned representations are not intrinsic to the system: Chinese Room, Symbol Grounding Problem.

Latest Trend

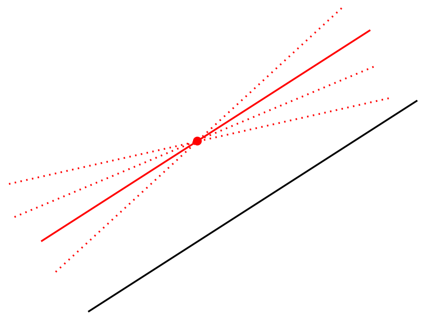
Deep learning is dominating ML/AI:

- Big data, fast computing (GPGPU), powerful algorithms, new tools (Tensorflow, etc.)
- Benchmark breaking performance in recognition tasks.
- Key idea is to build higher levels of **representations**.

Questions Regarding Deep Neural Networks

- What is deep neural network doing?
- How can we understand the internal workings?
- What do the hidden unit acitivities and wieghts represent?
- Why are these questions important?
  - For mission critical tasks where neural networks are used, we need assurances.
- Similar questions can be asked about the brain.

Asking the Right Questions



Parallel postulate

Euclid

Beltrami

- Asking the right question is critical:
  - How can we solve X? $\longrightarrow$ Unsolvable
  - Can we solve X? $\longrightarrow$ Solvable

cf. Choe and Mann (2012)

Reframing the Questions

Two related questions:

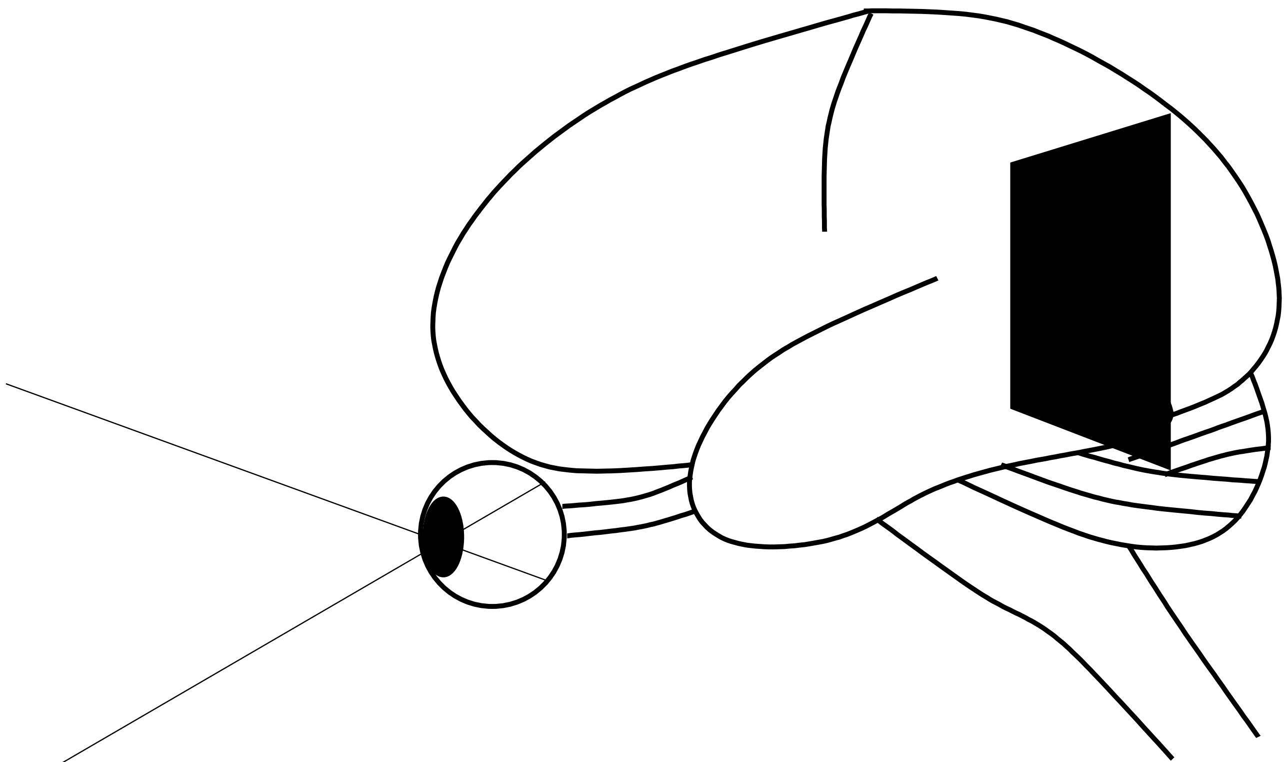
- How can we understand the neural network/brain?
- How can the neural network/brain itself understand its internal state?

These two questions lead to fundamentally different insights.

What Does This Mean?

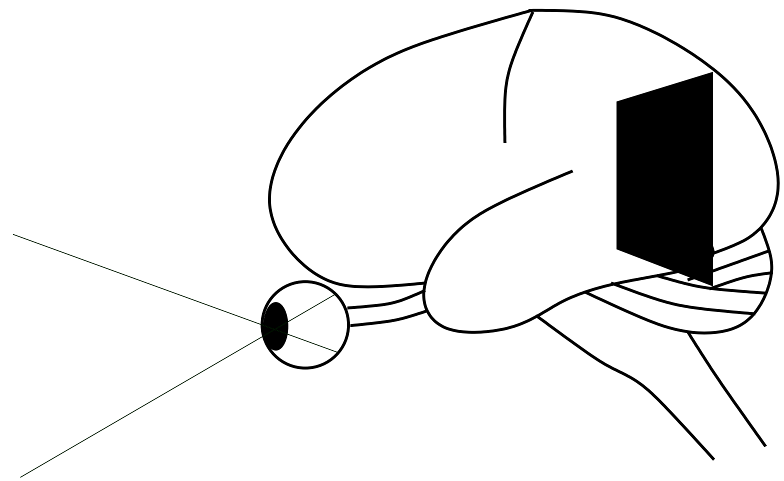
We are Clueless!

What If They Are Cortical Responses to Something



We are <u>Still</u> clueless!

They Are Visual Cortical Responses to Oriented Lines!



This is a problem of *grounding* (Harnad 1990), a problem that gets more severe as the representations become deeper and more complex. Note: Grounding = understanding.

Overview

- Grounding internal representations on action
- Learning internal representations together with grounding
- Perceptual vs. motor representations

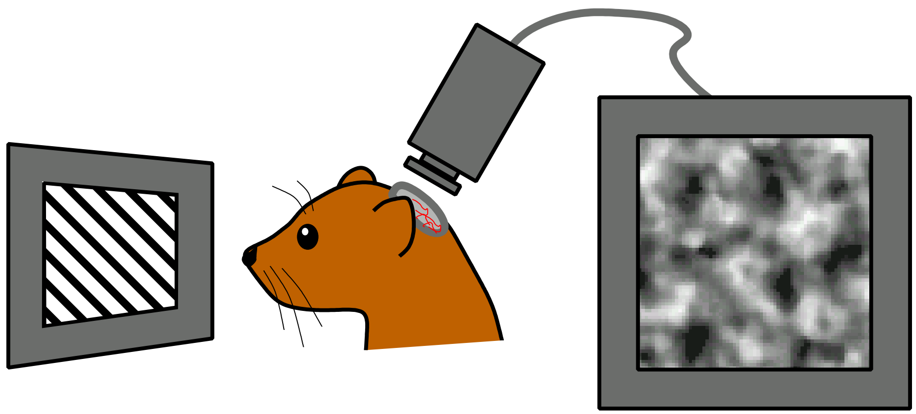
Part I: Grounding

Choe et al. (2007); Choe and Smith (2006); Choe and Bhamidipati (2004)

What Is Grounding?

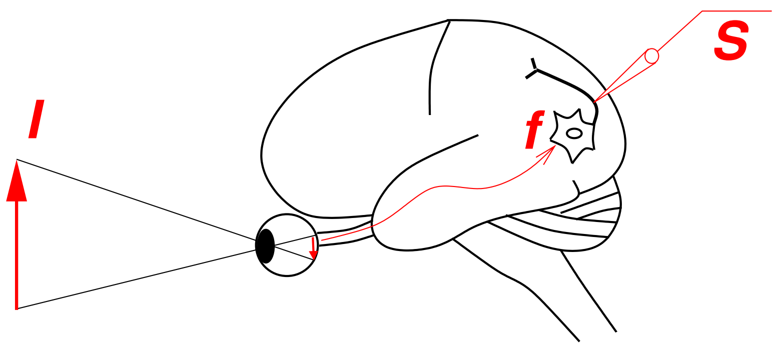
... How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols? ...
– Harnad (1990)

- Given a representation, figure out what it represents/means.
- Given an activity pattern in the brain, figure out what information it carries (decoding, decompression, etc.).

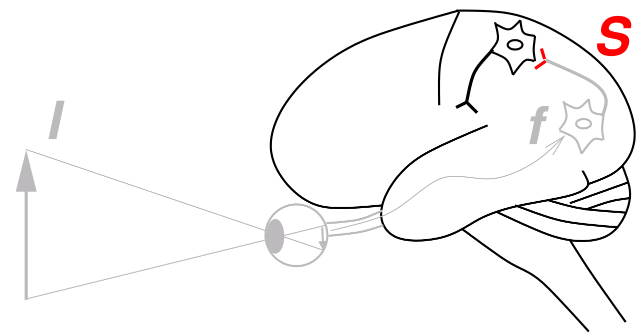


Miikkulainen et al. (2005); Weliky et al. (1995)

Grounding in the Brain



(a) External observer

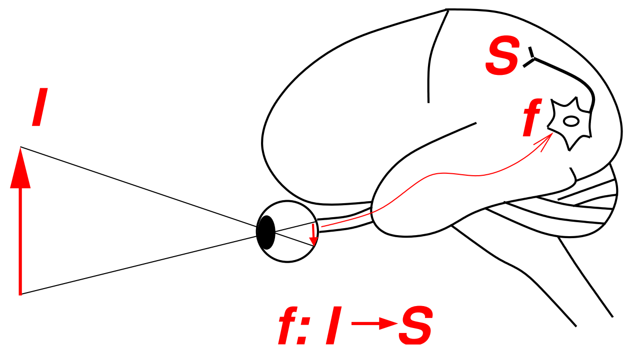


(b) Internal observer

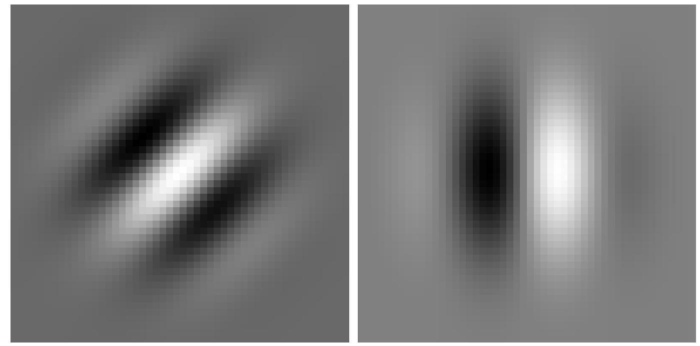
The problem of grounding, **within** the brain:

- **External observer** (e.g., a neuroscientist) **can figure out** how spike S relates to input I.
- **Internal observer cannot** seem to, which does not make sense at all.

Example: The Visual Cortex



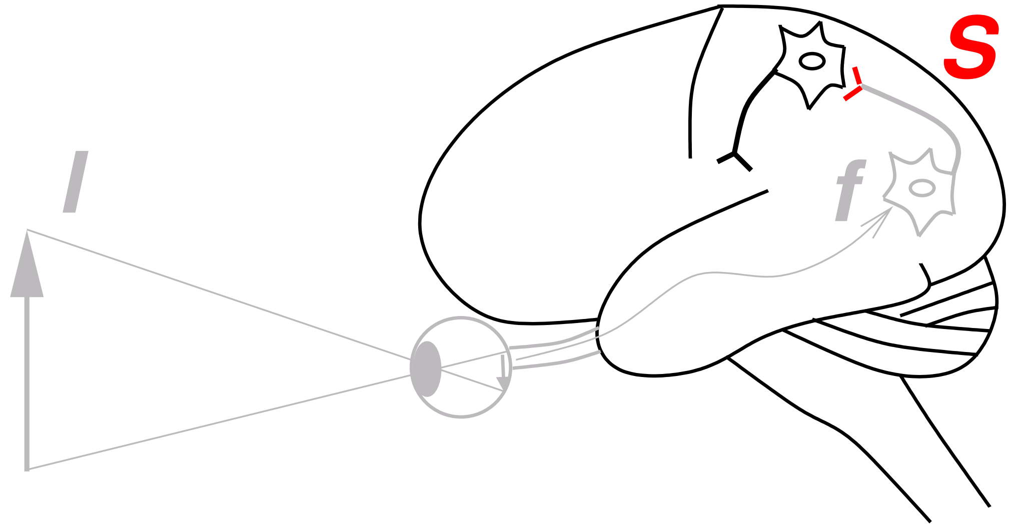
V1 Response to Input



Gabor-like RFs

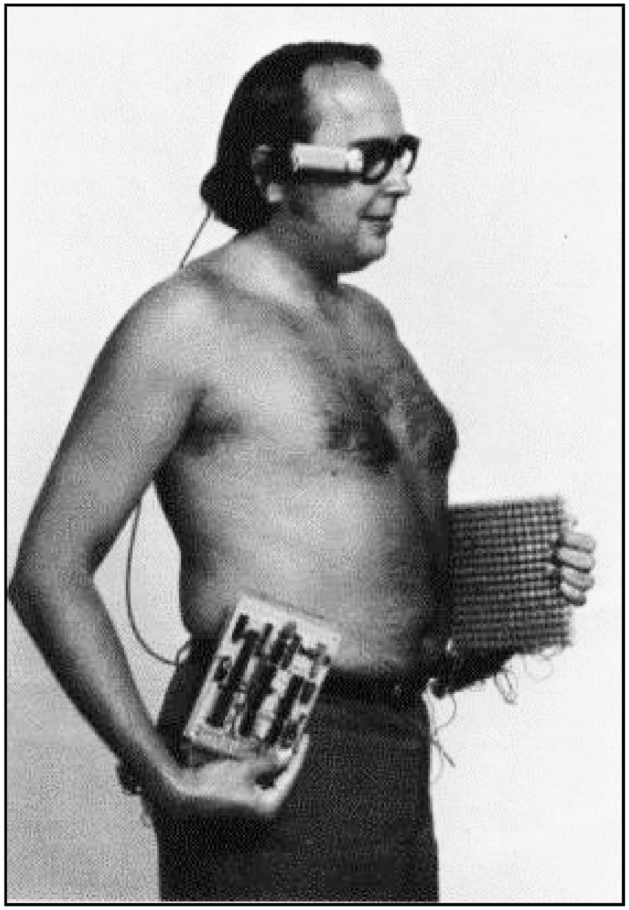
- With access to both I and S, Hubel and Wiesel (1959) figured out $f : I \to S$ in V1 (oriented Gabor-like receptive fields Jones and Palmer 1987).
- But even before that, and with access to only S, humans had no problem perceiving orientation.

Possible Solution: Allow Action



- A major problem in the picture is the **passiveness** of the whole situation.
- Adding action **can help solve** the problem.
- But **why** and **how**?

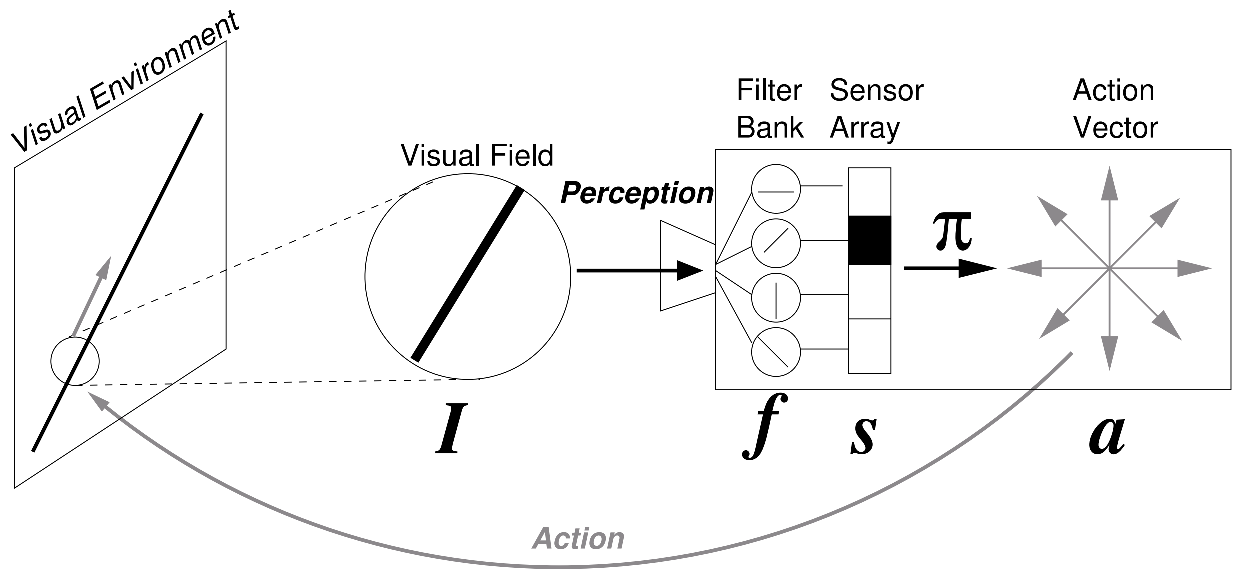
Experimental Evidence



Bach y Rita (1972; 1983)

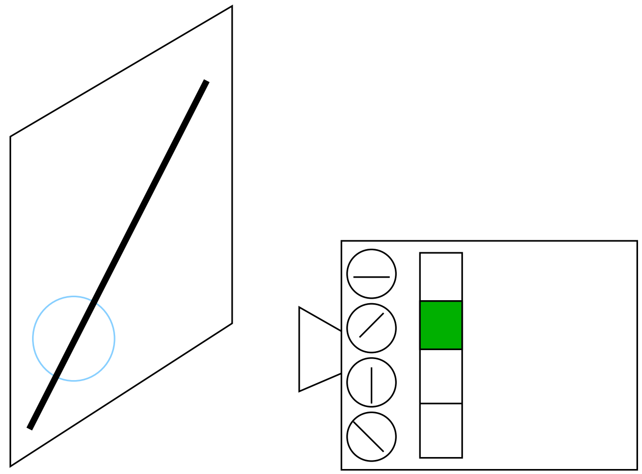
- Vibrotactile array linked to a video camera.
- Passive viewing results in **tactile** sensation.
- Moving the camera results in a **vision-like** sensation.
- Sensation as related to **voluntary/intentional action** may be the key!

Approach: Grounding Through Action



- Direct access to **encoded internal state** (sensory array) only.
- Action is enabled, which can **move the gaze**.
- How does this solve the grounding problem?

Action for Unchanging Internal State

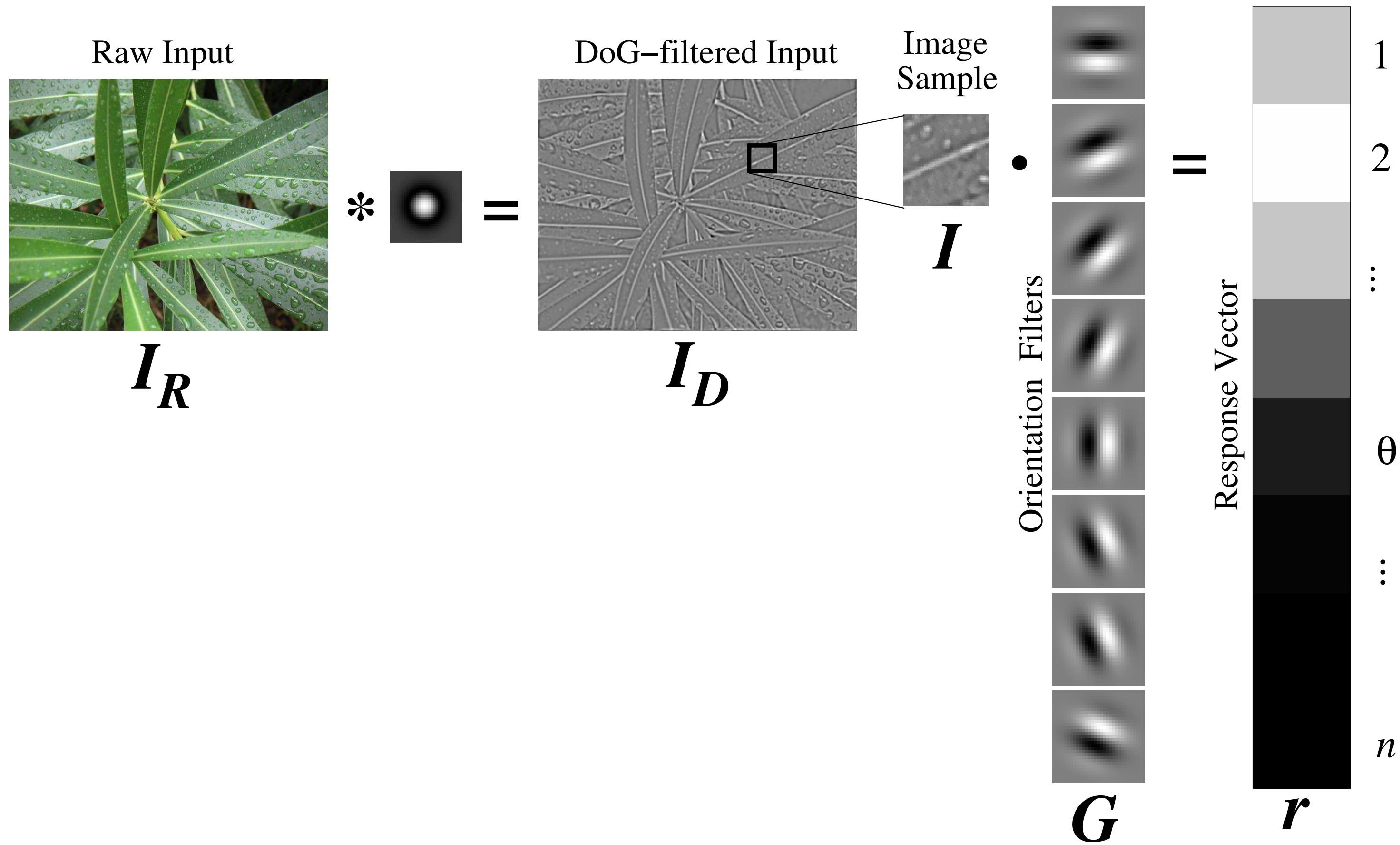


- Diagonal motion causes the *internal state* to **remain unchanging** over time.
- Property of such a movement **exactly reflects** the property of the input I: Semantics figured out through action.

Task

- Given an encoded sensory signal s, we want to learn action a that **maximizes the invariance** in the internal state over time.
- The learned action a will give **meaning** to s.
- This is basically a **reinforcement learning** task.

Methods: Orientation Response



Sensory state:

$$s = \underset{1 \le \theta \le n}{\arg\max}\ r_\theta .$$

Methods: Reinforcement Learning

Learn policy $\pi : S \rightarrow A$.

- Reward ρ: Similarity between previous and current internal state.

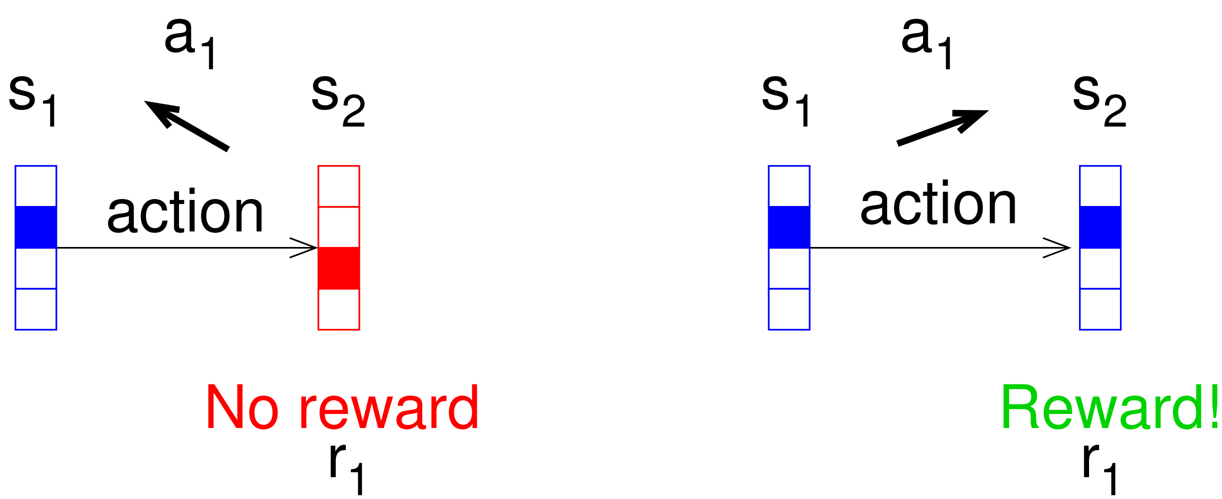
- Learning reward function $R(s, a)$:

$$R_{t+1}(s_t, a_t) = R_t(s_t, a_t) + \alpha\, \rho_{t+1},$$

  followed by normalization.

- Policy π derived from learned $R(s, a)$.

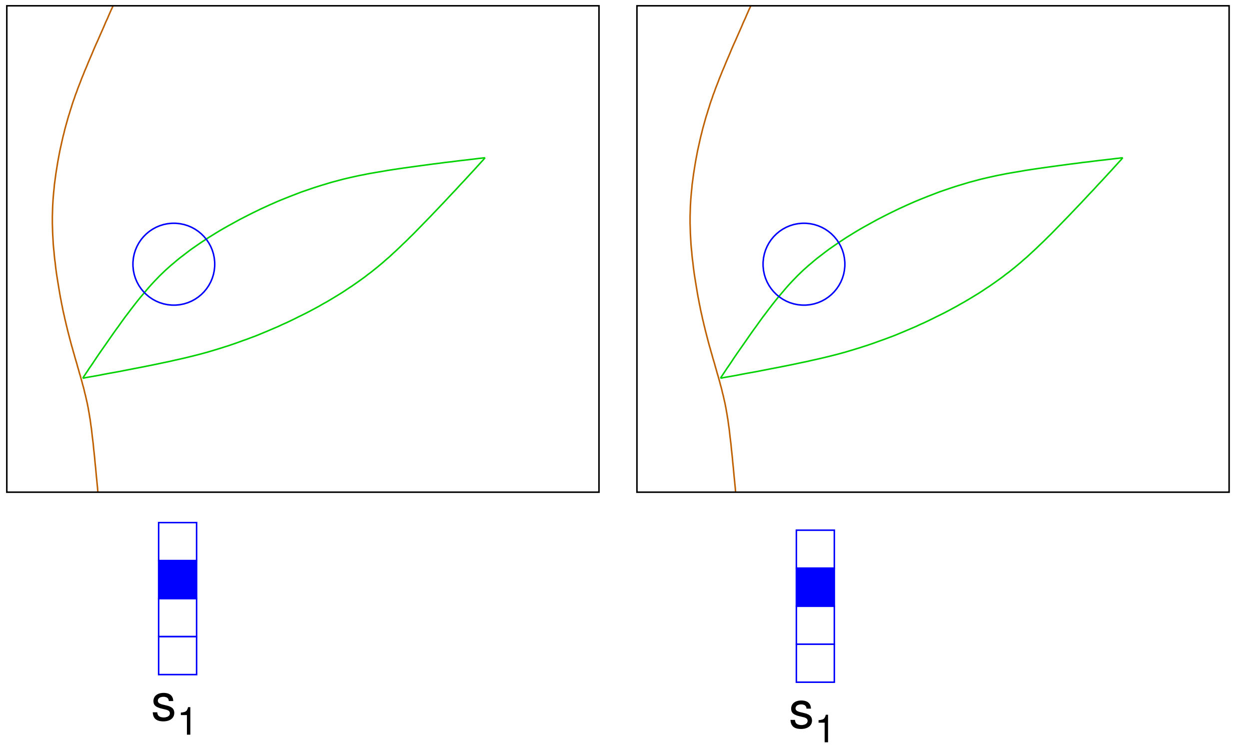
RL: Reward and Penalty ρ



Reward actions a that maintain invariance in s.

- If $s_1 = s_2$, Reward.
- If $s_1 \neq s_2$, Penalty.

RL: Reward and Penalty ρ



[Click figure to step through animation]

Reward actions a that maintain invariance in s.

- If $s_1 = s_2$, Reward.
- If $s_1 \neq s_2$, Penalty.

Reward Probability Table $R(s, a)$

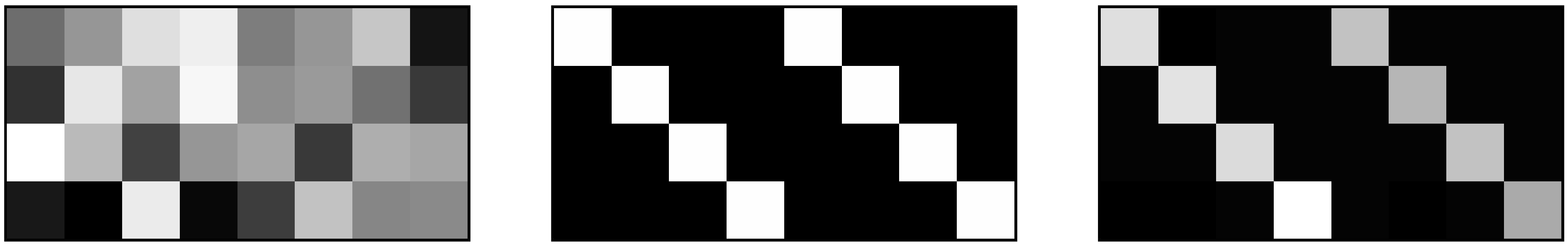
A: direction of motion

S: sensory state (orientation)

	→	↗	↑	↖	←	↙	↓	↘
─	0.5	0	0	0	0.5	0	0	0
/	0	0.5	0	0	0	0.5	0	0
|	0	0	R(s ,a)	0	0	0	0.5	0
\	0	0	0	0.5	0	0	0	0.5

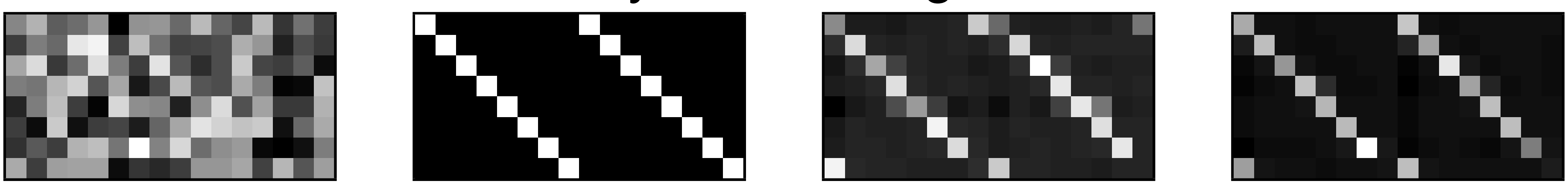
- Reward probability $R(s, a)$ can be tabulated.
- In an ideal case (world consists of straight lines only), we expect to see two diagonal matrices (shaded gray, above).

Results: Learned $R(s, a)$



(a) Initial (b) Ideal (c) Final

Synthetic image

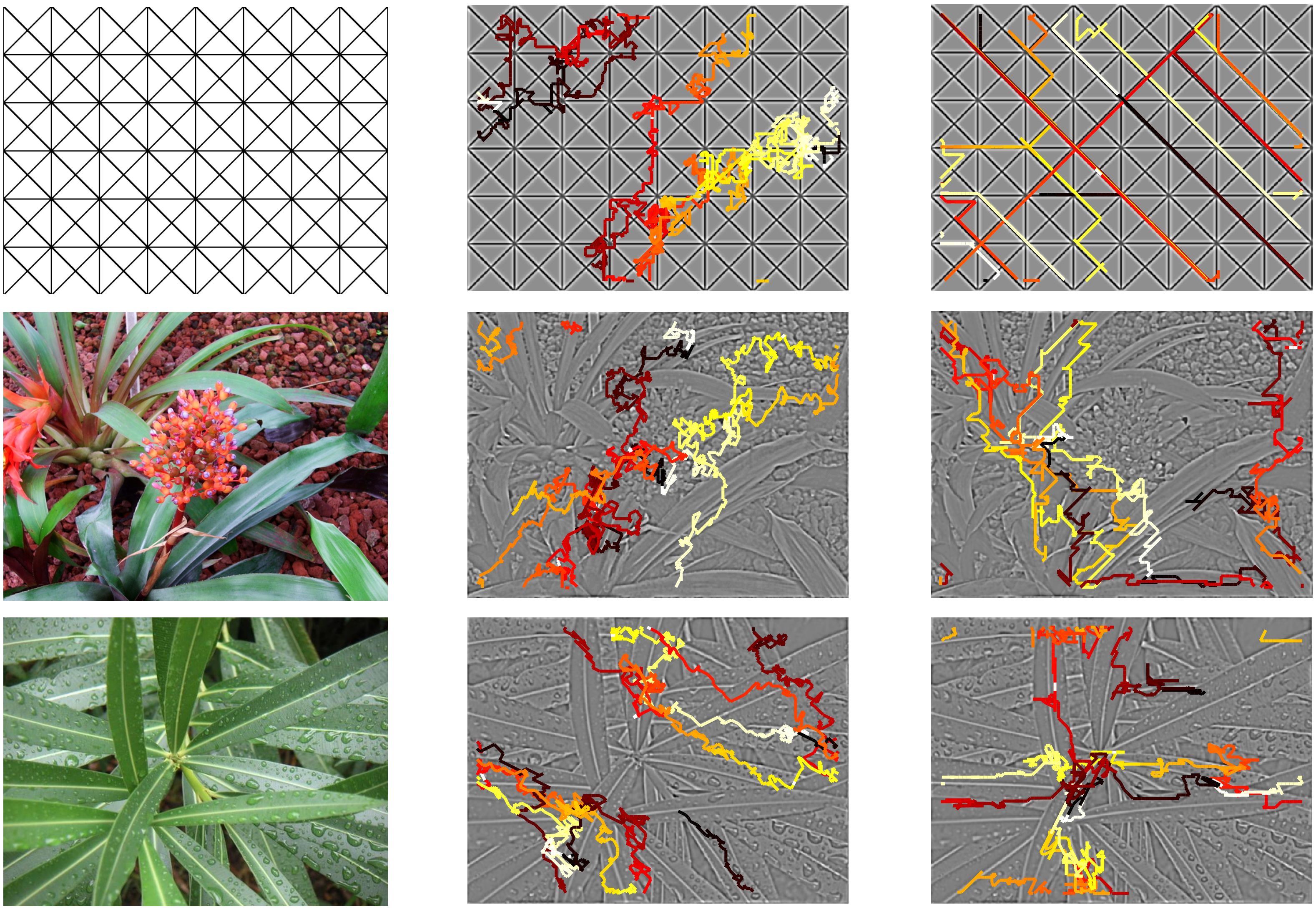


(a) Initial (b) Ideal (c) Plant (d) Oleander

Natural images

- Learned $R(s, a)$ close to ideal.

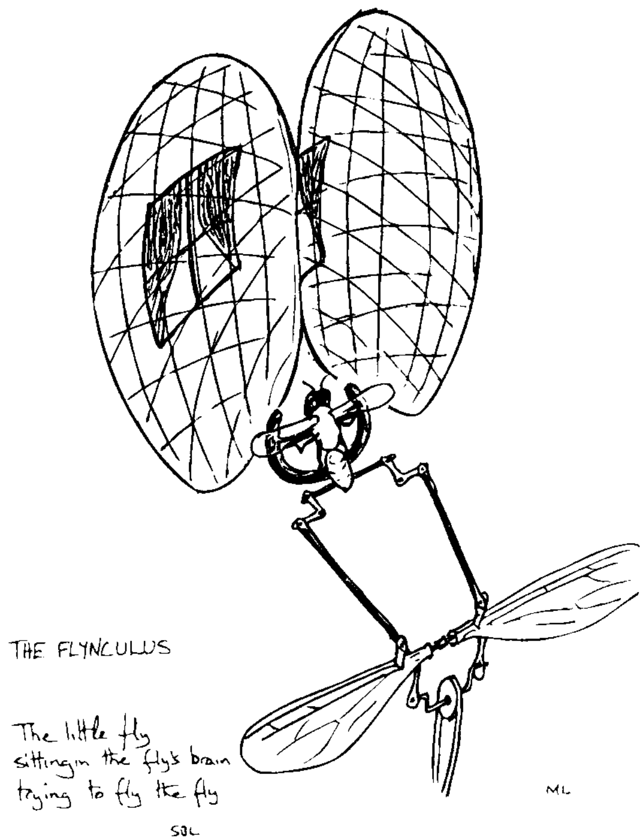
Results: Gaze Trajectory



(a) Input (b) Initial (c) Final

Results: Demo

Applications to Optic Flow



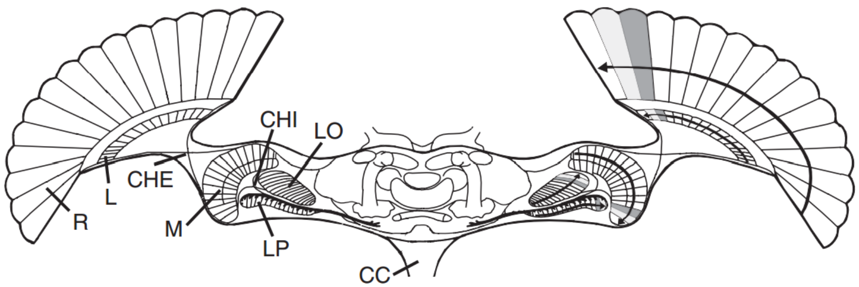
Same principle applied to the fly visual system:

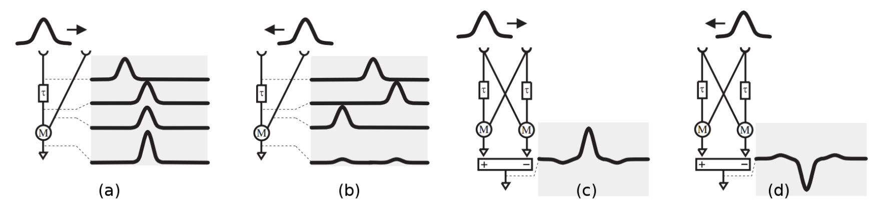
1. Fly Optic flow detectors (LPTC, Lobula Plate Tangential cells)
2. Learning the meaning of LTPC spikes: reinforcement learning based on internal state invarnance

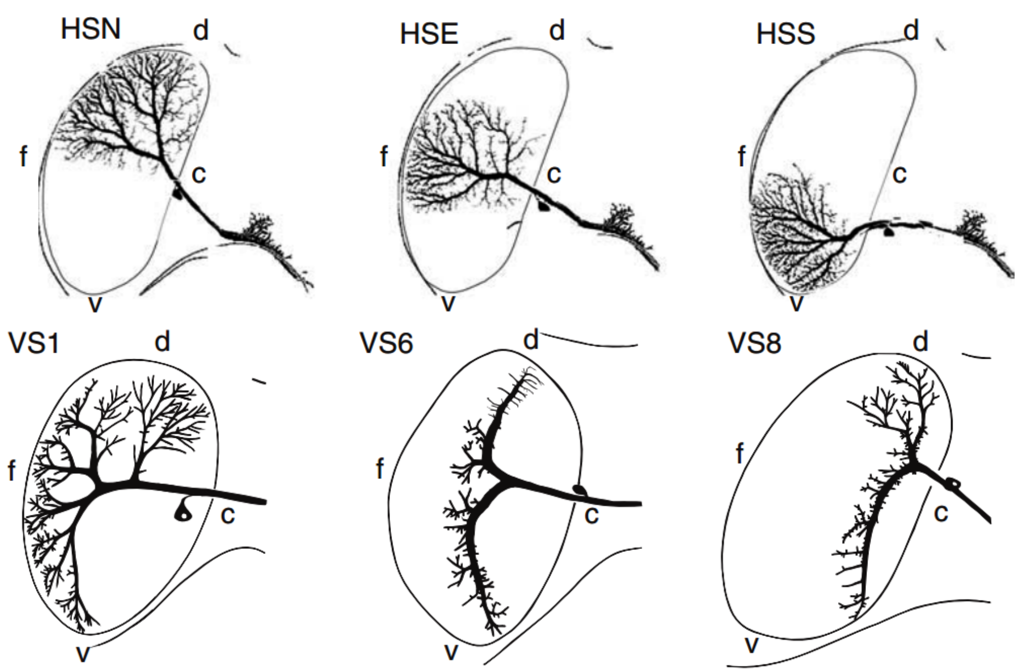
Cartoon from Rieke et al. (1997)

Parulkar and Choe IJCNN 2016 (Parulkar and Choe 2016).

Fly Visual System



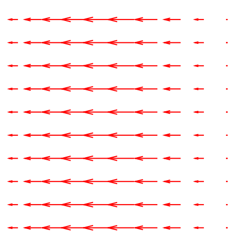




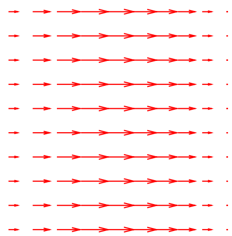
- Lamina (L) and Medulla (M):
  - Elementary Motion Detectors (EMD)
- Lobula Plate (LP) Tangential Cells (LPTCs)
  - HS and VS detect complex motion.

Borst and Egelhaaf (1989); Taylor and Krapp (2007)

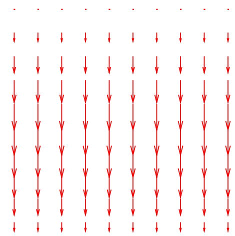
Fly Visual System Model



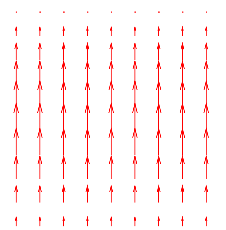
Rotation: Yaw
Right to Left
(RYRL)



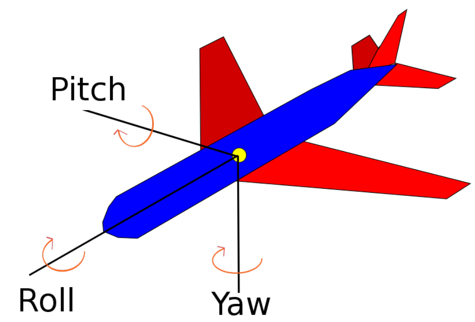
Rotation: Yaw
Left to Right
(RYLR)

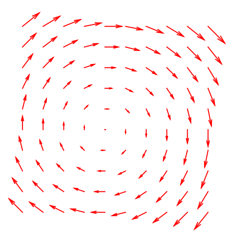


Rotation: Pitch
Up to Down
(RPUD)

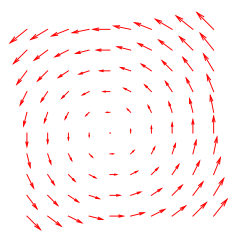


Rotation: Pitch
Down to Up
(RPDU)

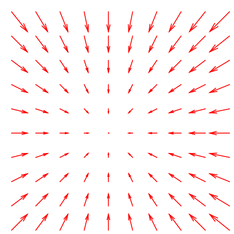




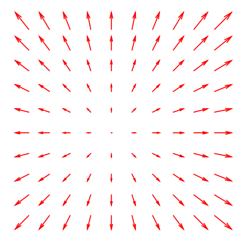
Rotation: Roll
Clockwise
(RRCL)



Rotation: Roll
X-clockwise
(RRAC)



Translation:
Radiate in
(TLRI)



Translation:
Radiate out
(TLRO)

- Initial optic flow computation: Lucas and Kanade (1981) method.
- HS: simple horizontal motion; VS: matched filter (roll and pitch [Krapp 2000])

Learning the Reward Table $R(s, a)$

Optical Flow States (s)

$R(s, a)$	Direction of Motion (a)							
	$\rightarrow$	$\leftarrow$	$\uparrow$	$\downarrow$	$\circlearrowleft$	$\circlearrowright$	$\blacktriangledown$	$\blacktriangle$
RYRL	1	0	0	0	0	0	0	0
RYLR	0	1	0	0	0	0	0	0
RPUD	0	0	1	0	0	0	0	0
RPDU	0	0	0	1	0	0	0	0
RRCL	0	0	0	0	1	0	0	0
RRAC	0	0	0	0	0	1	0	0
TLRI	0	0	0	0	0	0	1	0
TLRO	0	0	0	0	0	0	0	1

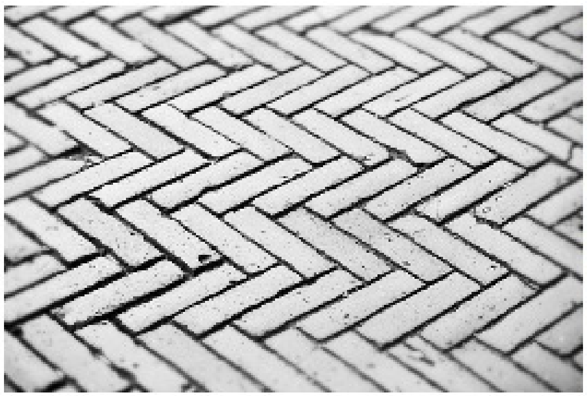
- Action is selected based on $P(a|s) = R(s, a)$.
- Learning (α: learning rate):

$$R_{t+1}(s_t, a_t) = R_t(s_t, a_t) + \alpha \rho_{t+1}, \text{ where}$$

$$\rho_{t+1} = 1 / \sqrt{\sum_i (r_{t+1,i} - r_{t,i})^2}$$

Finally, $R(s, a)$ is normalized over all a.

Experiments and Results: Input



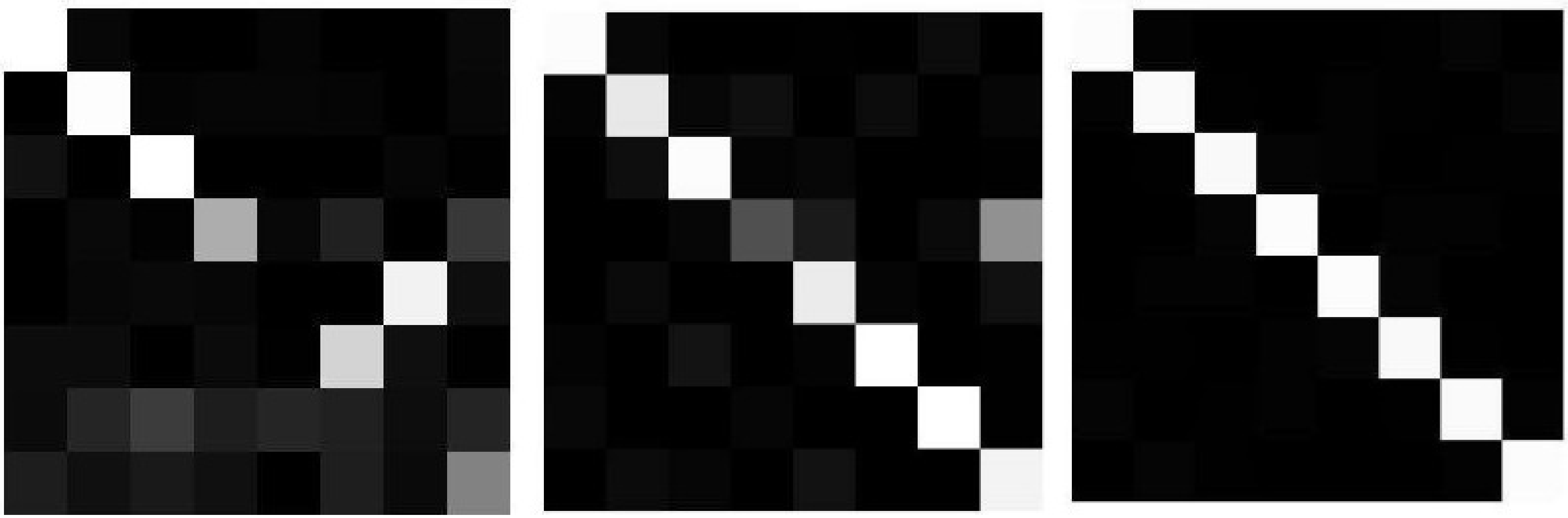
(a) Synthetic

(b) Natural 1

(c) Natural 2

- Model fly trained on three different inputs above.

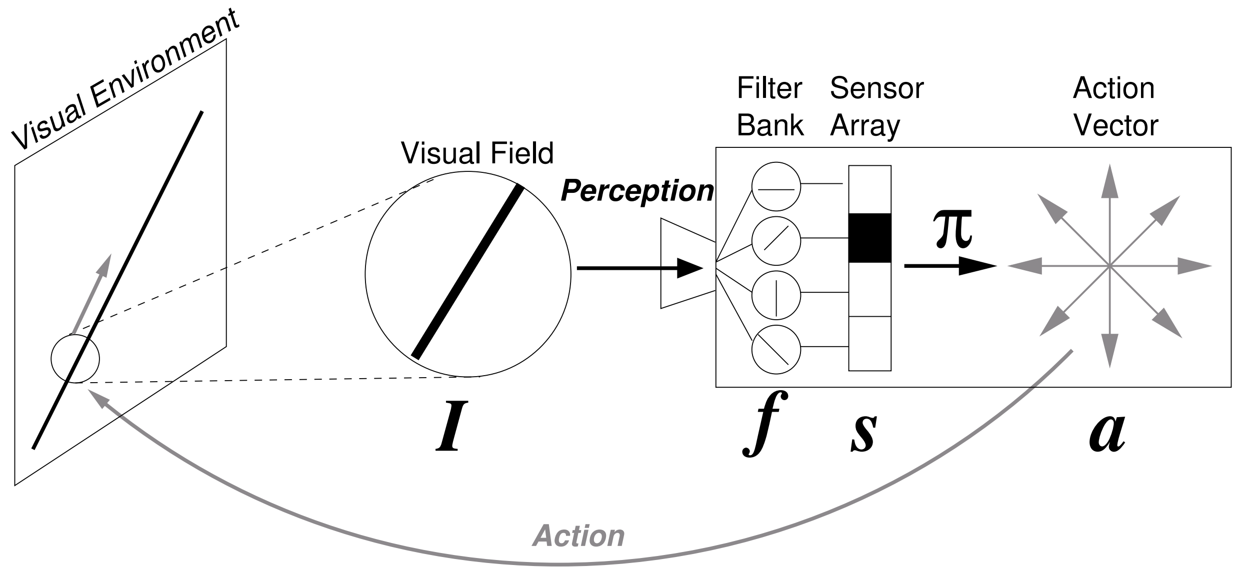
Experiments and Results: Learned R



(a) Synthetic (b) Natural 1 (c) Natural 2

- All three inputs lead to near-ideal $R(s, a)$.
- Given a certain internal state, action that has the same encoded property as that state is generated.

Part I: Summary

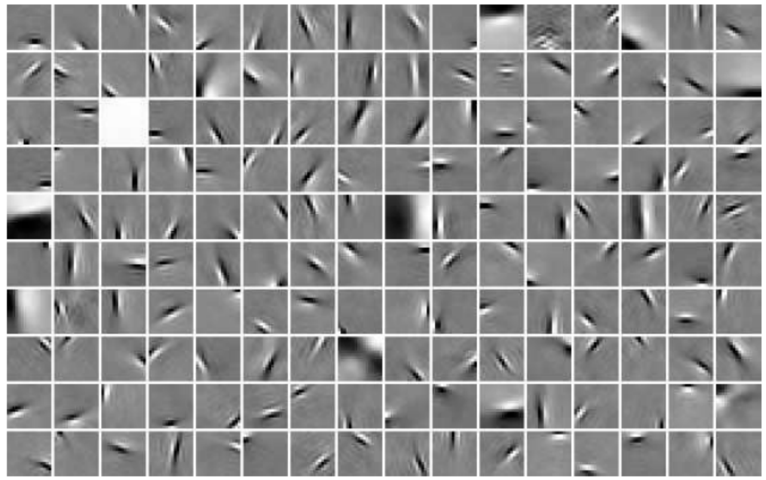


- (1) Using **invariance** as the only criterion, (2) particular **action pattern** was learned, (3) that has the **same property** as the input that triggered the sensors.

Part II: Learning Internal Representations

Yang and Choe (2007)

Theories of RF Formation



Hoyer and Hyvärinen (2000)

Well-developed understanding on how RFs form:

- Olshausen and Field (1997): Sparse coding; Barlow (1994): Redundancy reduction; Bell and Sejnowski (1997): Information maximization; Miikkulainen et al. (2005): Self-organization through Hebbian learning.

However, how is the resulting code to be used remains a question.

Questions

- The motor-based grounding experiment assumed that **receptive fields** are **given and fixed**.
- Can these be **learned** (developed) along with the grounding process?

Learning RFs along with Their Grounding (Decoding)

- Grounding (decoding): Same as Part I.
- RFs develop through local learning:

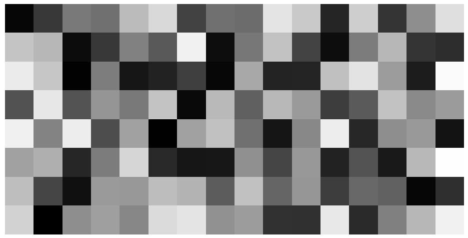
$$g_{ij} = \frac{g_{ij} + \alpha(I_{ij} - g_{ij})}{\sum_{mn} g_{mn} + \alpha(I_{mn} - g_{mn})},$$

where g_{ij} is the afferent connection weight and I_{ij} the input pixel value.

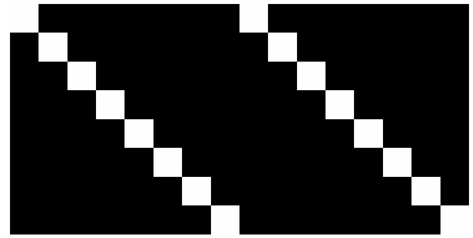
Experiments

1. Effects of different action policy on RF learning.
   - Random $R(s, a)$
   - Ideal $R(s, a)$
2. Simultaneous learning of RF and action policy.
   - RF learning through normalized Hebbian learning
   - Reinforcement learning of $R(s, a)$ based on internal-state invariance

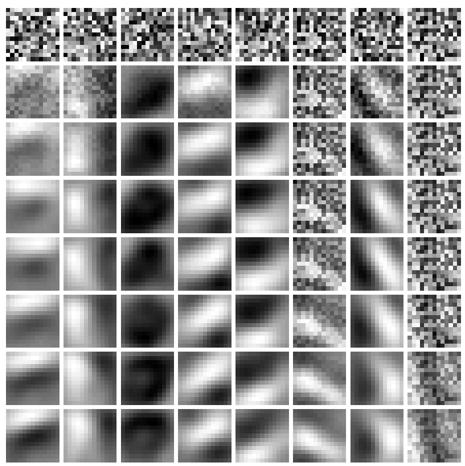
Effects of $R(s, a)$ on RF Learning



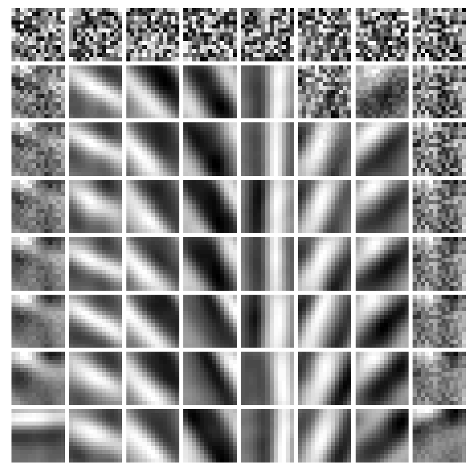
Fixed Random R



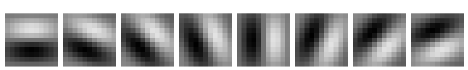
Fixed Ideal R



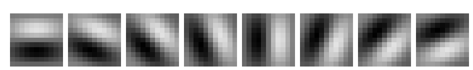
RF w/ Random Policy



RF w/ Ideal Policy

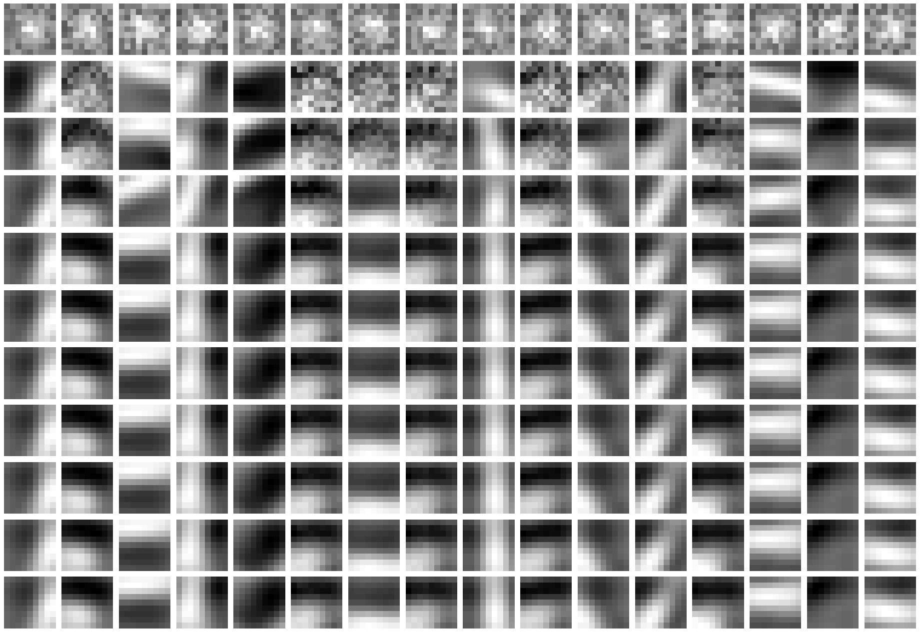


Reference RFs

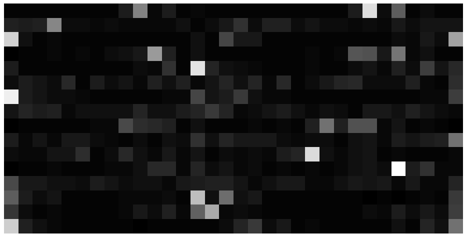


Reference RFs

Simul. Learning of RFs & $R(s, a)$



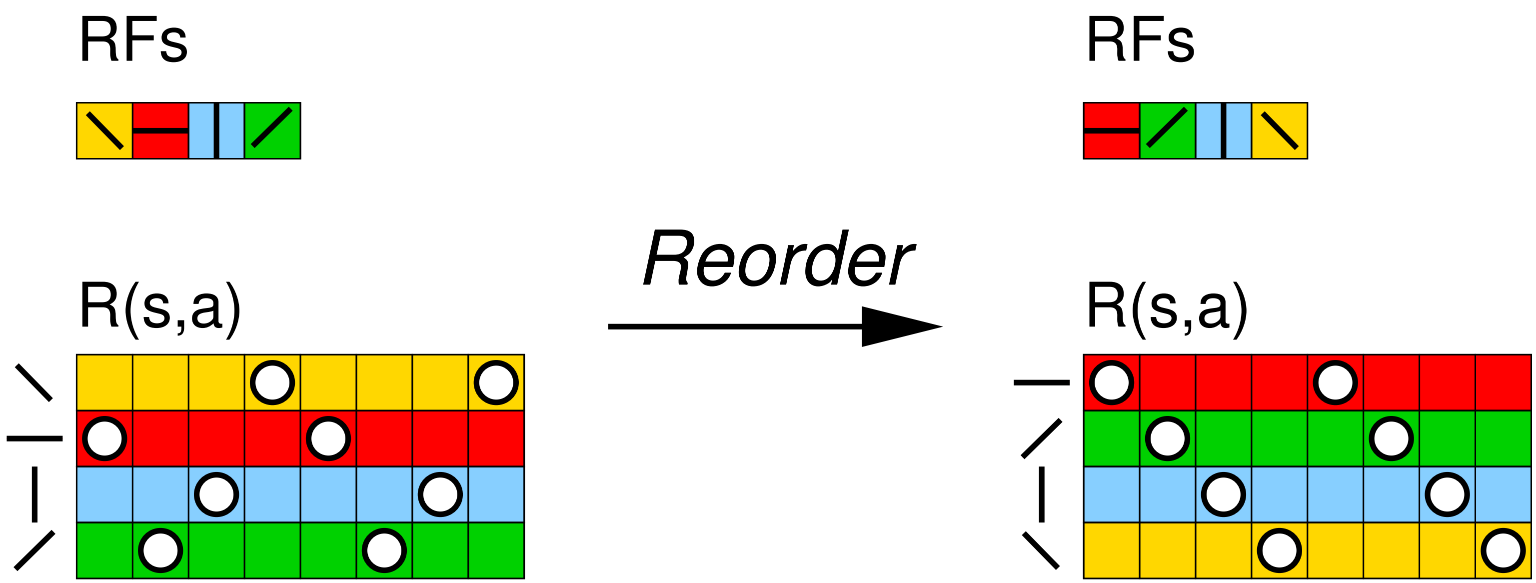
Learned RFs



Learned $R(s, a)$

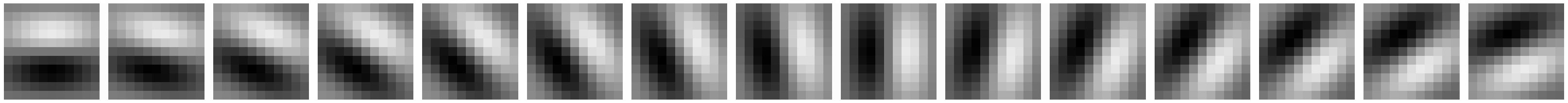
- Seemingly unordered RFs and $R(s, a)$ results.

Reordering RFs

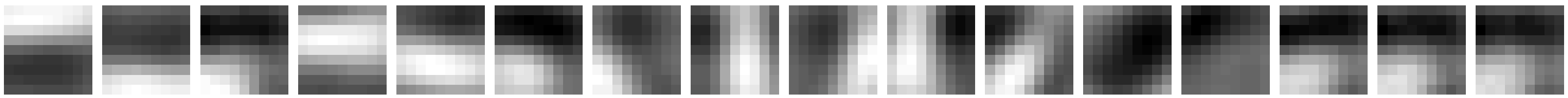


- The $R(s, a)$ result looks bad because each row's corresponding RF orientation is not ordered.
- Reordering RF orientation reorders the rows in $R(s, a)$.

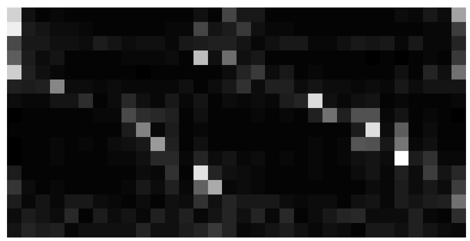
Reordered RFs and $R(s, a)$



Reference RFs



Reordered final RFs



Reordered final $R(s, a)$

- However, reordering the RFs and their corresponding $R(s, a)$ rows shows the true underlying structure! (Not perfect, but a good start!)

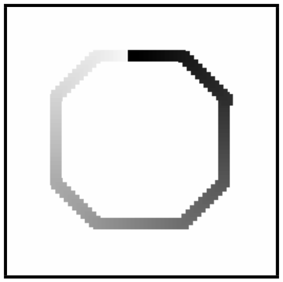
Part II: Summary

- Action policy strongly influences RF properties, by altering the input statistics.
- Certain action policies may give better RFs, faster.
- Receptive fields and action policy can learn simultaneously, from scratch, thus allowing encoding/decoding to evolve together.

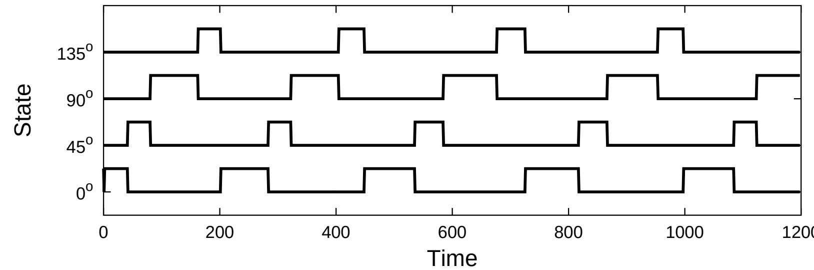
Part III: Perceptual vs. Motor Representations

Misra and Choe (2007)

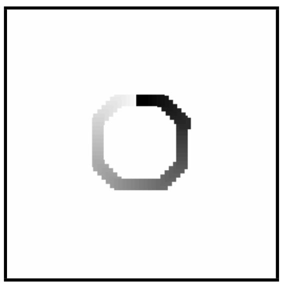
Learning About Shapes



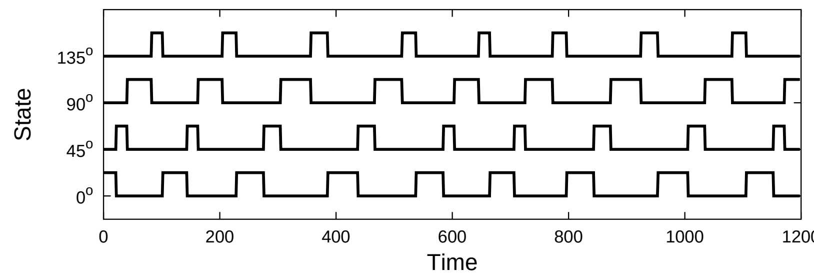
(a) Eye position (large input)



(b) Internal state (large input)



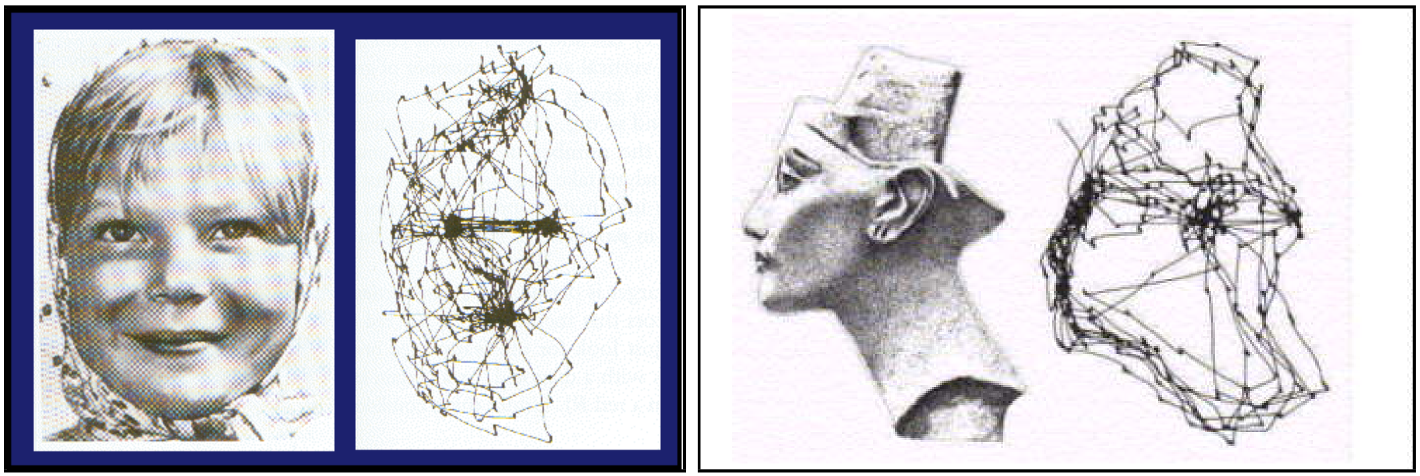
(c) Eye position (small input)



(d) Internal state (small input)

- For complex objects, a history of sensory activity may be needed (i.e., some form of memory).
- Invariance can be detected in the spatiotemporal pattern of sensor activity.

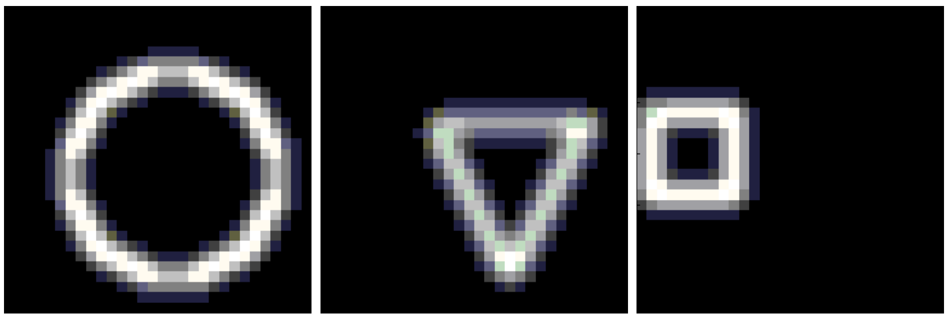
Motor System and Object Recognition



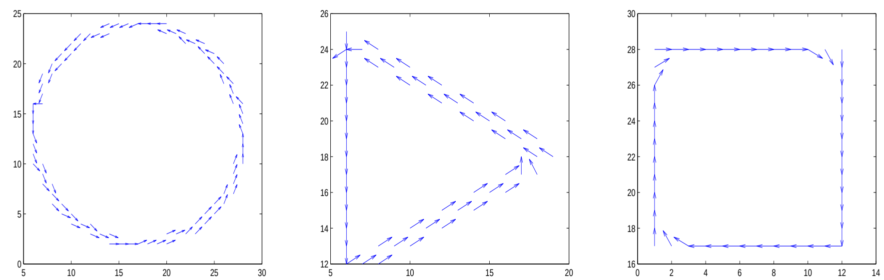
Yarbus (1967)

- When we look at objects, our gaze wanders around.
- Could such an interaction be necessary for object recognition?

Advantage of Motor-Based Memory (Habit, or Skill)



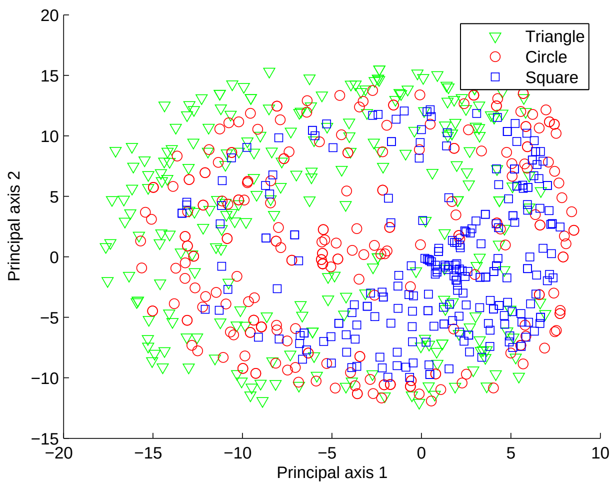
(a) Sensor-based Representation



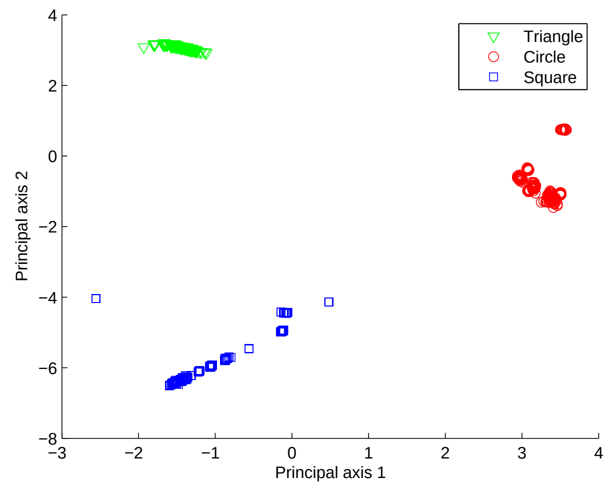
(b) Motor-based Representation

- Sensor-based representations may be hard to learn and inefficient.
- Motor-based approaches may generalize better.
- Comparison: Make both into a 900-D vector and compare backpropagation learning performance.

Class Separability



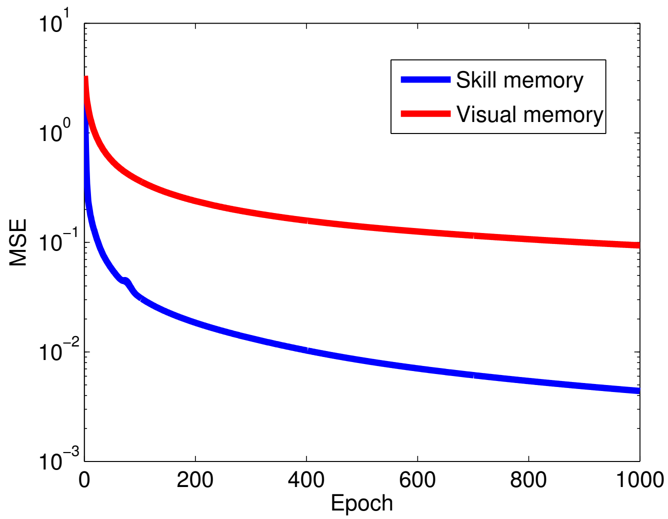
(a) Visual Memory



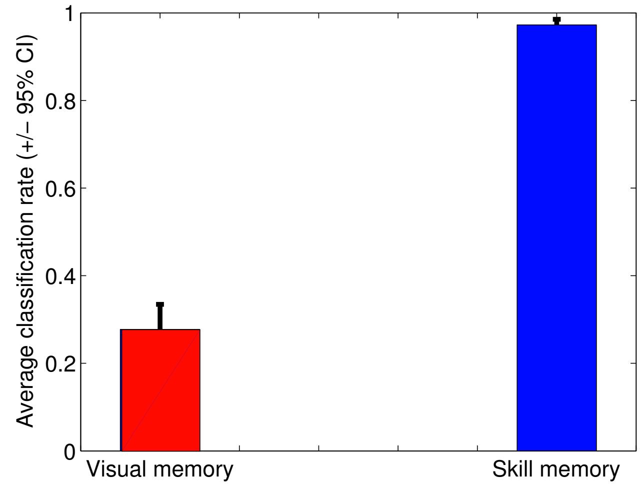
(b) Motor Memory

- Comparison of PCA projection of 1,000 data points in the visual and motor memory representations.
- Motor memory is clearly separable.

Speed and Accuracy of Learning



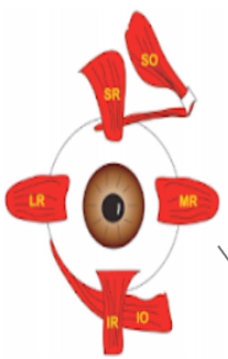
(a) Training Speed

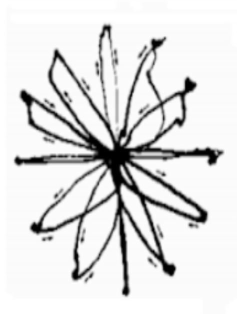


(b) Generalization Accuracy

- Motor-based memory resulted in faster and more accurate learning (10 trials).

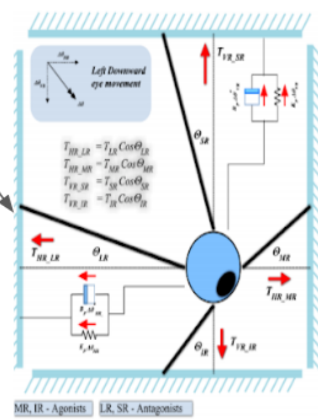
Motor Fidelity and Perceptual Performance



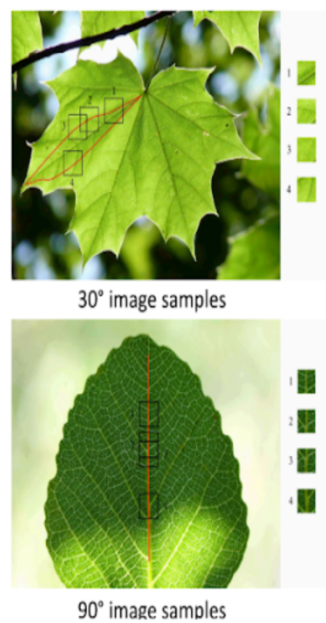


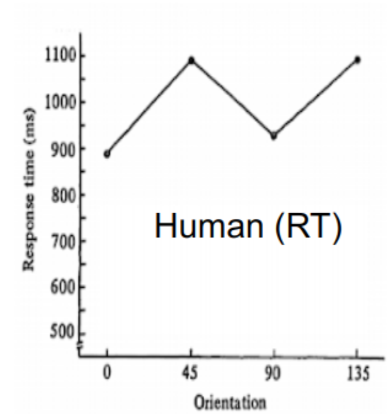
Eye movement more accurate along orthogonal direction

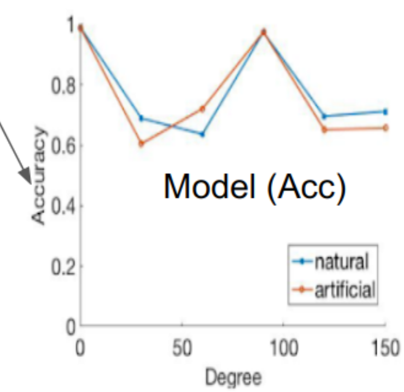
Simulated
Eye Movement



Data Generation
& Training







Nguyen et al. ICDL 2018

Part III: Summary

Motor-based representations of shape are

- More separable in the representational space,
- Faster to learn,
- Better at novel tasks (generalization), compared to sensory representations.

Perceptual performance depends on motor fidelity

Wrap Up

Related Works (Selected)

- Graziano (2009): Motor areas encoding a map of whole gestures (motor primitives?).
- Fuster (1997): perception-action link at all levels.
- Rizzolatti et al. (2001): Mirror neurons
- Salinas (2006): Sensory RF coding dictated by downstream requirements.
- Sejnowski (2006): Importance of “projective fields”.

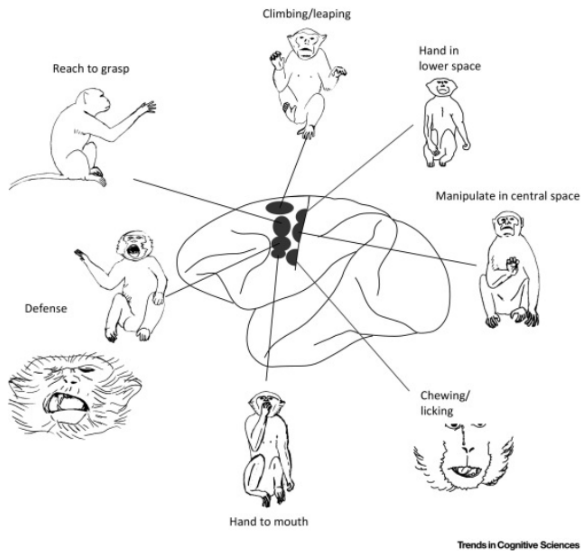
Discussion

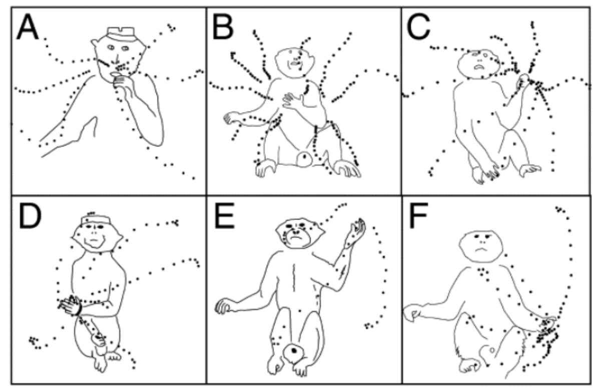
- Main contribution: Discovery of the invariance criterion for sensorimotor grounding, development, and recognition.
- Importance of self-generated action in autonomous understanding.
- Richer motor primitive repertoire can lead to richer understanding (compare different animals).
- Tool use can dramatically augment motor primitive repertoire.

Implications on Representation and Recognition

- Motor function is key for autonomous understanding of internal representations.
- Learning of internal representations can be influenced by motor policy.
- Internal representations and understanding can be learned at the same time.
- Motor-based representations lead to more reliable recognition.

Predictions





Motor cortex represents complex motions in a topographical organization

- Perceived orientation of a line can be altered by eye movement in the direction of incompatible orientation.
- Motor structures (cerebellum, basal ganglia) may be intimately involved in semantics.
- Geometrical understanding may be limited by the motor primitive repertoire.

Conclusions

- Motor action is important for understanding and grounding.
- Simple criterion (state invariance) can help link sensory coding with meaningful action.
- RFs can be developed along with grounding.
- Motor-based representations are more effective for shape recognition.
- Deep learning needs to look more into motor aspects of learning and autonomous understanding.

Credits

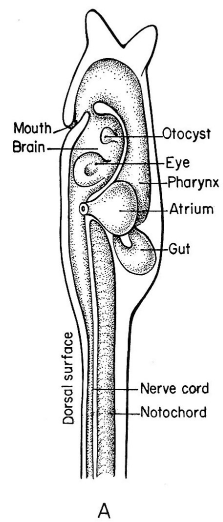
- Contributors: Kuncara A. Sukasdadi, S. Kumar Bhamidipati, Noah Smith, Stu Heinrich, Navendu Misra, Huei-Fang Yang, Daniel C.-Y. Eng
- Choe et al. (2008, 2007); Choe and Smith (2006); Choe and Bhamidipati (2004); Choe (2011)

Why Do We Have a Brain?

- Survival and reproduction? Think again!

Tree

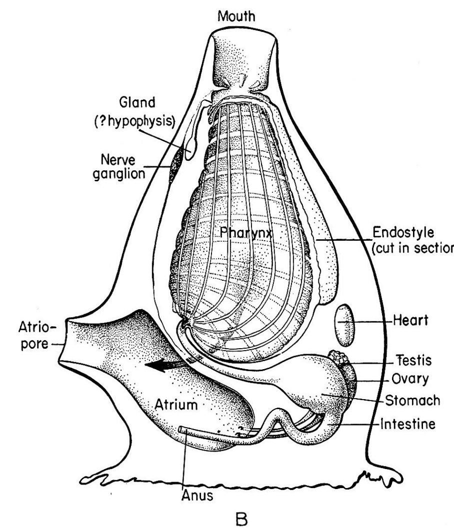
(no Brain)



Tunicate

Free-floating

(w/ Brain)



Tunicate

Settled

(w/o Brain)

Llinás (2001)

Sources: http://homepages.inf.ed.ac.uk/jbednar/ and http://bill.srnr.arizona.edu/classes/182/Lecture-9.htm

References

Bach y Rita, P. (1972). *Brain Mechanisms in Sensory Substitution*. New York: Academic Press.

Bach y Rita, P. (1983). Tactile vision substitution: Past and future. *International Journal of Neuroscience*, 19:29–36.

Barlow, H. (1994). What is the computational goal of the neocortex? In Koch, C., and Davis, J. L., editors, *Large Scale Neuronal Theories of the Brain*, 1–22. Cambridge, MA: MIT Press.

Bell, A. J., and Sejnowski, T. J. (1997). The "independent components" of natural scenes are edge filters. *Vision Research*, 37:3327.

Borst, A., and Egelhaaf, M. (1989). Principles of visual motion detection. *Trends in neurosciences*, 12(8):297–306.

Choe, Y. (2011). Action-based autonomous grounding. In Modayil, J., Precup, D., and Singh, S., editors, *AAAI-11 Workshop on Lifelong Learning from Sensorimotor Experience*, 56–57. Palo Alto, CA: AAAI Press. AAAI Workshop Technical Report WS-11-15.

Choe, Y., and Bhamidipati, S. K. (2004). Autonomous acquisition of the meaning of sensory states through sensory-invariance driven action. In Ijspeert, A. J., Murata, M., and Wakamiya, N., editors, *Biologically Inspired Approaches to Advanced Information Technology*, Lecture Notes in Computer Science 3141, 176–188. Berlin: Springer.

Choe, Y., and Mann, T. A. (2012). From problem solving to problem posing. *Brain-Mind Magazine*, 1:7–8.

Choe, Y., and Smith, N. H. (2006). Motion-based autonomous grounding: Inferring external world properties from internal sensory states alone. In Gil, Y., and Mooney, R., editors, *Proceedings of the 21st National Conference on Artificial Intelligence(AAAI 2006)*, 936–941.

Choe, Y., Yang, H.-F., and Eng, D. C.-Y. (2007). Autonomous learning of the semantics of internal sensory states based on motor exploration. *International Journal of Humanoid Robotics*, 4:211–243.

Choe, Y., Yang, H.-F., and Misra, N. (2008). Motor system's role in grounding, receptive field development, and shape recognition. In *Proceedings of the Seventh International Conference on Development and Learning*, 67–72. IEEE.

Fuster, J. M. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and the Frontal Lobe*. Lippencott-Raven. Third edition.

Graziano, M. (2009). *The Intelligent Movement Machine: An Ethological Perspective on the Primate Motor System*. Oxford, UK: Oxford University Press.

Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42:335–346.

Hoyer, P. O., and Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems*, 11:191–210.

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148:574–591.

Jones, J. P., and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

Krapp, H. G. (2000). Neuronal matched filters for optic flow processing in flying insects. *International review of neurobiology*, 44:93–120.

Llinás, R. R. (2001). *I of the Vortex*. Cambridge, MA: MIT Press.

Lucas, B. D., and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*, vol. 81, 674–679.

Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. Berlin: Springer. URL: http://www.computationalmaps.org. 538 pages.

Misra, N., and Choe, Y. (2007). Shape recognition through dynamic motor representations. In Kozma, R., and Perlovsky, L., editors, *Neurodynamics of Higher-Level Cognition and Consciousness*, 185–210. Berlin: Springer.

Olshausen, B. A., and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37:3311–3325.

Parulkar, A., and Choe, Y. (2016). Motor-based autonomous grounding in a model of the fly optic flow system. In *Proceedings of the International Joint Conference on Neural Networks*, 4354–4361.

Rieke, F., Warland, D., de Ruter van Steveninck, R., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press. First edition.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2:661–670.

Salinas, E. (2006). How behavioral constraints may determine optimal sensory representations. *PLoS Biology*, 4:2383–2392.

Sejnowski, T. J. (2006). What are the projective fields of cortical neurons? In van Hemmen, L. J., and Sejnowski, T. J., editors, *Twenty Three Problems in Systems Neuroscience*, 394–405. Oxford, UK: Oxford University Press.

Taylor, G. K., and Krapp, H. G. (2007). Sensory systems and flight stability: what do insects measure and why? *Advances in insect physiology*, 34:231–316.

Weliky, M., Kandler, K., Fitzpatrick, D., and Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron*, 15:541–552.

Yang, H.-F., and Choe, Y. (2007). Co-development of visual receptive fields and their motor-primitive-based decoding scheme. In *Proceedings of the International Joint Conference on Neural Networks 2007 Post conference Workshop on Biologically-inspired Computational Vision (BCV) 2007*.