Motor System's Role in Grounding, Development, and Recognition in Vision

Forum for AI at UTCS: *October 10, 2008*

Yoonsuck Choe

Department of Computer Science

Texas A&M University

Based on Choe, Yang, and Misra (ICDL 2008)

Motivation and Overview

Important aspects of vision may be hidden in its coupling with motor function.

1. Grounding of internal representations in the visual system.
2. Development/co-development of visual receptive fields with their grounding.
3. Visual recognition facilitated by motor exploration.

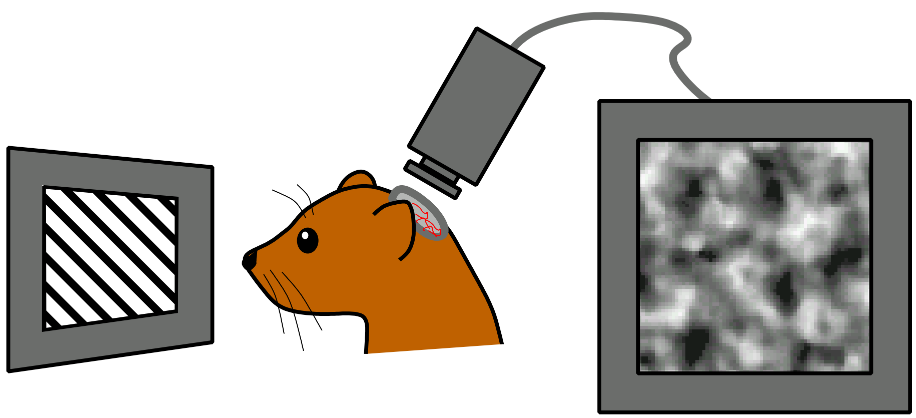
Part I: Grounding

Choe et al. (2007)

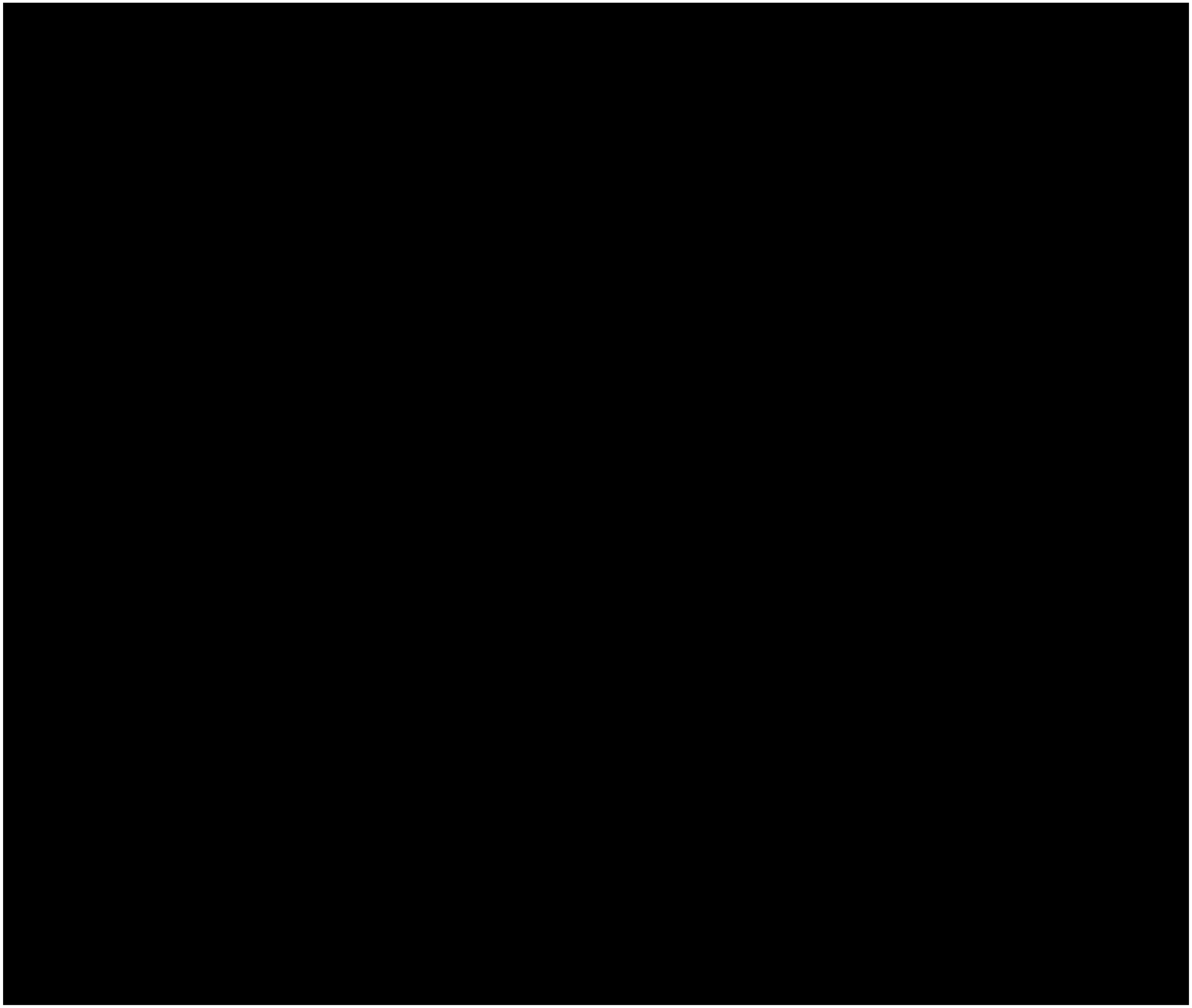
What Is Grounding?

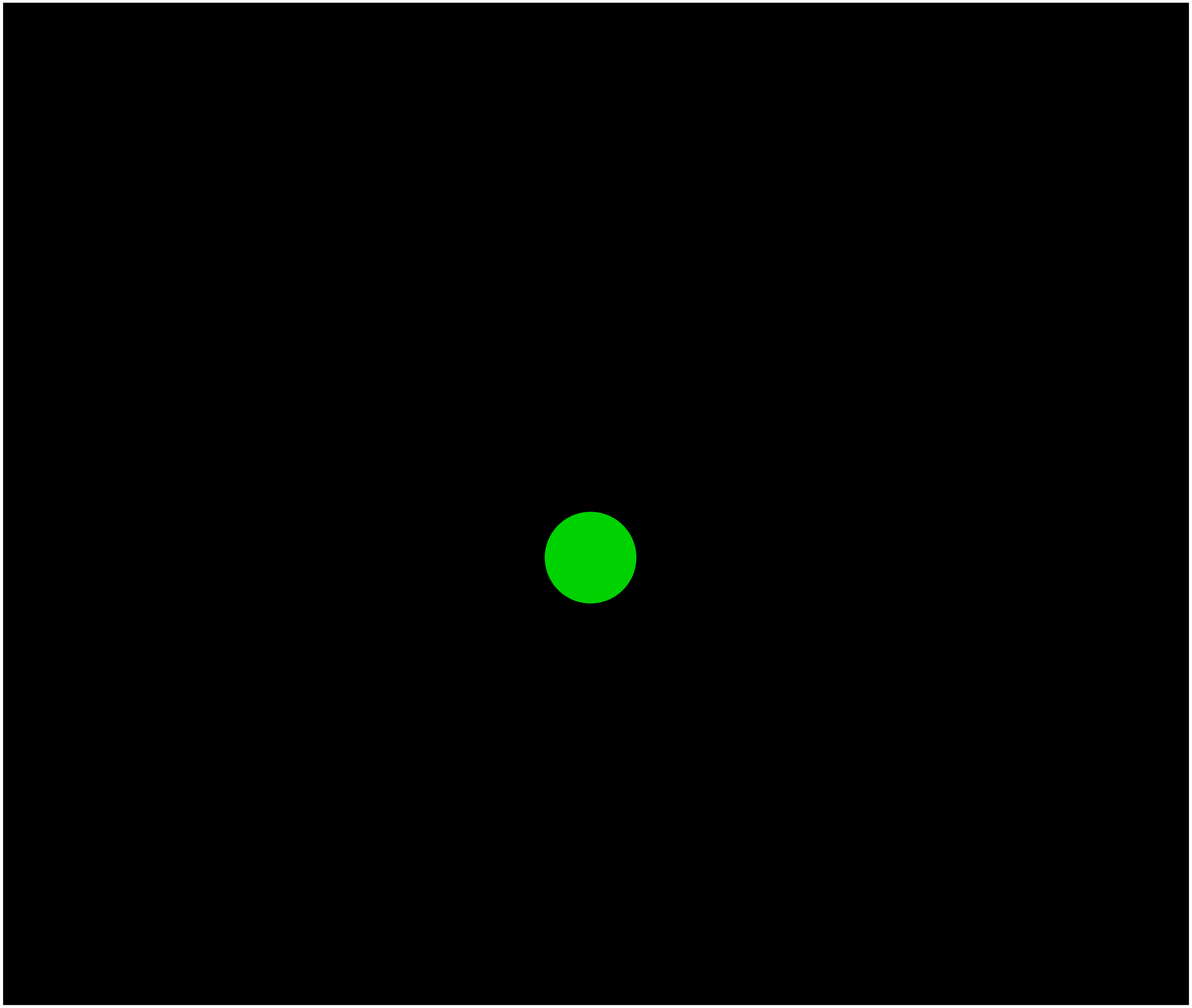
... How can the meanings of the meaningless symbol tokens, manipulated solely on the basis of their (arbitrary) shapes, be grounded in anything but other meaningless symbols? ...
– Harnad (1990)

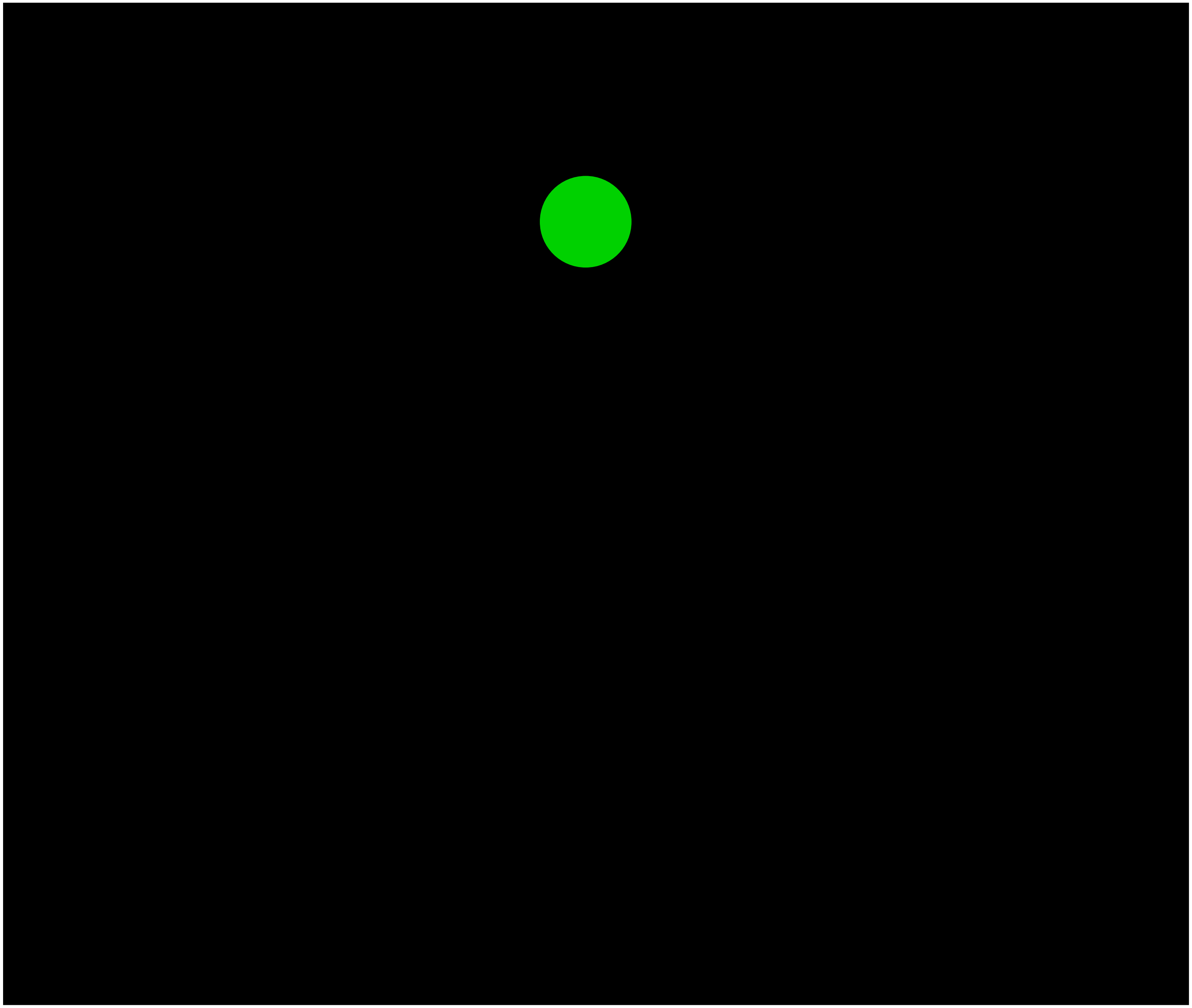
- Given a representation, figure out what it represents/means.
- Given an activity pattern in the brain, figure out what information it carries.

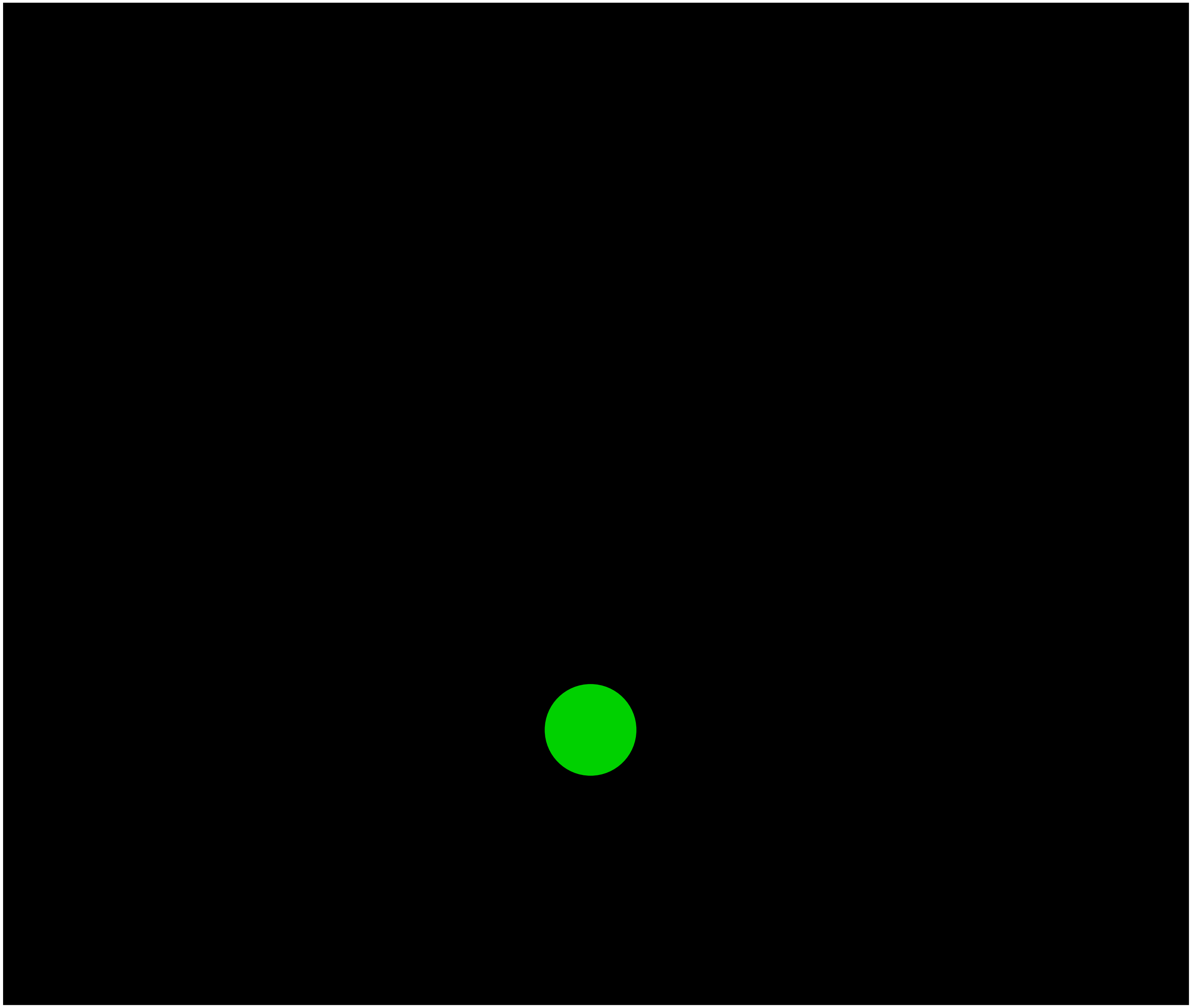


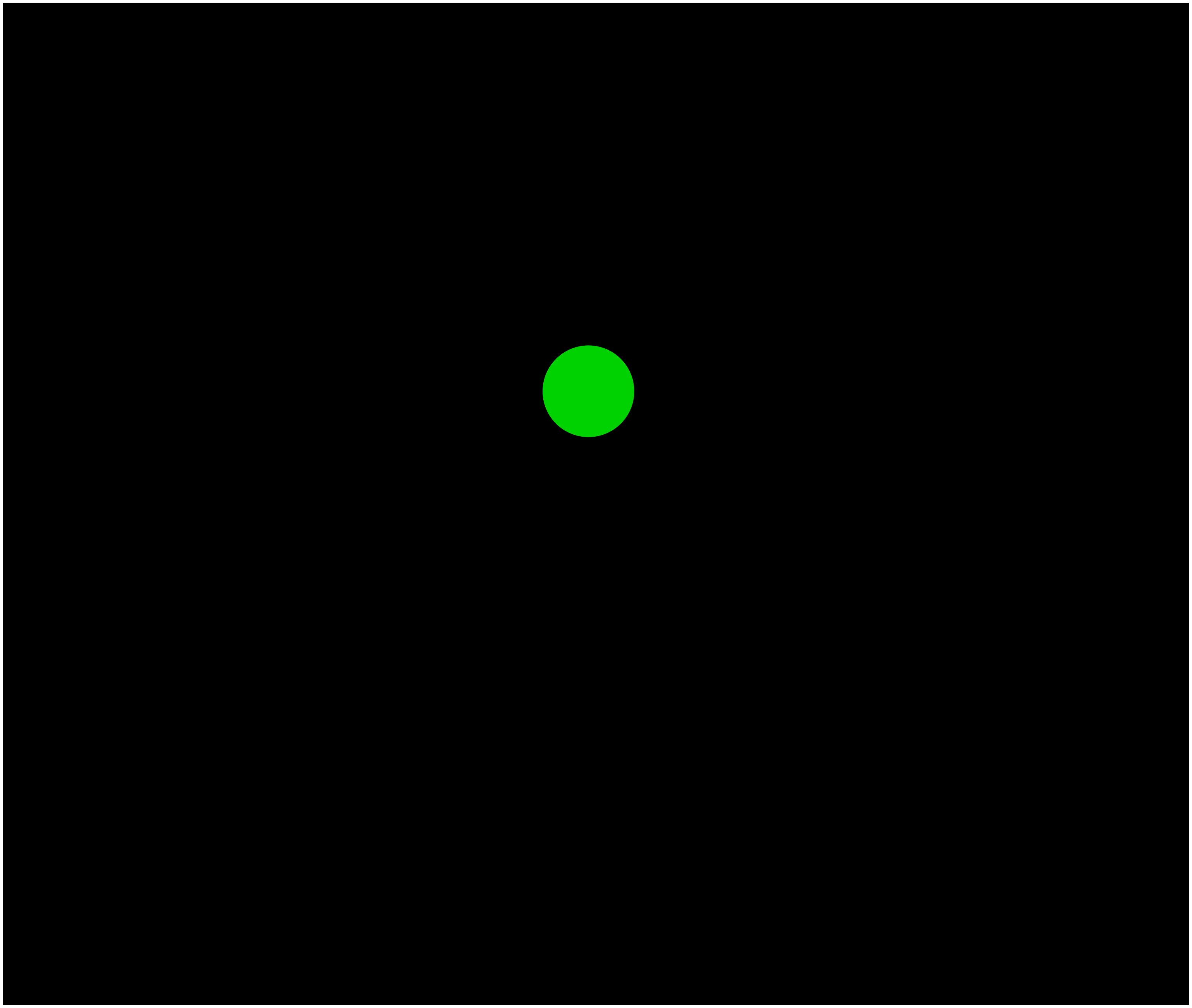
Miikkulainen et al. (2005); Weliky et al. (1995)







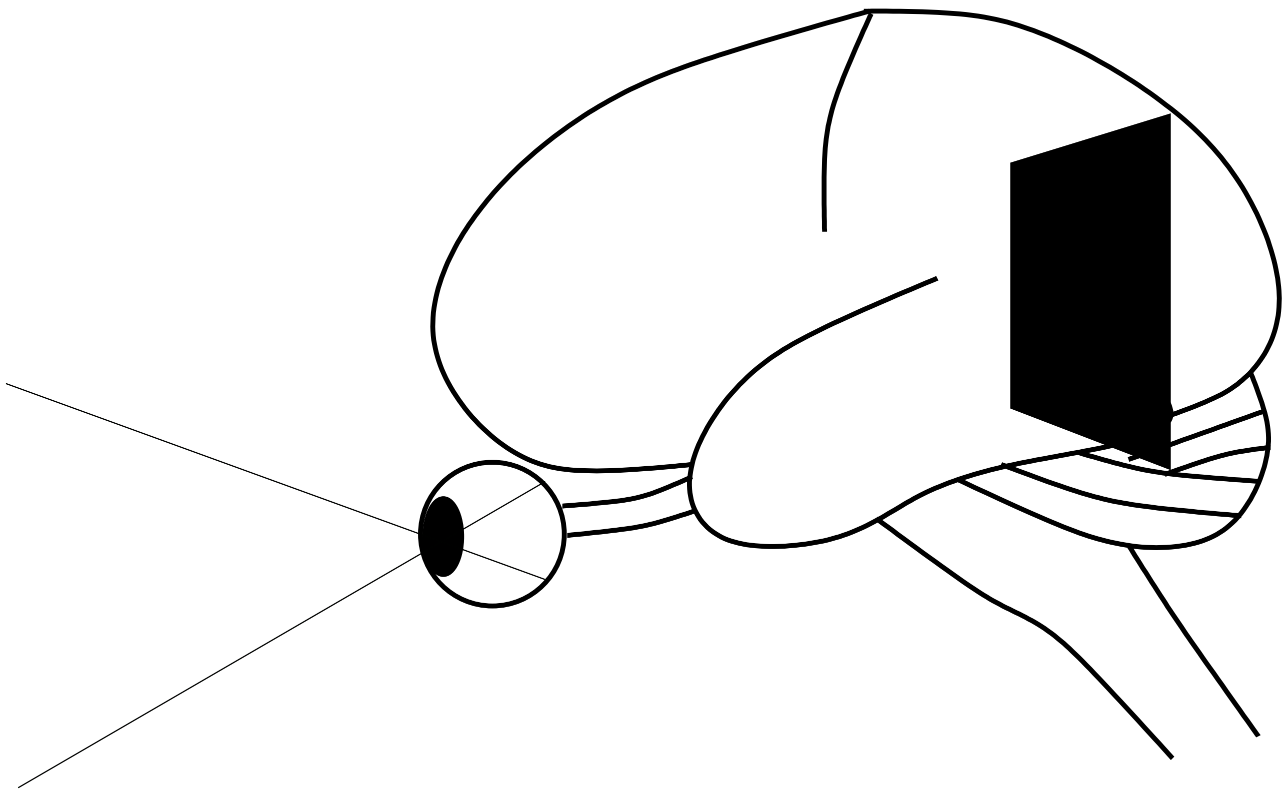




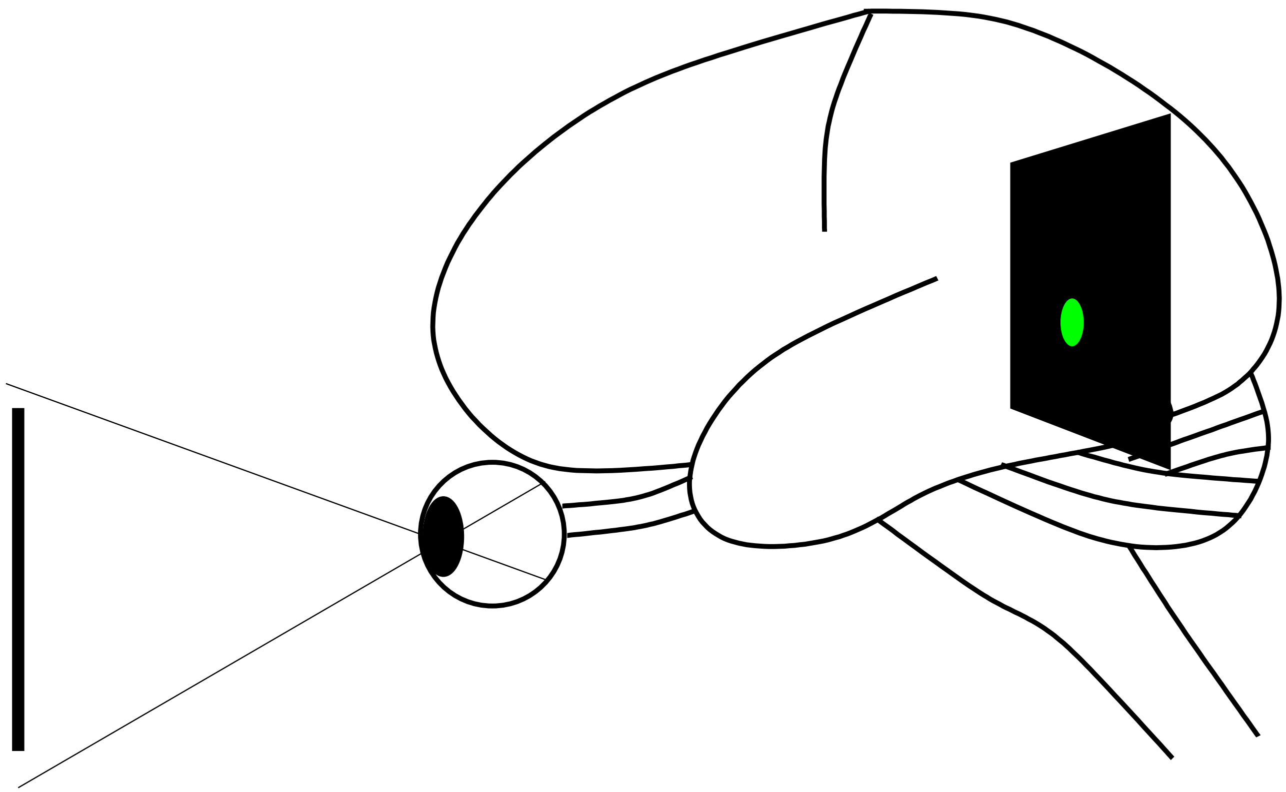
What Do Those Green Lights Represent?

- It is hard to get any idea at all.
- Actually, this is how it might be like looking at the **brain's activity from the inside** of the brain.

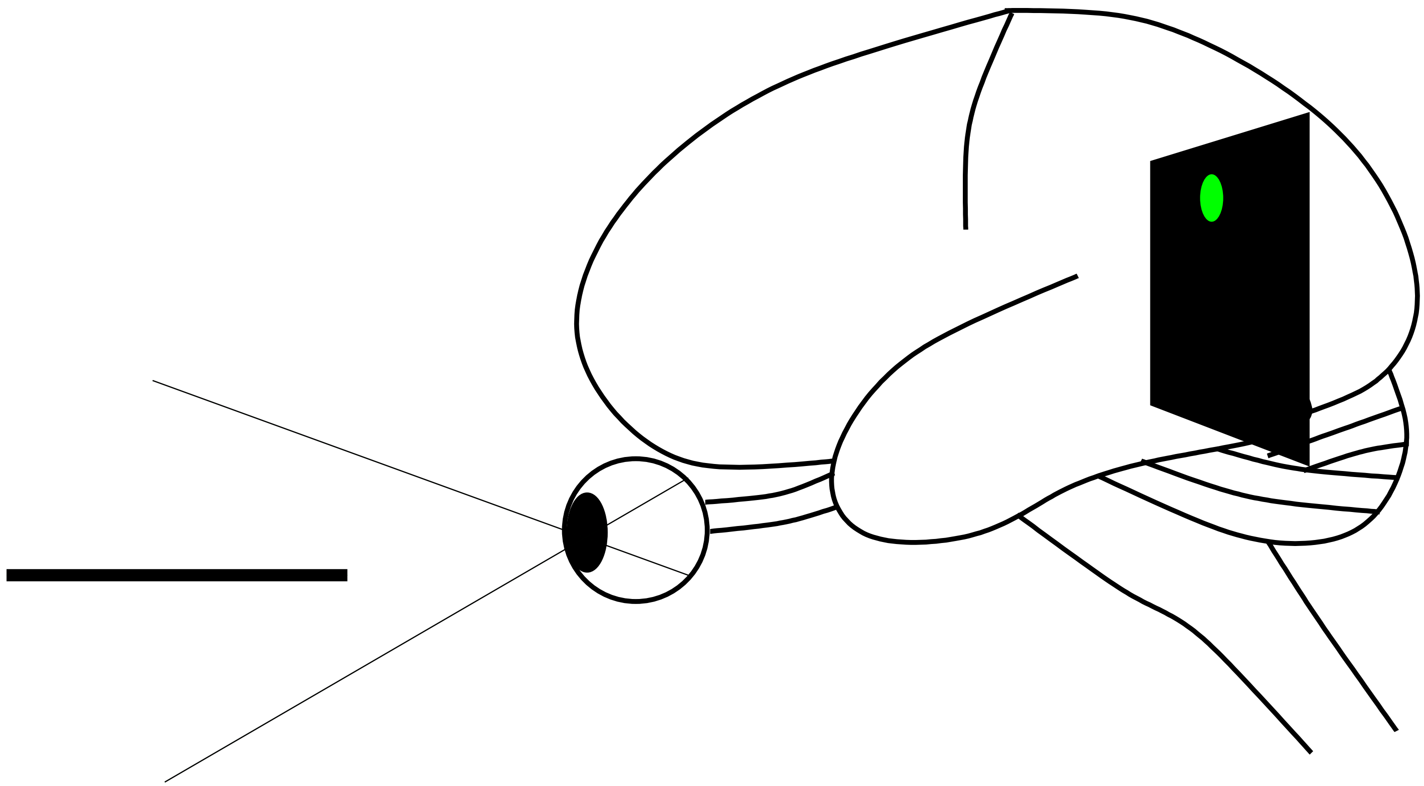
They Are Visual Cortical Responses to Oriented Lines



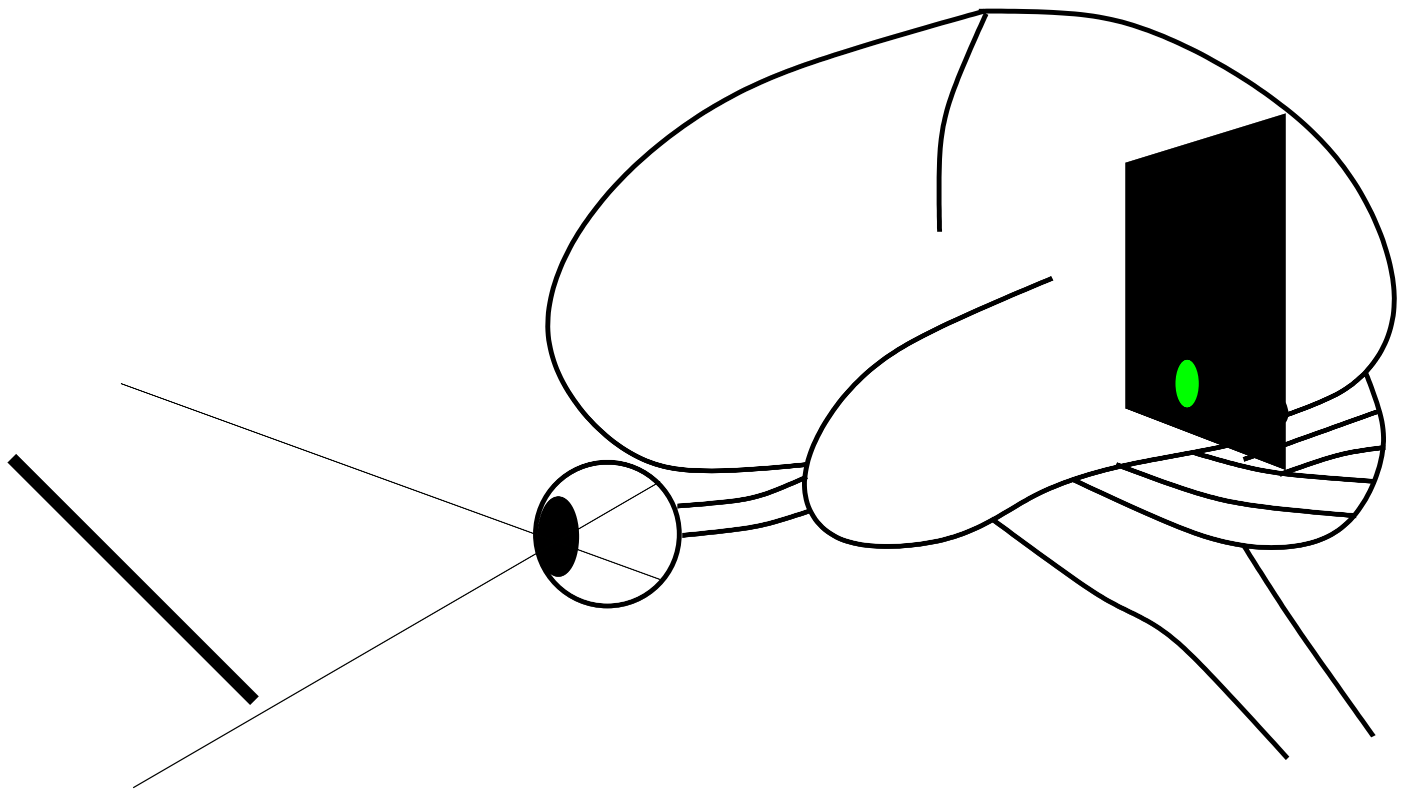
They Are Visual Cortical Responses to Oriented Lines



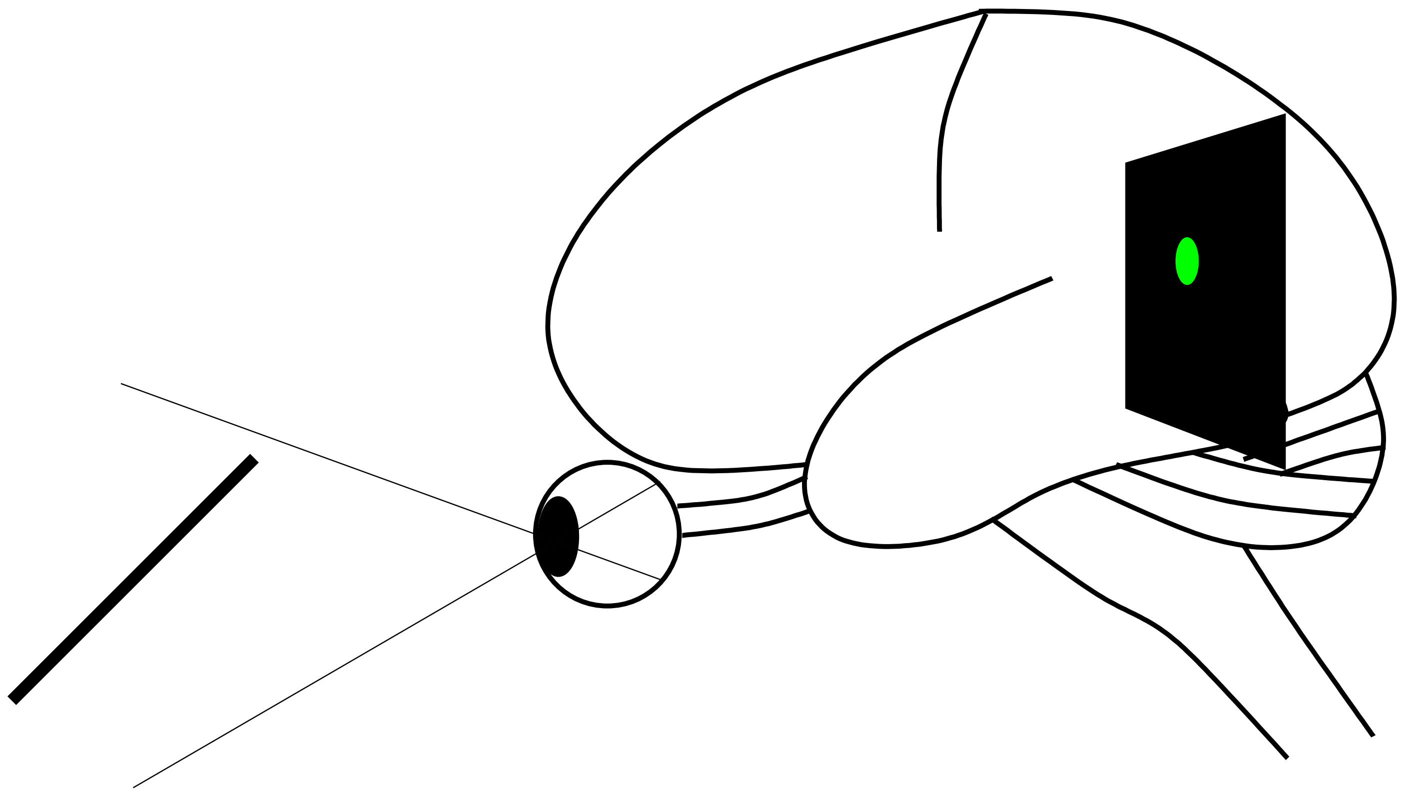
They Are Visual Cortical Responses to Oriented Lines



They Are Visual Cortical Responses to Oriented Lines



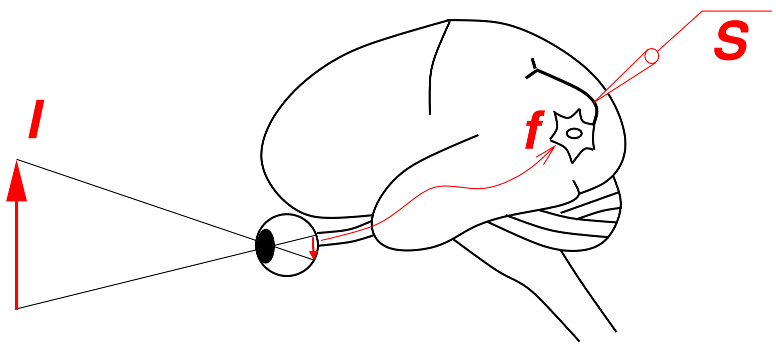
They Are Visual Cortical Responses to Oriented Lines



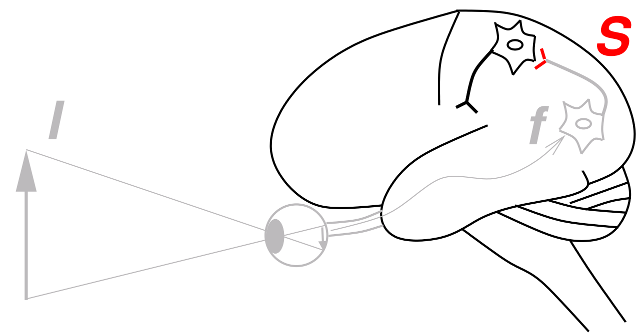
A Clearer Picture Emerges, or Does It?

- By having access to the external input, we get a much better understanding of the nature of those green lights.
- However, this poses a dillema: **How can the rest of the brain understand the green lights without access to those inputs?**.

Grounding in the Brain



(a) External observer

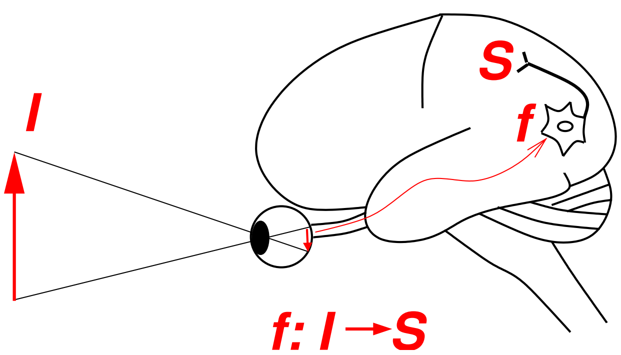


(b) Internal observer

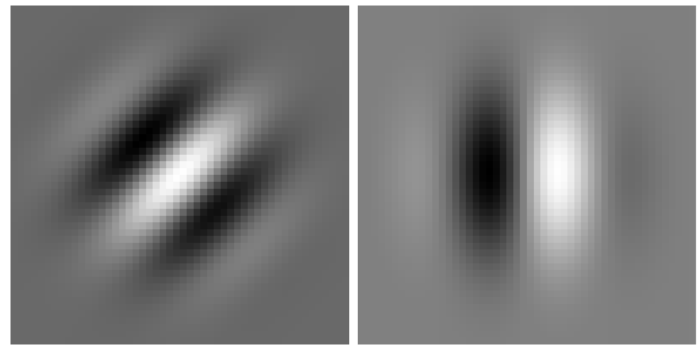
The problem of grounding, **within** the brain:

- **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



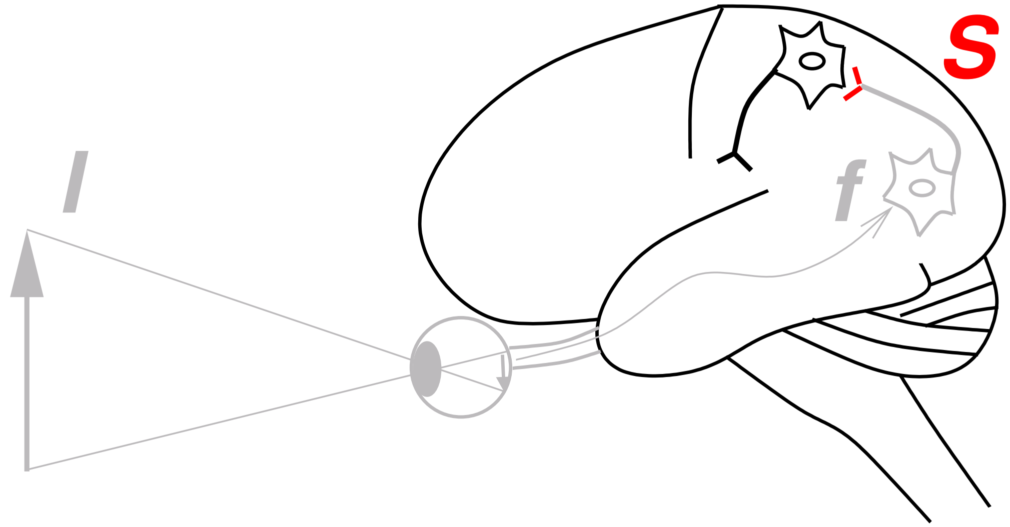
V1 Response to Input



Gabor-like RFs

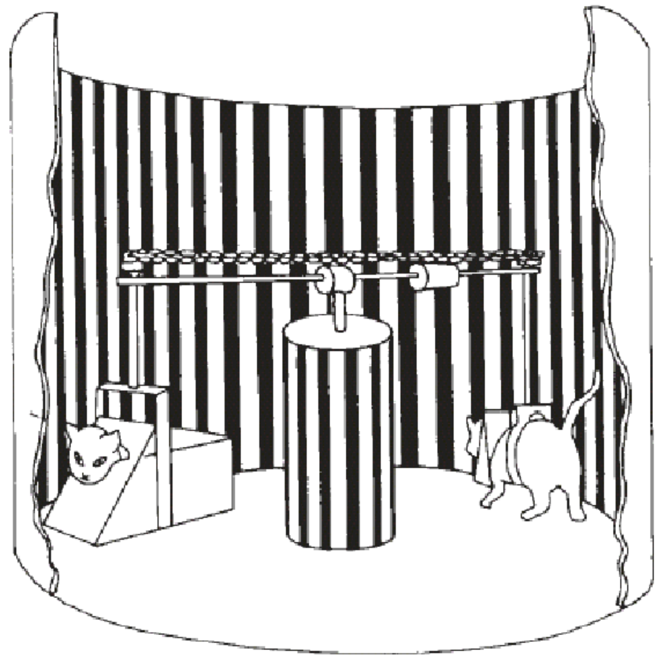
- With access to both I and S, Hubel and Wiesel (1959) figured out $f : I \rightarrow S$ in V1 (oriented Gabor-like receptive fields Jones and Palmer 1987).
- But even before that, and with access to only S, humans had no problem perceiving orientation.

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**?

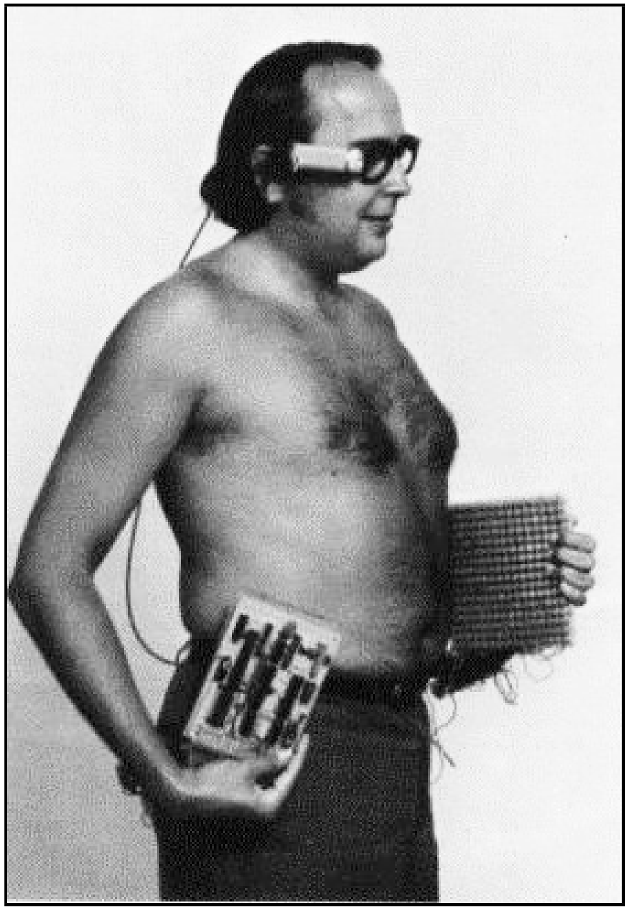
Experimental Evidence



Held and Hein (1963)

- Active animal developed normal vision.
- Passive animal did not.
- Suggests the importance of action in vision.

Experimental Evidence



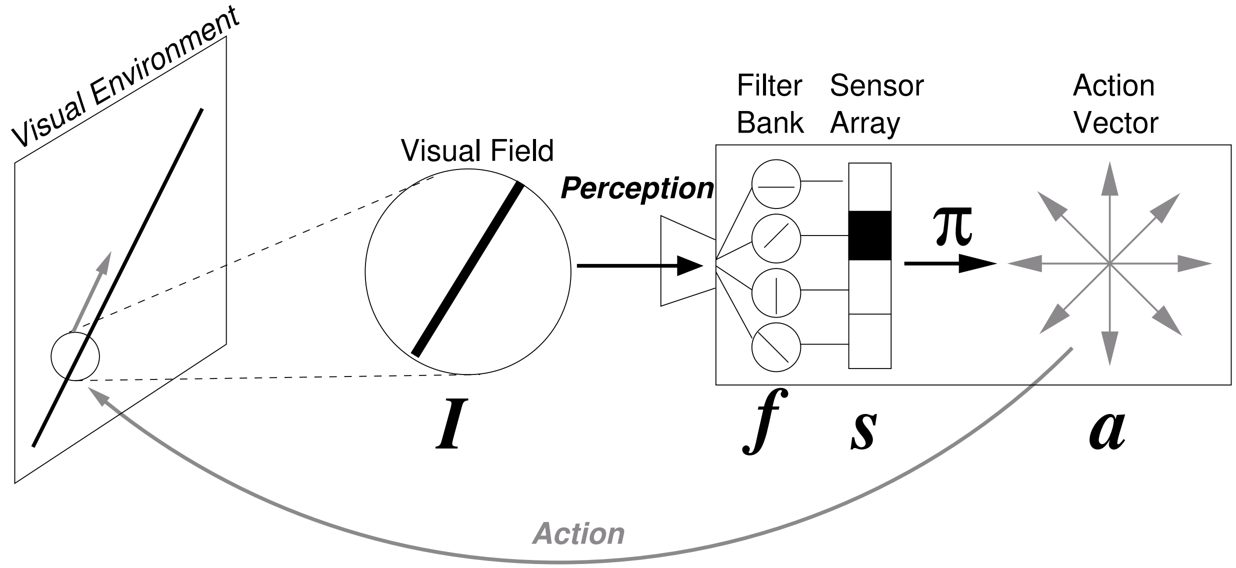
Bach y Rita (1972; 1983)

- 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!

Theoretical Insights

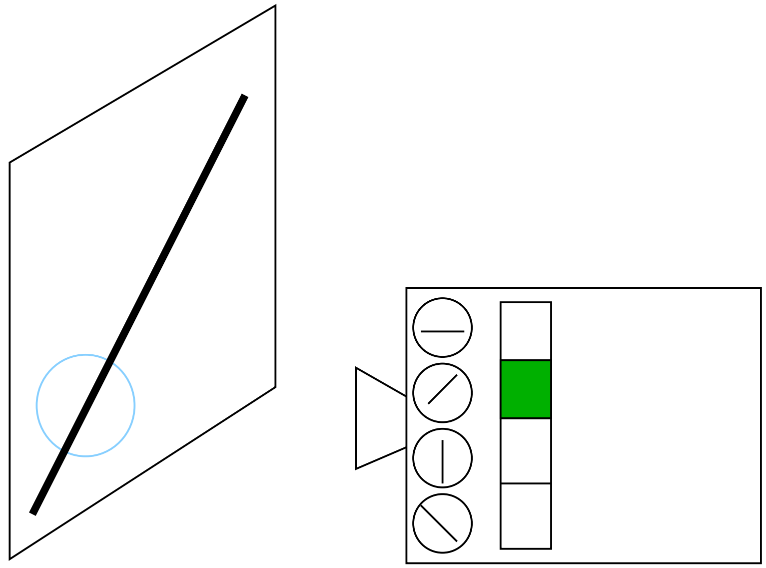
- Philipona et al. (2003) showed that properties of ambient space (such as the dimensionality) can be inferred based on internal sensory input alone.
- The key concept is about the compensability between ego-motion and the change in the environmental input conveyed to exteroceptors.

Approach: Grounding Through Action



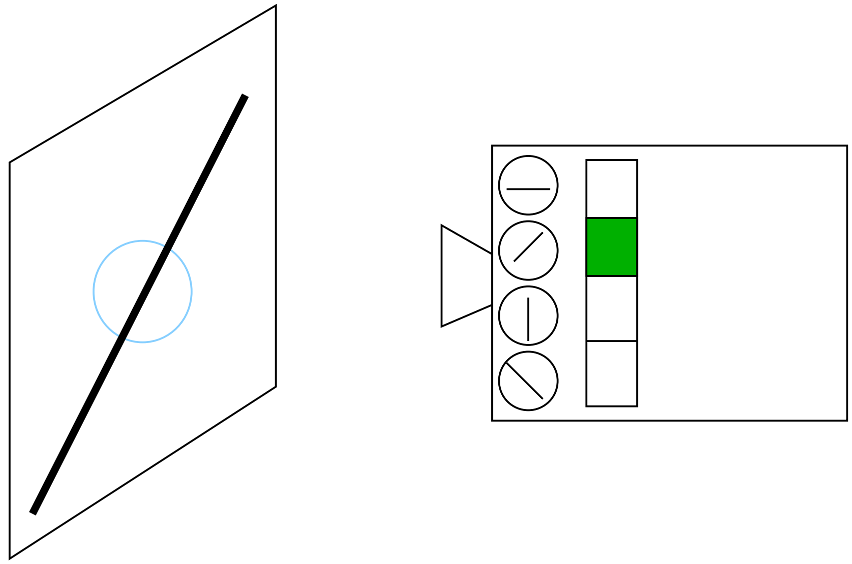
- Direct access to **encoded internal state** (sensory array) only.
- Action is enabled, which can **move the gaze**.
- How does this solve the grounding problem?

Action for Unchanging Internal State



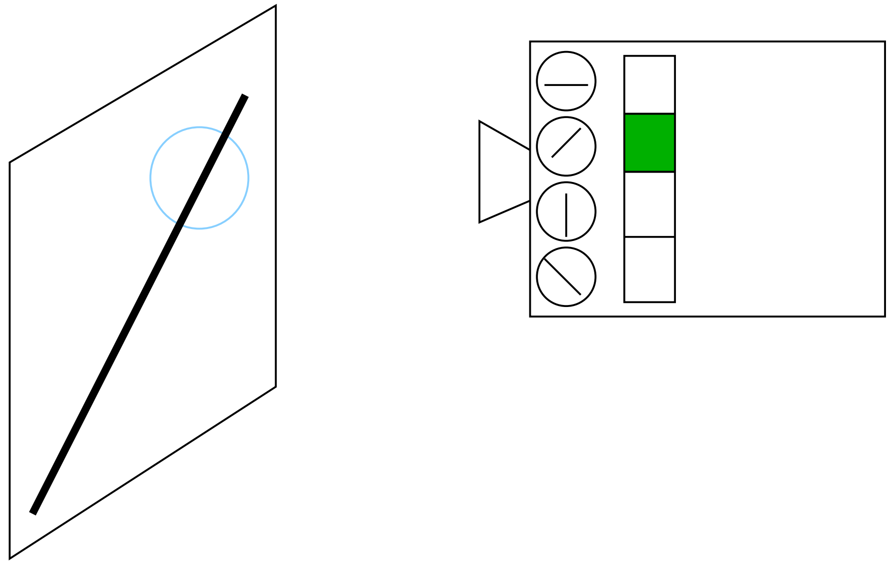
- Diagonal motion causes the *internal state* to **remain unchanging** over time.
- Property of such a movement **exactly reflects** the property of the input I: Semantics figured out through action.

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.

Action for Unchanging Internal State

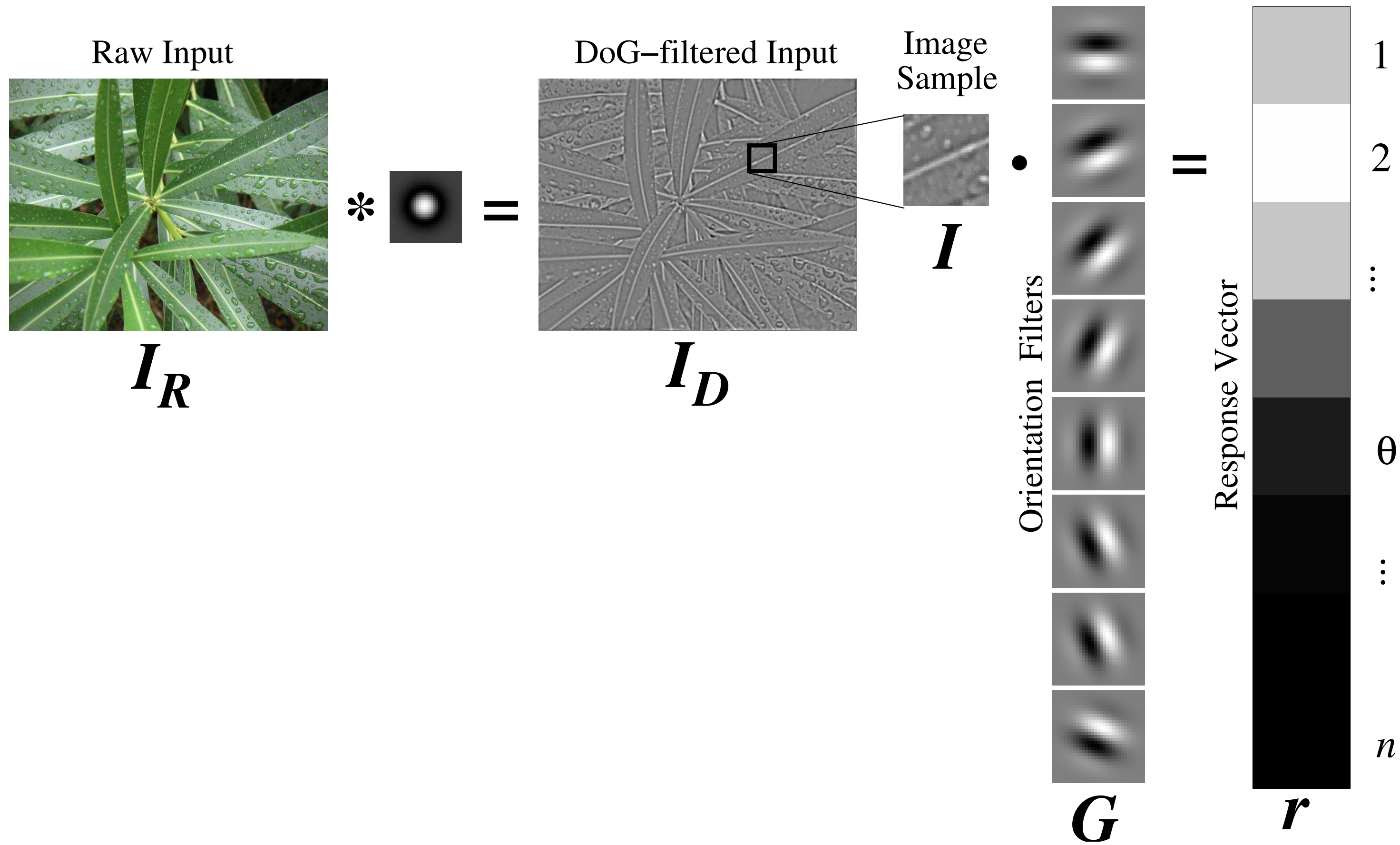


- Diagonal motion causes the *internal state* to **remain unchanging** over time.
- Property of such a movement **exactly reflects** the property of the input I: Semantics figured out through action.

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

- Policy π: Given reward probability $R(s, a) = P(a|s)$ and state s, stochastically generate action a with probability $P(a|s)$.
- Reward: measure similarity between previous and current response vector $\mathbf{r}$

$$\rho_{t+1} = \mathbf{r}_t \cdot \mathbf{r}_{t+1}$$

- Learning $R(s, a)$:

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

and then normalize over all actions for a given state.

Reward Probability Table

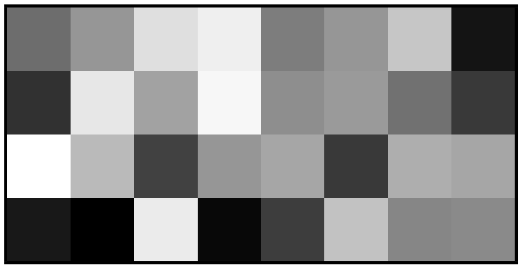
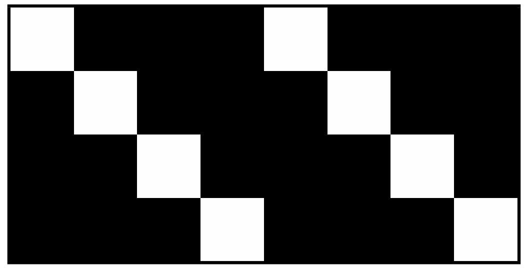
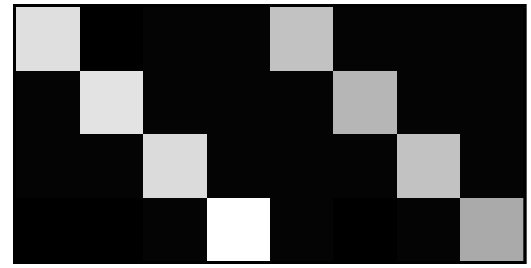
A: direction of motion

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

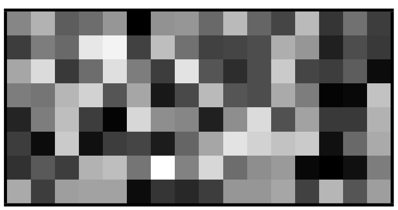
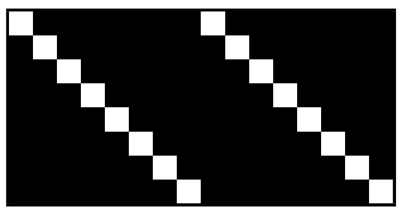
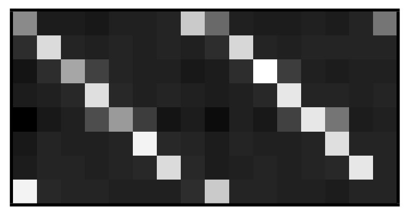
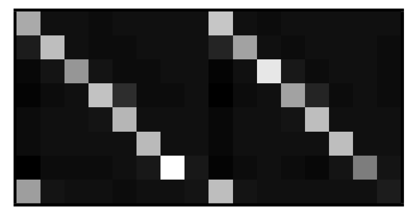
- Reward probability $R(s, a)$ can be tabulated.
- In an ideal case (world consists of straight lines only), we expect to see two diagonal matrices (shaded gray, above).

Results: Learned $R(s, a)$

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

Synthetic image

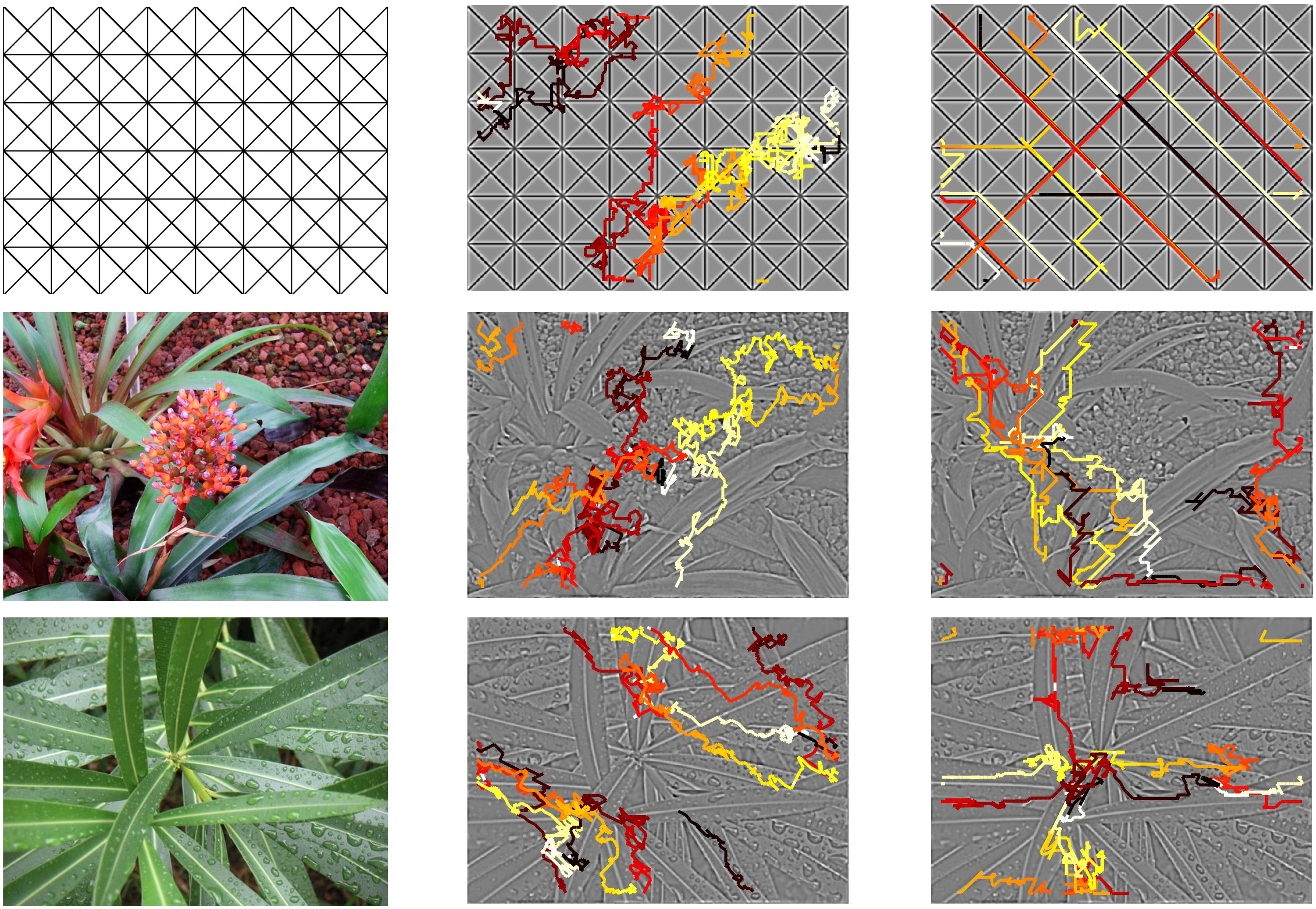





(a) Initial (b) Ideal (c) Plant (d) Oleander

Natural images

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

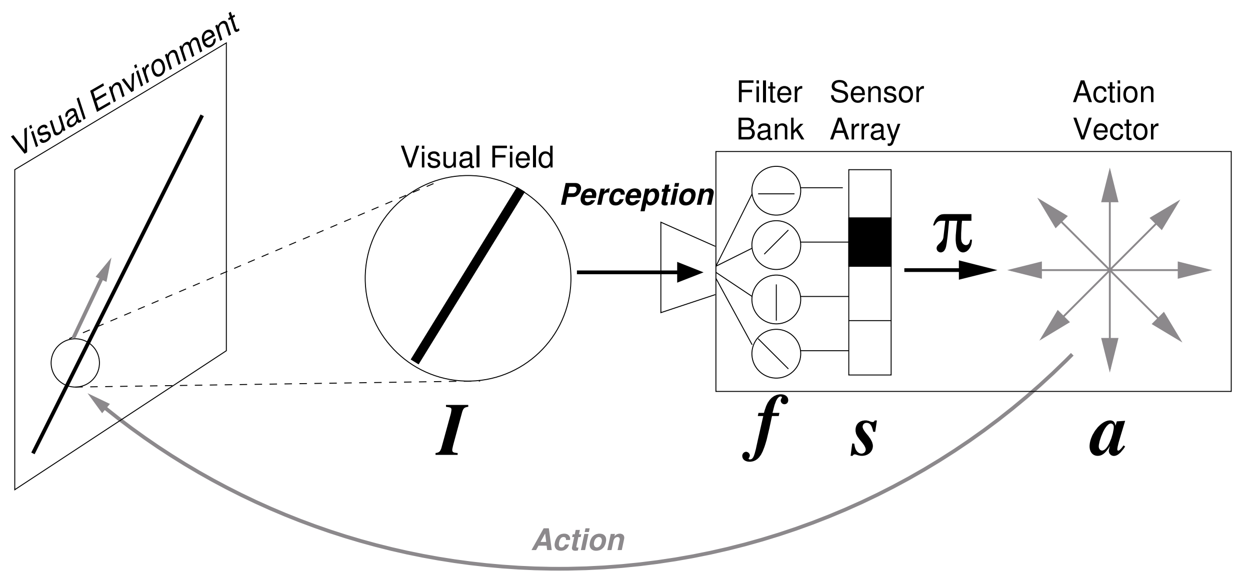
Results: Gaze Trajectory



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

Results: Demo

Part I: Summary

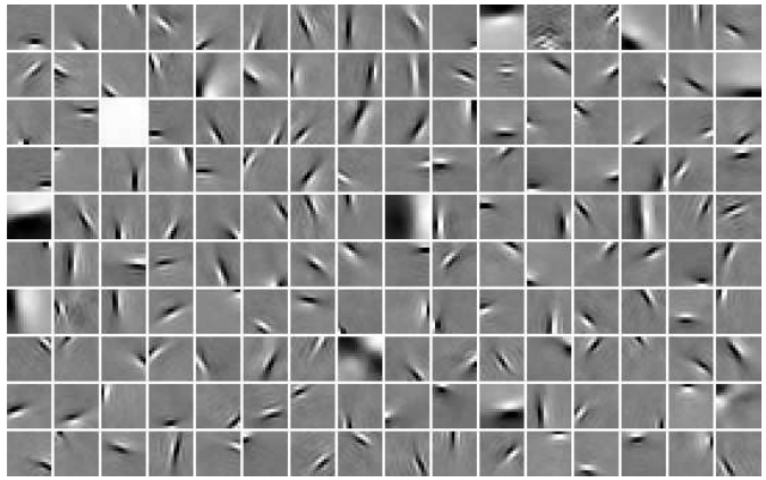


- Using **invariance** as the only criterion, particular **action pattern** that has the **same property** as the input that triggered the sensors was learned.

Part II: Receptive Field Learning

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 normalized Hebbian learning:

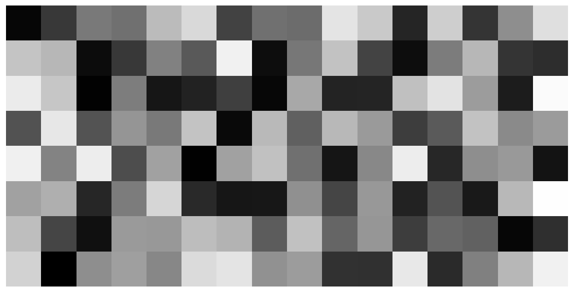
$$g_{ij} = \frac{g_{ij} + \alpha(I_{ij} - g_{ij})}{\sum_{mn} g_{mn} + \alpha(I_{mn} - g_{mn})},$$

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

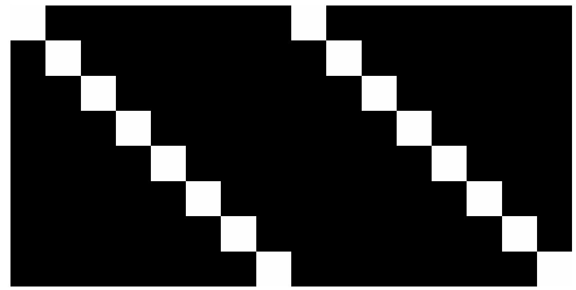
Experiments

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

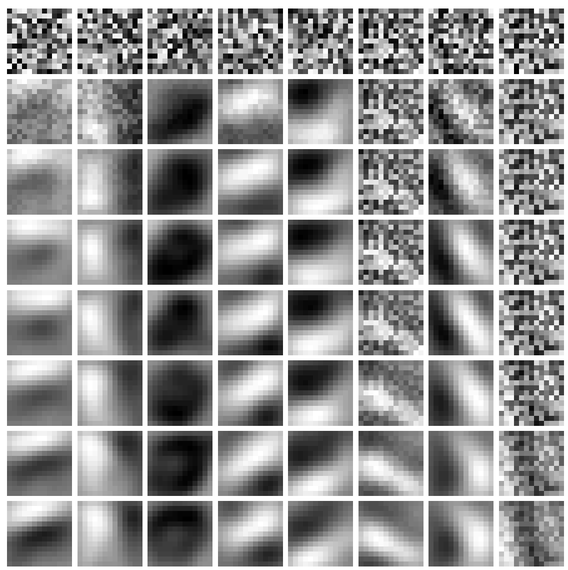
Effects of $R(s,a)$ on RF Learning



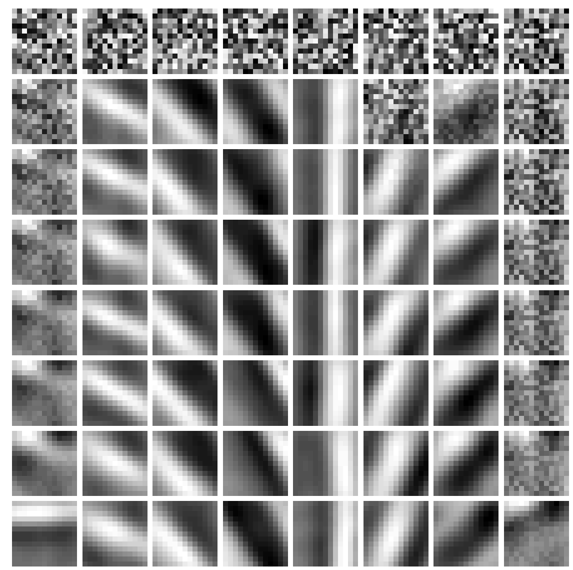
Fixed Random R



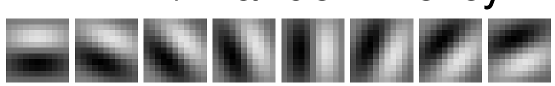
Fixed Ideal R



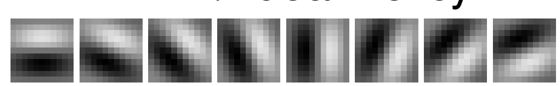
RF w/ Random Policy



RF w/ Ideal Policy

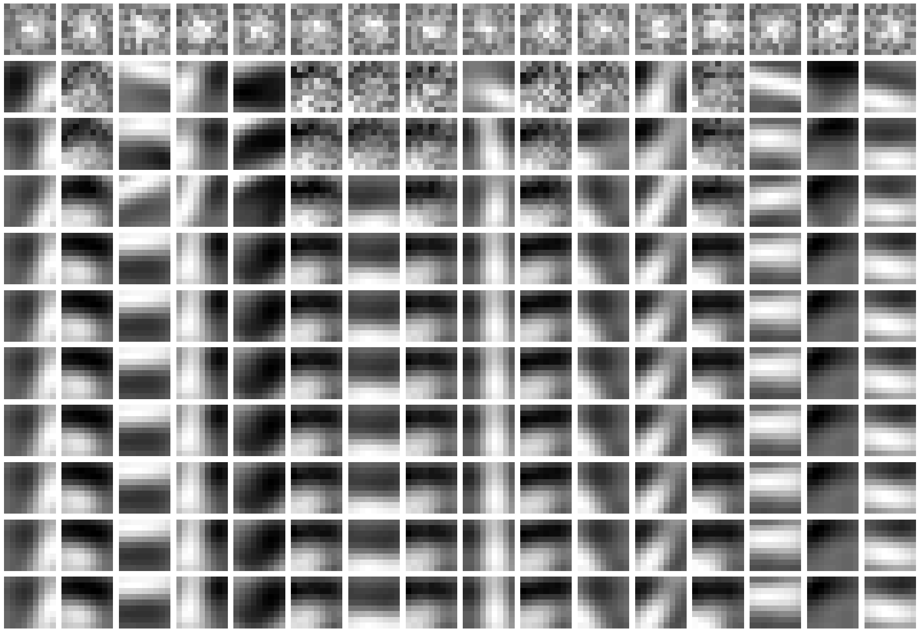


Reference RFs

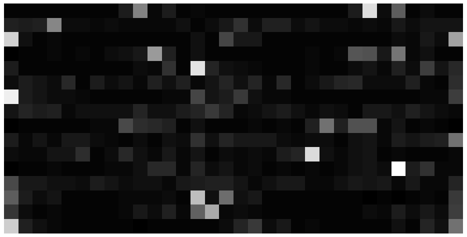


Reference RFs

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



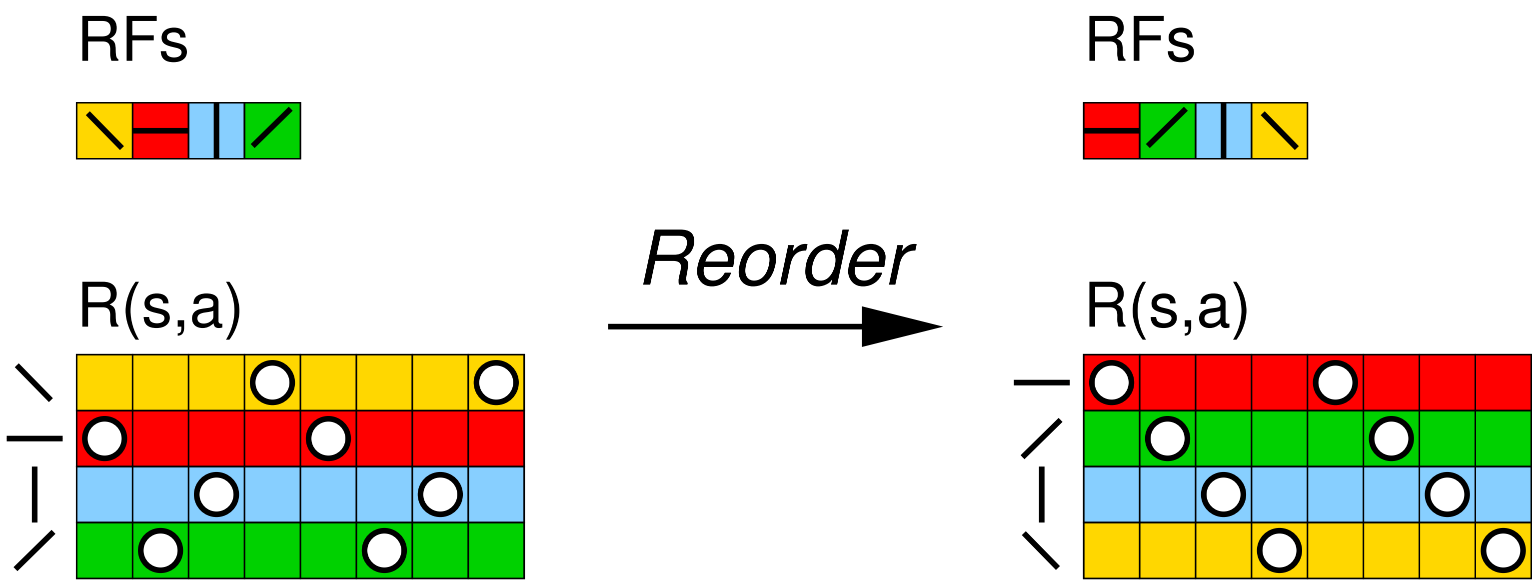
Learned RFs



Learned $R(s,a)$

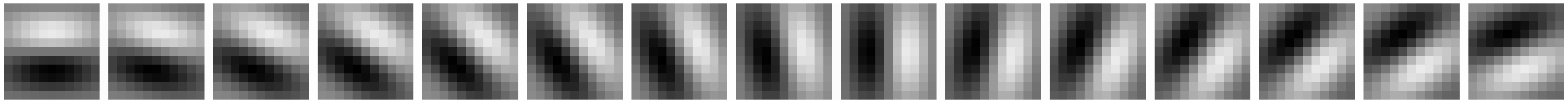
- Seemingly unordered RFs and $R(s,a)$ results.

Reordering RFs

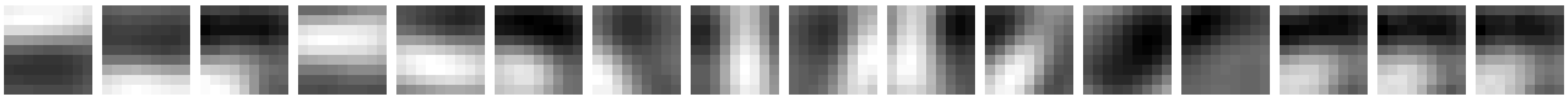


- The $R(s, a)$ result looks bad because each row's corresponding RF orientation is not ordered.
- Reordering RF orientation reorders the rows in $R(s, a)$.

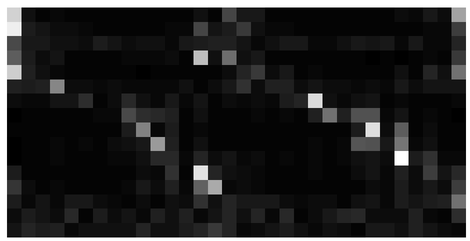
Reordered RFs and $R(s, a)$



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!)

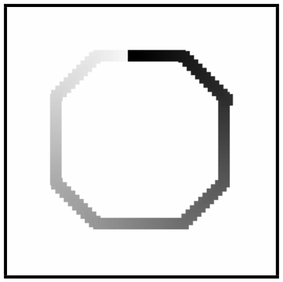
Part II: Summary

- 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.

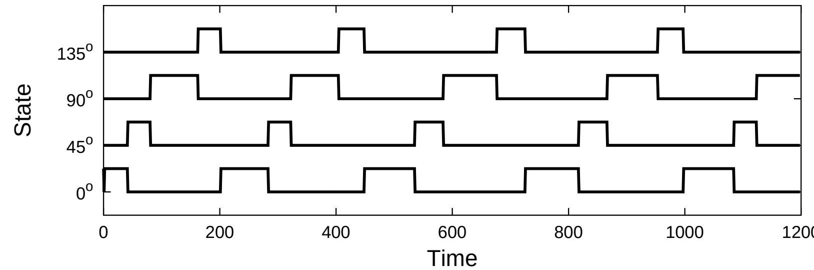
Part III: Shape Recognition

Misra and Choe (2007)

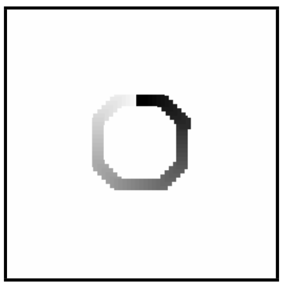
Learning About Shapes



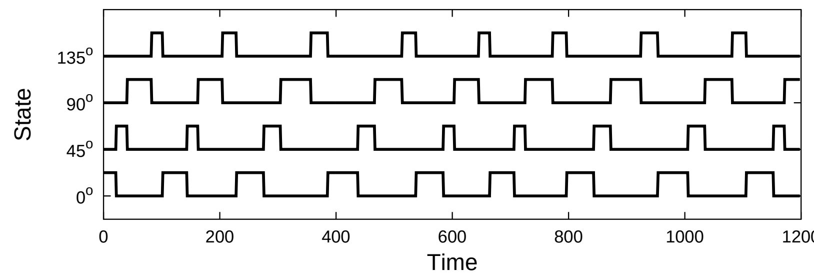
(a) Eye position (large input)



(b) Internal state (large input)



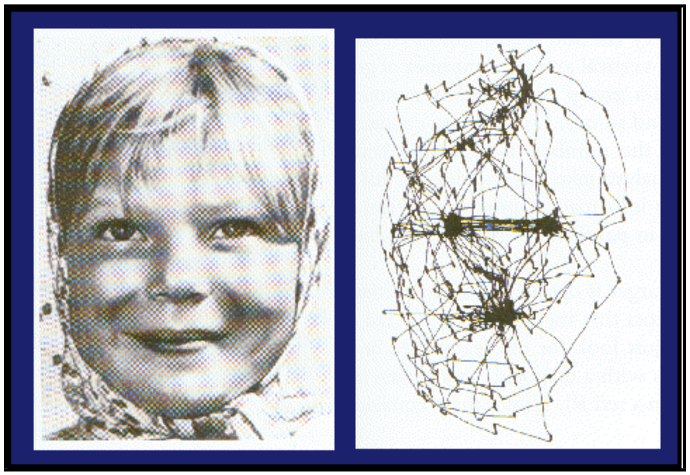
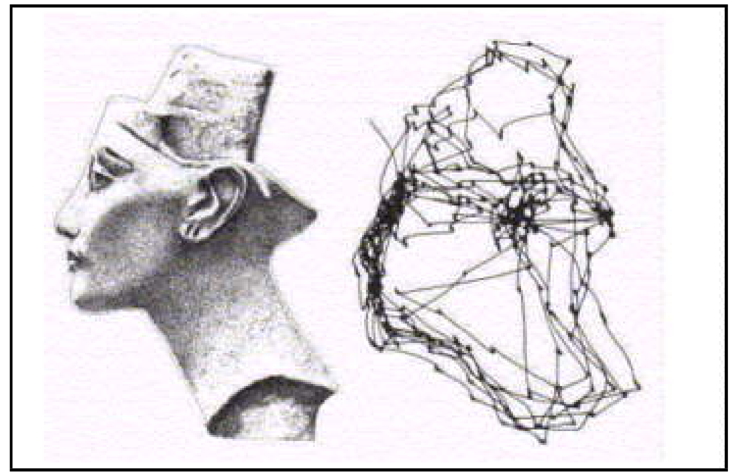
(c) Eye position (small input)



(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.

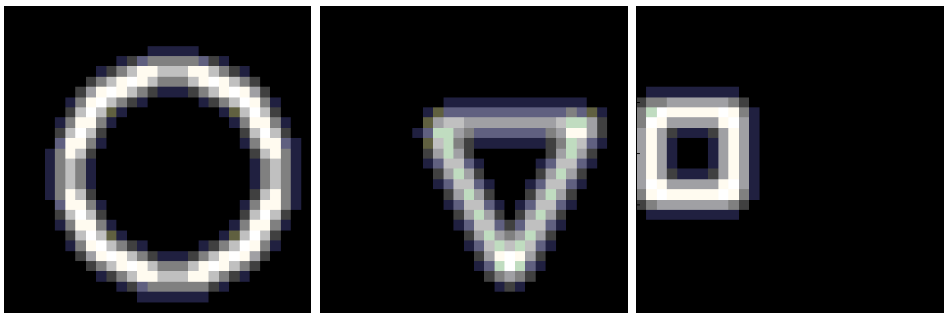
Motor System and Object Recognition

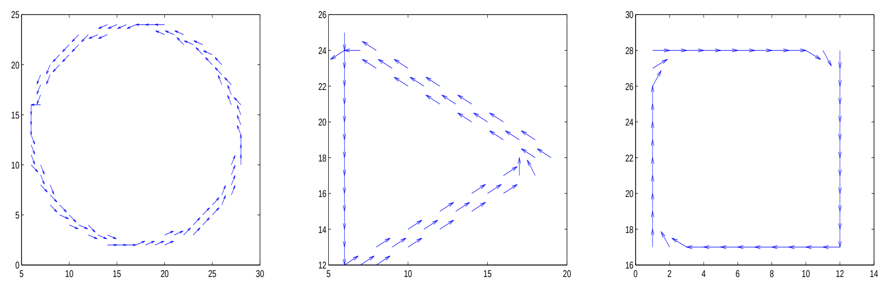
Yarbus (1967)

- 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)



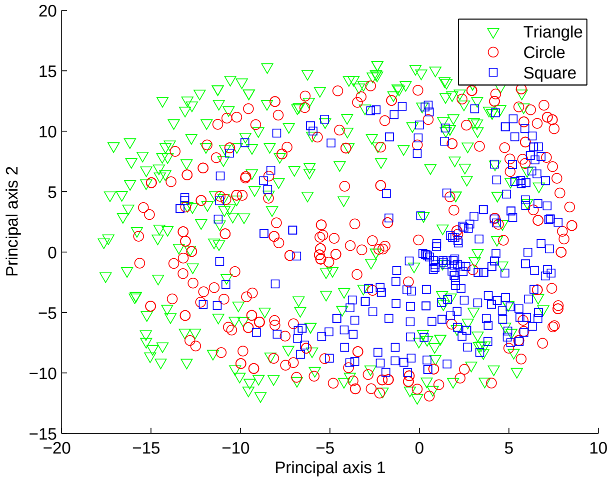
(a) Sensor-based Representation



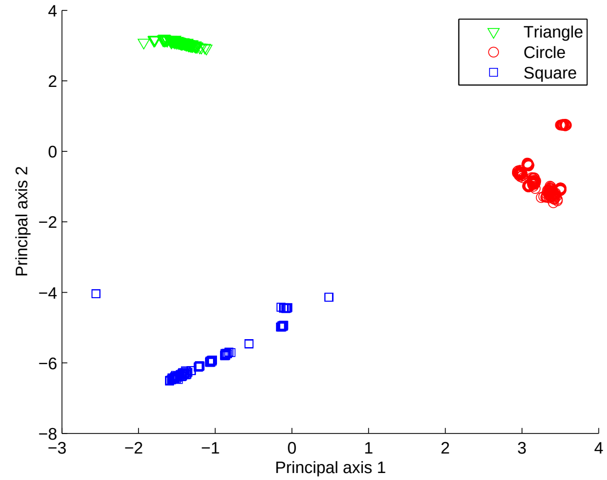
(b) Motor-based Representation

- 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



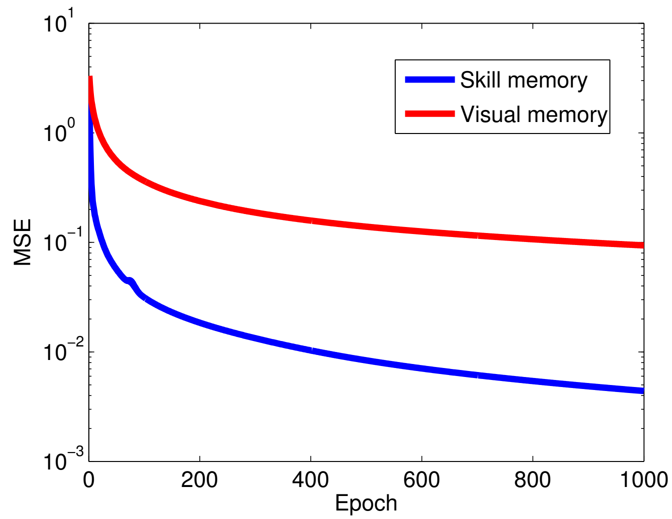
(a) Visual Memory



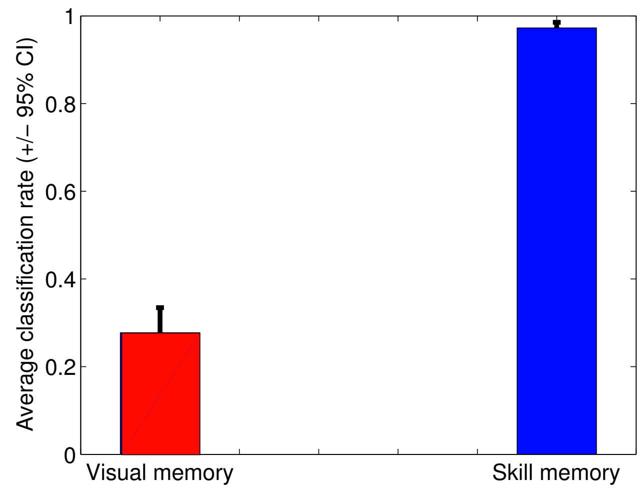
(b) Motor Memory

- 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).

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.

Related Works (Selected)

- Pierce and Kuipers (1997): Learning from raw sensor/actuators (See related work on bootstrap learning).
- Miikkulainen et al. (2005): Visual cortical development and function
- Ballard (1991): Animate vision
- Rizzolatti et al. (2001): Mirror neurons
- Salinas (2006): Sensory RF coding dictated by downstream requirements.
- Sejnowski (2006): Importance of “projective fields”.

Discussion

- Why is **knowing** ones own **action** any easier than **perceptual interpretation**?: Knowledge of own action may be more **immediate** than perception (cf. Moore 1996, citing Bergson).
- What gives rise to **voluntary, intentional action** and why is it special? (Freeman 1999; Kozma and Freeman 2003; Taylor 1999).
- A different view of **invariance**: Not (only) something to be detected in the environment (cf. Gibson 1950), but something that we **actively seek within**.

Discussion (Cont'd)

- Why not just **analyze the input directly?**: The raw input is only available at the immediate sensory surface.
- What about **other sensory modalities** (such as touch, olfaction, or audition)?
- The learning scheme **depends** on **structure** in the environment: If the environment didn't have structure, the agent can never learn.

Discussion (Cont'd)

- Relation to **mirror neurons** (Rizzolatti et al. 2001)?
- Role of **attention** (e.g. Rensink et al. 1997; Taylor 1999)?: Attention may be needed when ambiguities are present.
- Do **motor primitives** restrict the kind of sensory property that can be learned? What kinds of motor primitive do we have?

Discussion (Cont'd)

- What about meaning **other than sensorimotor-like**, such as reinforcement signals (Rolls 2001) or "feeling" (Harnad 2001)?
- **Grounding on perception alone** may not be sufficient: cf. Perceptual symbol system (Barsalou et al. 2003).
- What to make of the segregation in the **dorsal–ventral pathway?**

  (Goodale and Milner 1992).

Predictions

- 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

We must ask how the brain understands itself.

- Action is important for understanding/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.

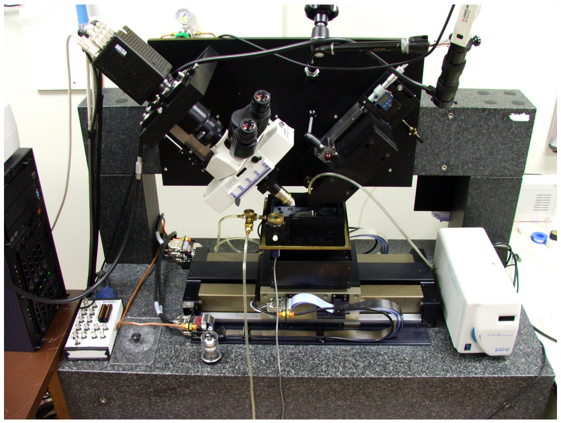
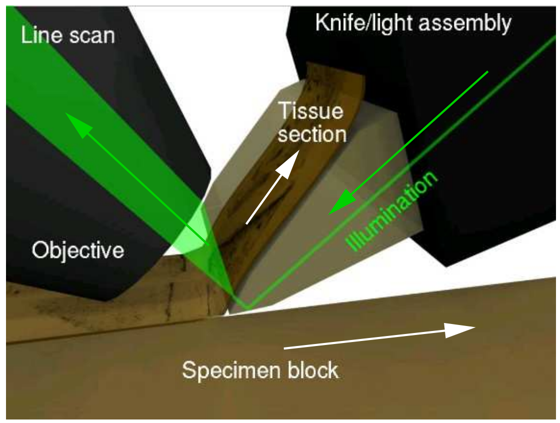
Credits

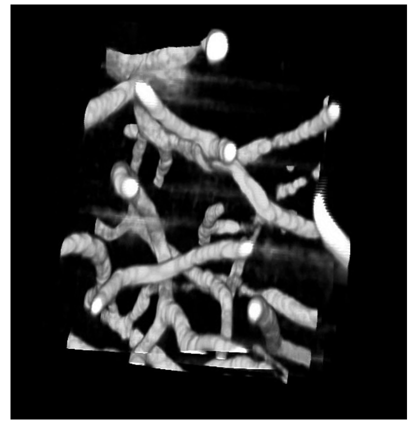
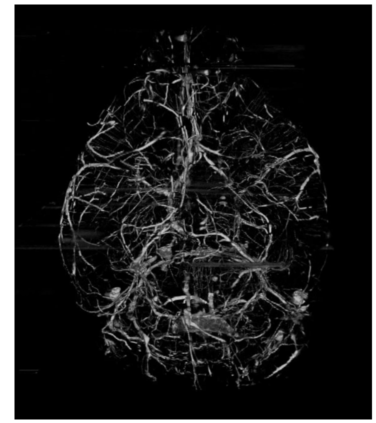
- Contributors: Kuncara A. Suksadadi, S. Kumar Bhamidipati, Noah Smith, Stu Heinrich, Navendu Misra, Huei-Fang Yang, Daniel C.-Y. Eng

Other Projects at Texas A&M

- Knife-Edge Scanning Microscope (KESM) Project
- Flash-lag effect, delay compensation, and facilitating synapses
- Evolutionary precursor of agency: internal state predictability
- And more ...

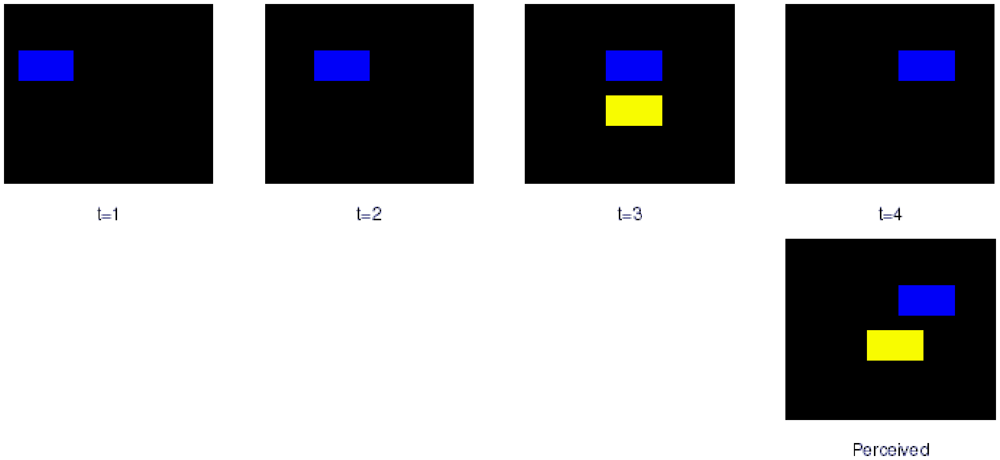
Knife-Edge Scanning Microscope Project

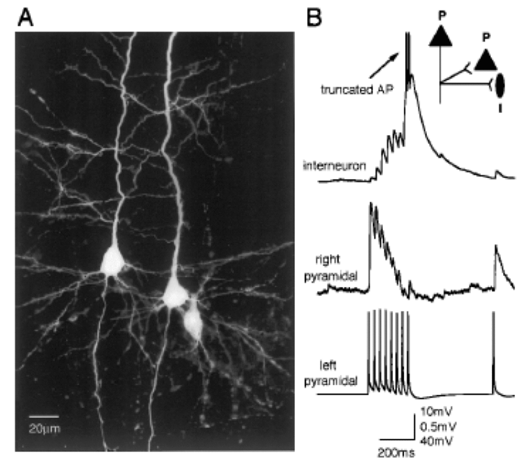



- Cut and image whole mouse brain at sub-micrometer resolution.
- Fully automated: one mouse brain imaged in less than 2 weeks.
- Resulting data: 2 to 20 TB per mouse brain.
- Analysis of the data is a major issue.

FLE, Delay Compensation, & Facilitating Synapses

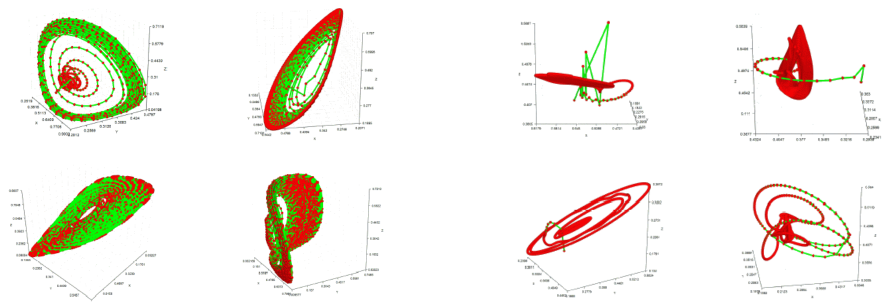


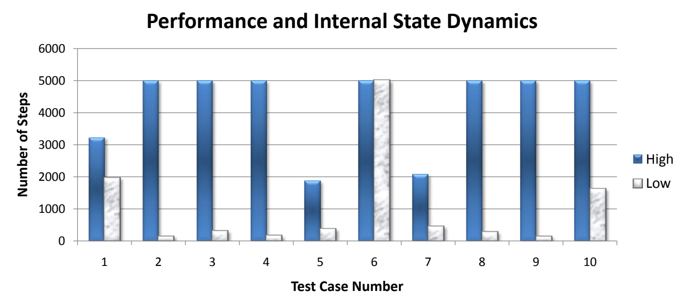


Lim and Choe (2008, 2005, 2006)

- Delay in the nervous system on the order of 100 ms.
- Flash-lag effects suggest a compensatory mechanism.
- Facilitating synapses may be the neural substrate.

Evolutionary Precursor of Agency





Kwon and Choe (2008)

- Agency $>$ authorship $>$ 100% predictability of own action.
- For this, internal state trajectory must be predictable.
- Same task performance but more predictable internal state trajectory have an advantage when the task becomes more difficult.

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.

Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48:57–86.

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.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, 7:84–91.

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

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. http://dx.doi.org/10.1142/S0219843607001102.

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*.

Freeman, W. J. (1999). *How Brains Make Up Their Minds*. London, UK: Wiedenfeld and Nicolson Ltd. Reprinted by Columbia University Press (2001).

Gibson, J. J. (1950). *The Perception of the Visual World*. Boston: Houghton Mifflin.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15:20–25.

Harnad, S. (2001). TTT guarantees only grounding: But Meaning = Grounding + Feeling. *Think*, 12(045).

Held, R., and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56:872–876.

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.

Kozma, R., and Freeman, W. J. (2003). Basic principles of the KIV model and its application to the navigation problem. *Journal of Integrative Neuroscience*, 2:125–145.

Kwon, J., and Choe, Y. (2008). Internal state predictability as an evolutionary precursor of self-awareness and agency. In *Proceedings of the Seventh International Conference on Development and Learning*.

Lim, H., and Choe, Y. (2005). Facilitatory neural activity compensating for neural delays as a potential cause of the flash-lag effect. In *Proceedings of the International Joint Conference on Neural Networks*, 268–273. Piscataway, NJ: IEEE Press.

Lim, H., and Choe, Y. (2006). Facilitating neural dynamics for delay compensation and prediction in evolutionary neural networks. In Keijzer, M., editor, *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation, GECCO-2006*, 167–174. Nominated for Best Paper Award in the Artificial Life/Evolutionary Robotics/Adaptive Behavior track.

Lim, H., and Choe, Y. (2008). Delay compensation through facilitating synapses and its relation to the flash-lag effect. *IEEE Transactions on Neural Networks*. In press.

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

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.

Moore, F. C. T. (1996). *Bergson: Thinking Backwards*. Cambridge, UK: Cambridge University Press.

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.

Philipona, D., O'Regan, J. K., and Nadal, J.-P. (2003). Is there something out there? Inferring space from sensorimotor dependencies. *Neural Computation*, 15:2029–2050.

Pierce, D. M., and Kuipers, B. J. (1997). Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:162–227.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8:368–373.

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

Rolls, E. T. (2001). Representation in the brain. *Synthese*, 129:153–171.

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, J. G. (1999). *The Race for Consciousness*. Cambridge, MA: MIT Press.

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*. [Online] https://umdrive.memphis.edu/iftekhar/public/IJCNN/BCV.htm.