Cours 2022-2023:

Quel code neural pour les représentations mentales?
Vector codes and the geometry of mental representations

Stanislas Dehaene
Chaire de Psychologie Cognitive Expérimentale

Cours n°1

Cellules grand-mère ou vecteurs neuronaux :
Les représentations mentales sont-elles localisées ou distribuées ?

Grand-mother cells or neuronal vectors:
Are mental representations localized or distributed ?

How do neurons encode our thoughts?

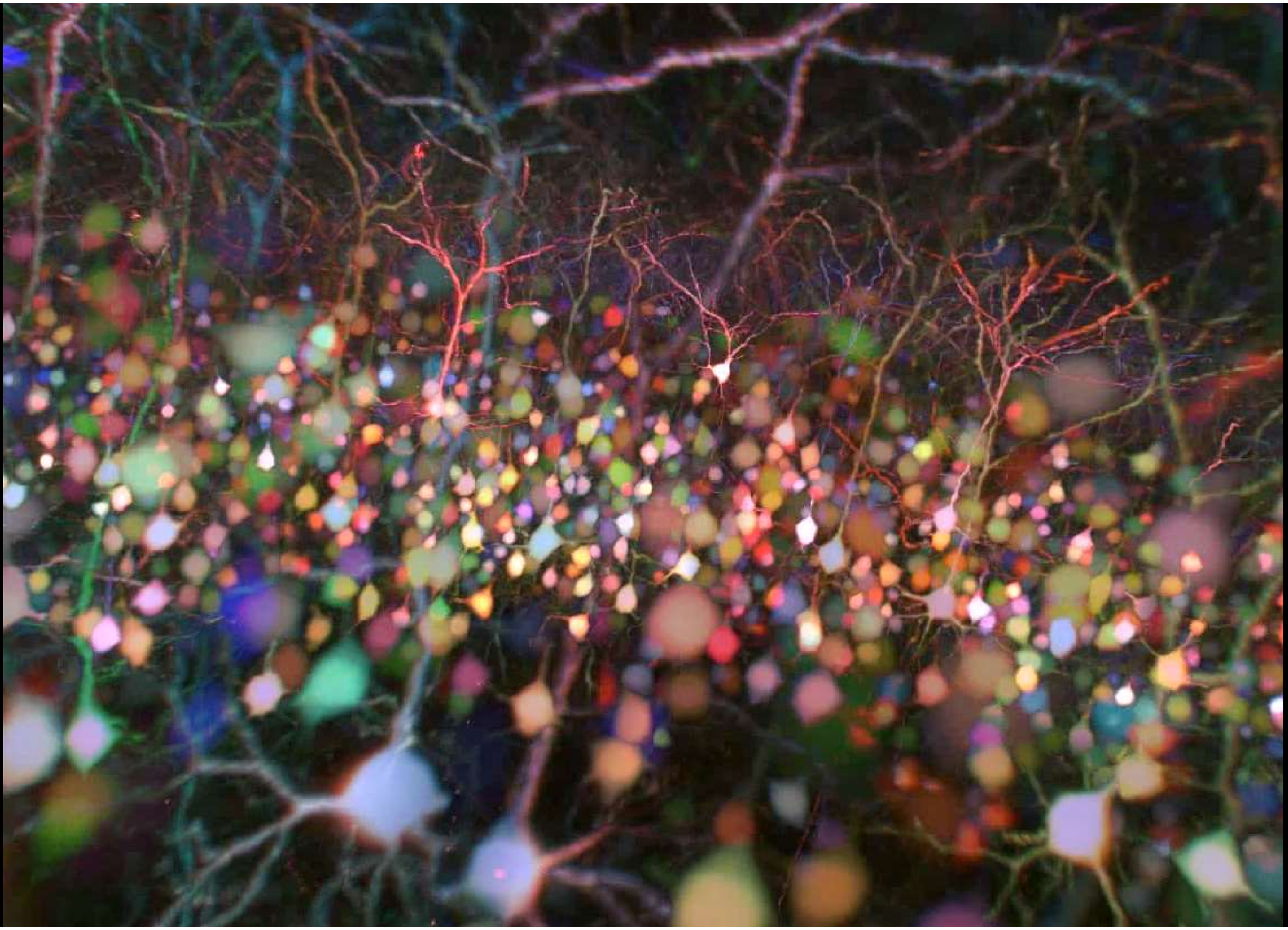
What is the code?

And if we understood it, could we decode it, or even manipulate it?

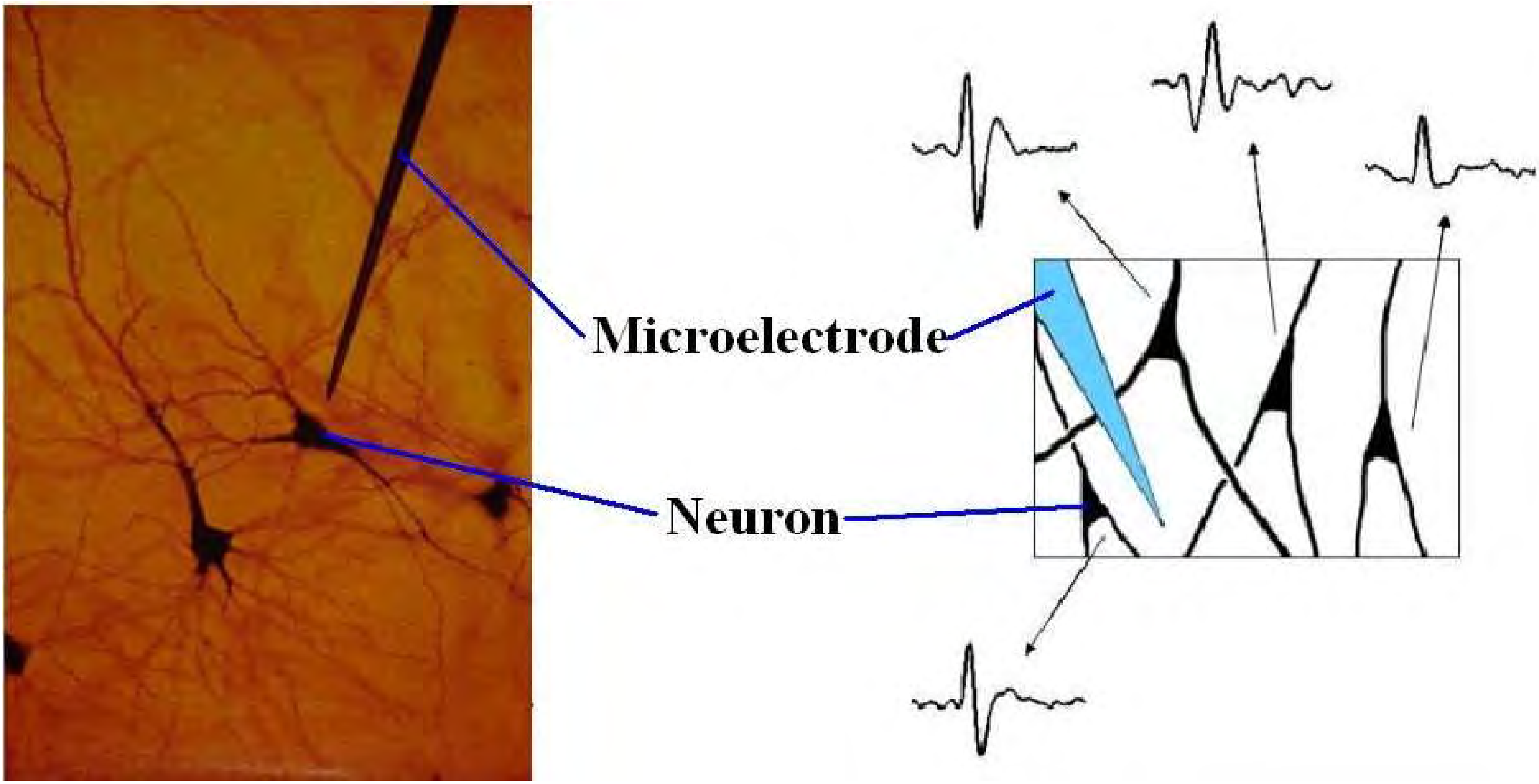
We need to identify bridging laws between psychology and neuroscience.

The bridge is unlikely to have a single arch, directly from individual concepts to individual neurons
One concept
≠ one neuron

In this course, we will explore the hypothesis that:
- Neural populations encode vector spaces
- Mental representations are points in high-dimensional space



Single neuron electrophysiology



http://newton.umsl.edu/tsytsarev_files/Lecture02.htm

Single-cell electrophysiology : the gold standard for decades of discoveries in neuroscience

Neurons are « feature detectors ».

Hubel and Wiesel in area V1 : simple, complex, hypercomplex cells.

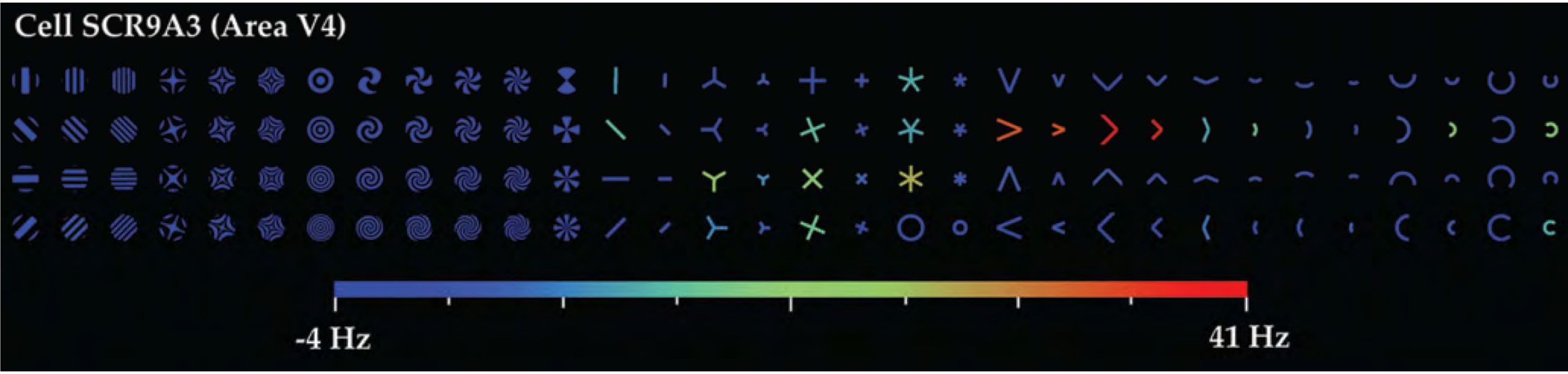
Area V4: shape or color detectors.

Area V5/MT: motion detectors.

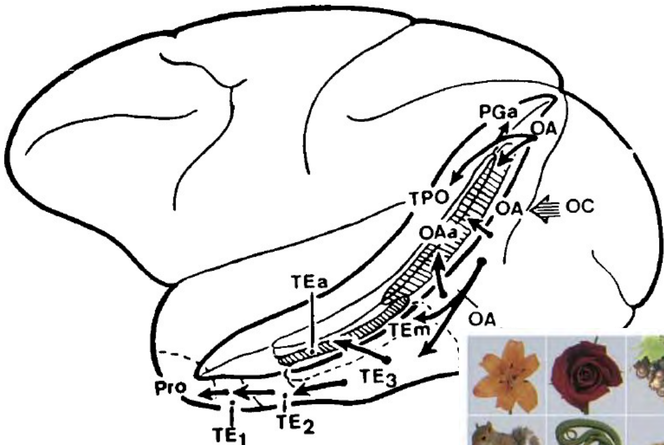
Infero-temporal cortex: face detectors.

Etc. etc.

Sensitivity to higher-level features in ventral visual cortex



Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas v2 and v4. Cerebral Cortex (New York, N.Y.: 1991), 17(5), 1100-1116. https://doi.org/10.1093/cercor/bhl020

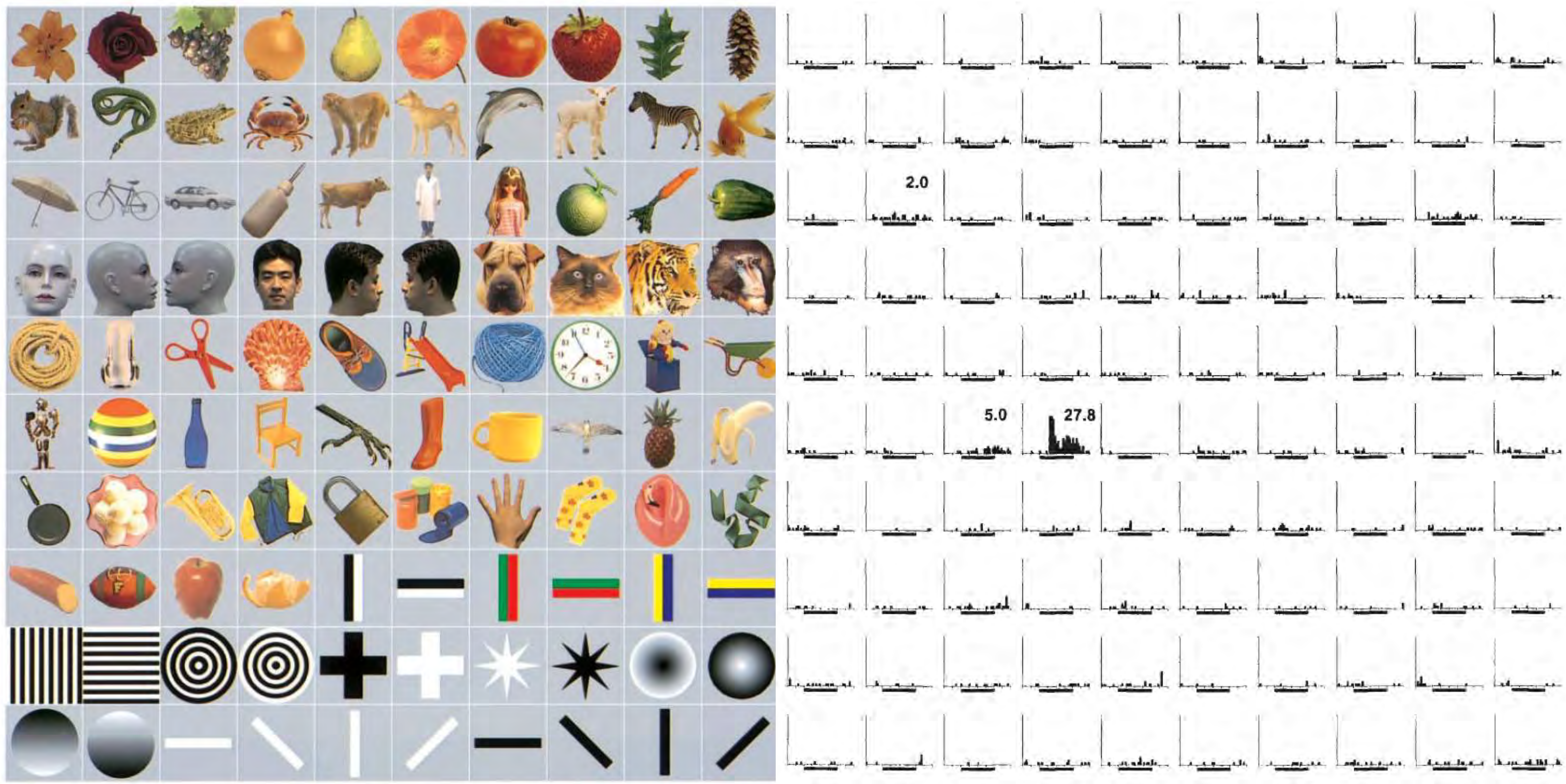


From Rolls, Neuron 2000

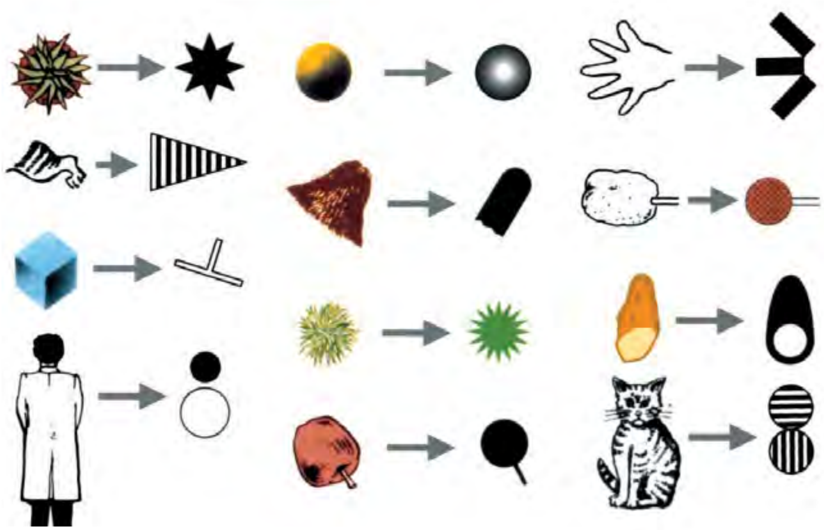
Characterizing neurons along the ventral visual pathway: An emphasis on highly selective responses

Gross, Tanaka, Logothetis, Poggio, Perrett, Orban, Rolls, etc.

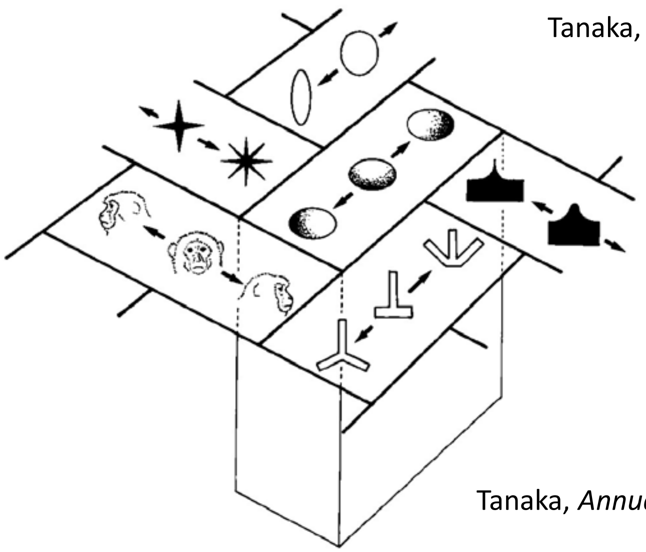
From Tamura, H., & Tanaka, K. (2001). Cerebral Cortex.



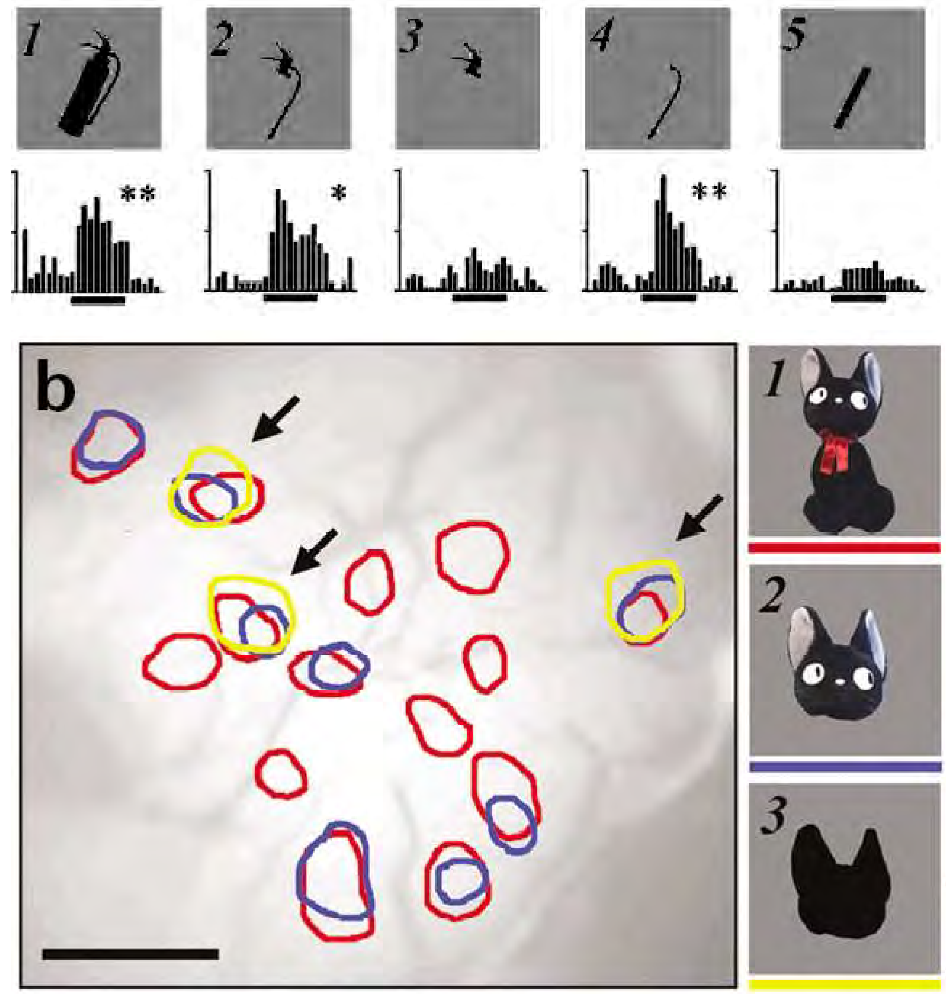
A search for the defining features for a given cell



Tanaka, 1996



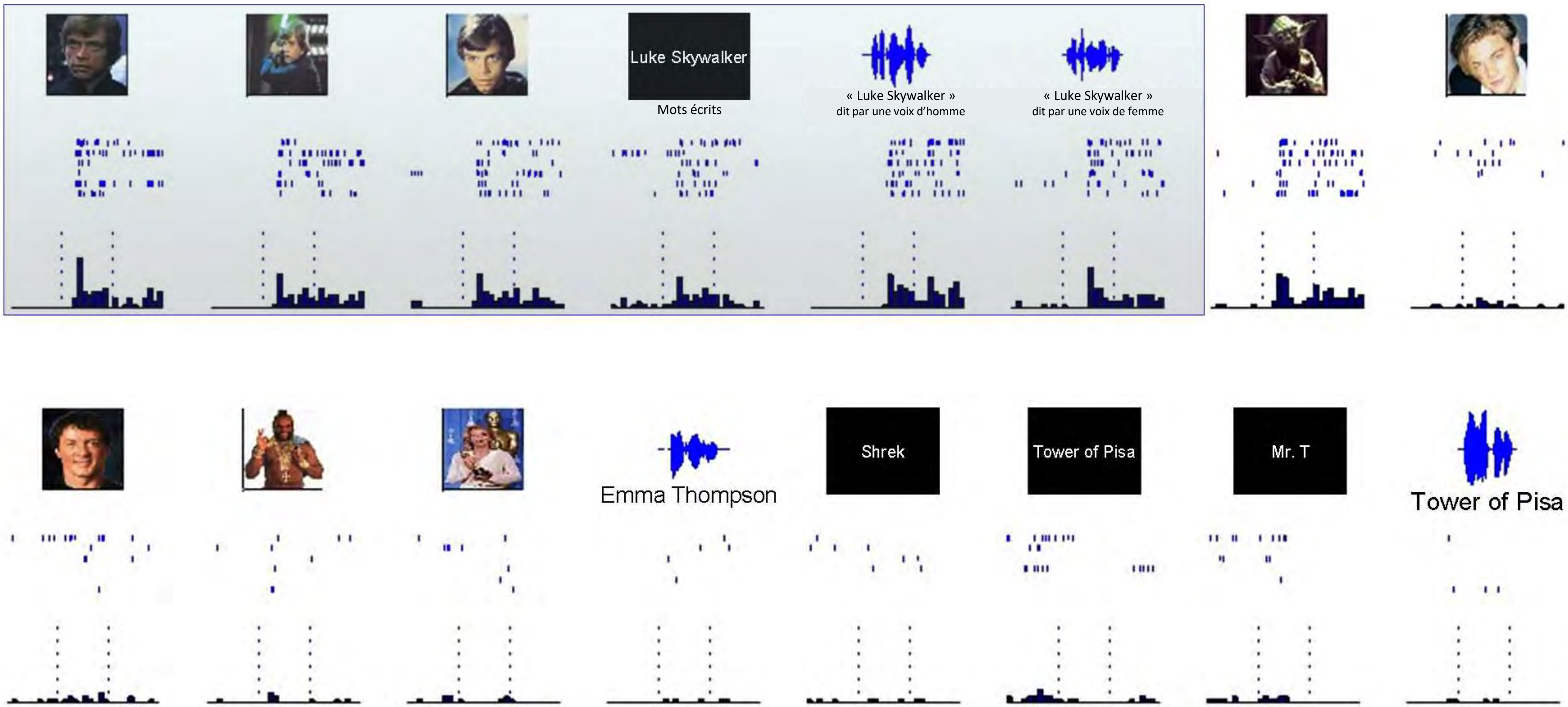
Tanaka, *Annual Rev. Neurosci.* 1996



Tsunoda, … & Tanifuji, *Nature Neuroscience* 2001

Concept cells in human anterior medial temporal lobe

Quiroga, R. Q. (2012). Concept cells : The building blocks of declarative memory functions. *Nature Reviews Neuroscience*, *13*(8), 587-597. Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102-1107.



« Grand mother » cells and arguments against them

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*(3), 87-91. https://doi.org/10.1016/j.tics.2007.12.003

Pushing things to the limit, some neuroscientist have entertained "the hypothesis that a single cell might respond to one and only one object or person, independently of, for example, its angle of gaze, location on the retina or facial expression."

→ 'grandmother cells', as named by Jerry Lettvin ; a.k.a. as 'pontificial cells' (Sherrington), 'gnostic cells' (Konorski) or 'cardinal cells' (Barlow)

But note that a "grandmother" code is ***not*** Rodrigo Quian-Quiroga's interpretation : he argues for a **sparse distributed code for concepts**.

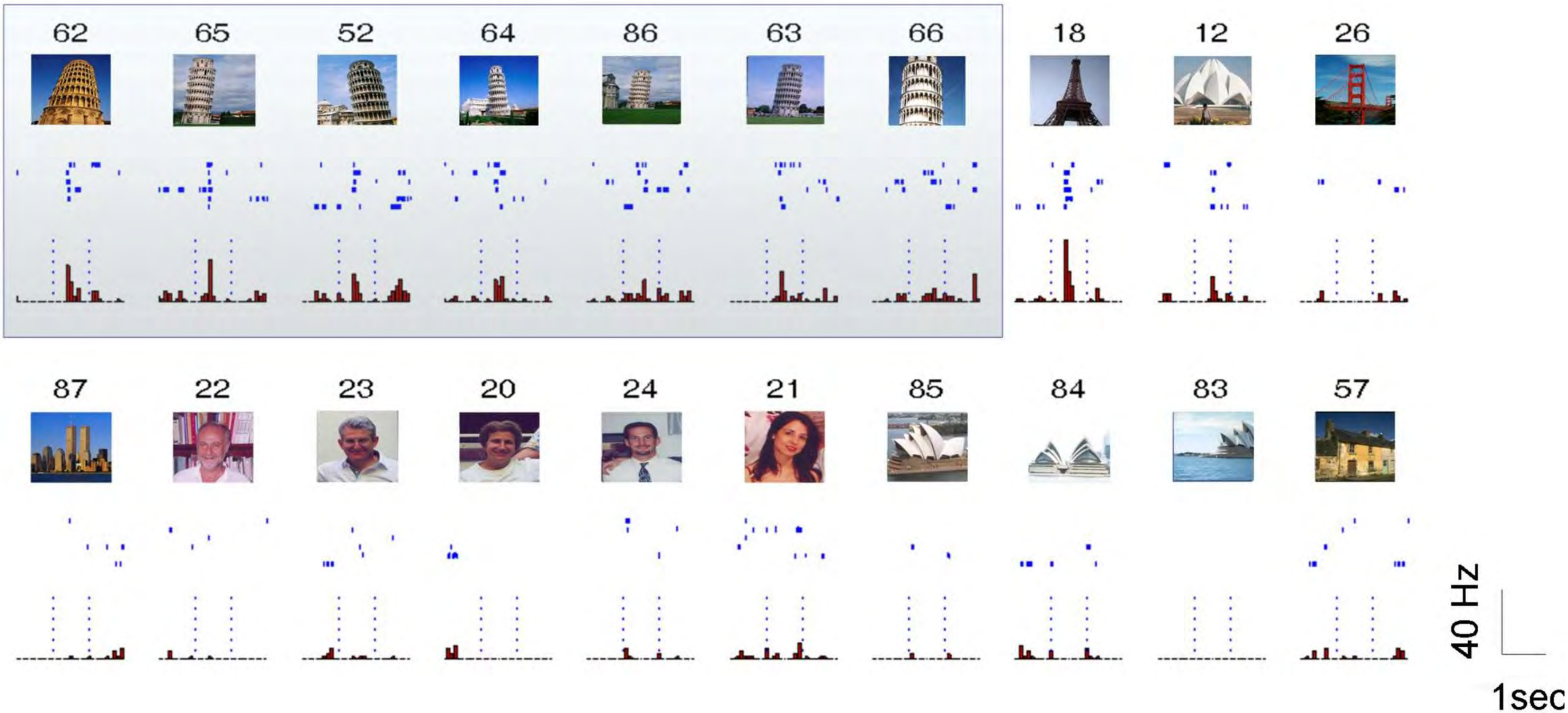
Arguments against a grandmother code:

- If a concept was coded by a single neuron, the probability of finding it would be vanishingly small. The fact that, by random recording, we find such a cell implies that there must be tens of thousands of them
- In Quiroga's work, only ~200 pictures are explored → we cannot exclude that each neuron responds to more than one picture…

and indeed this does occur.

A neuron that jointly responds to the Pisa tower and the Eiffel tower

Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102-1107.



More arguments against « grand mother » cells

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*(3), 87-91. https://doi.org/10.1016/j.tics.2007.12.003

Pushing things to the limit, some neuroscientist have entertained "the hypothesis that a single cell might respond to one and only one object or person, independently of, for example, its angle of gaze, location on the retina or facial expression."

→ 'grandmother cells', as named by Jerry Lettvin ; a.k.a. as 'pontificial cells' (Sherrington), 'gnostic cells' (Konorski) or 'cardinal cells' (Barlow)

Arguments against a grandmother code:

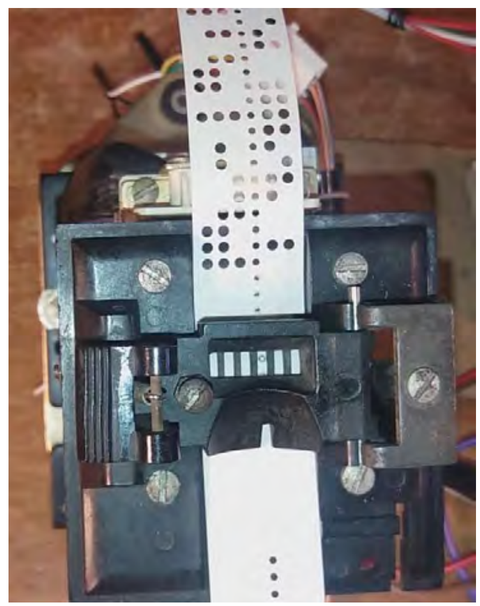
- If a concept was coded by a single neuron, the probability of finding this neuron would be vanishingly small. The fact that, by random recording, we find such a cell implies that there must be tens of thousands of them
- Only ~200 pictures are explored → we cannot exclude that each neuron responds to more than one picture, and indeed this does occur.
- The fact that ~200 pictures suffice to find one that makes a given neuron fire, implies that each neuron may respond to as many as 50-150 concepts.
- Only 40% of medial temporal lobe cells respond in this way – others may have a different, possibly more distributed coding scheme.

Other arguments:

- Even if a neuron responds very selectively, it owes its selectivity to ~10,000 synaptic inputs that *collectively* suffice to characterize a given concept → a pre-synaptic distributed code.
- Single-cell coding makes learning extremely difficult. How could the brain keep a list of "unused" or "unaffected" cells and assign a new one to each new concept ?
- Problem of compositionality: There are simply not enough neurons to encode all the concepts or images that we can process.

Philip K Dick : “The electrical ant” (1968)

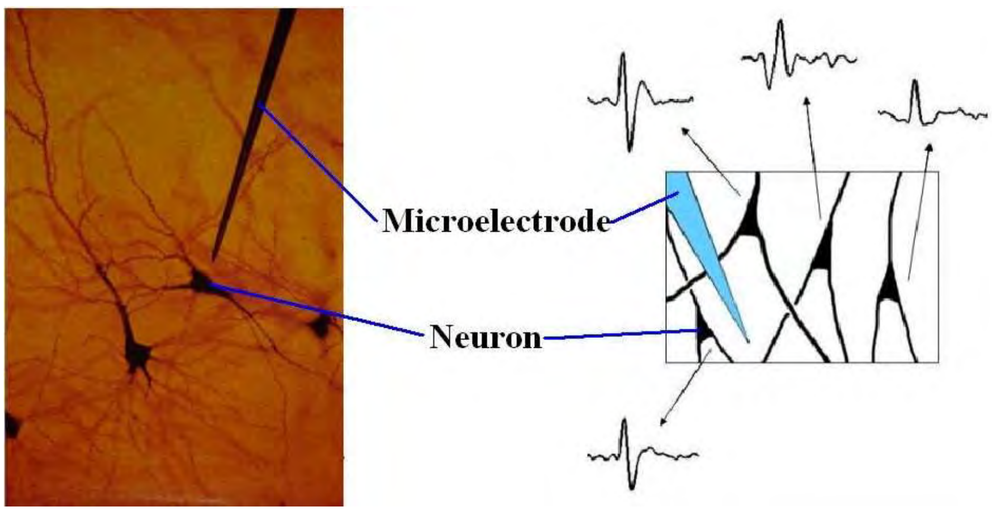
In this novel, the hero learns that he is not a human, but an organic robot. Furthermore, his subjective perception is entirely pre-programmed in a punched tape located inside his chest – each hole corresponds to a percept or concept! He discovers this by plugging some of the holes – and, for a short period, the corresponding parts of his subjective world vanish.
He reasons that, if all the holes were uncovered at the same time, he would experience all possible perceptions at once.



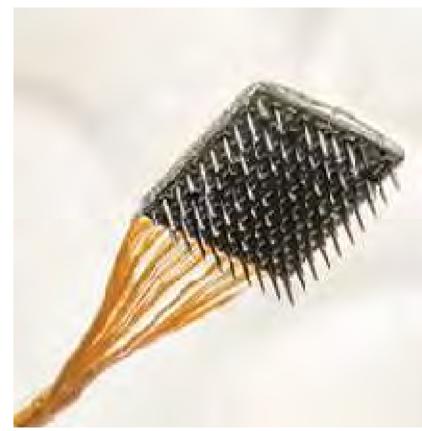
“He saw apples, and cobblestones, and zebras. He felt warmth, the silky texture of cloth; he felt the ocean lapping at him and a great wind, from the North, plucking at him as if to lead him somewhere.
(…) Butter relaxed into liquid on his tongue, and at the same time hideous odors and tastes assailed him: the bitter presence of poisons and lemons and blades of summer grass. He drowned; he fell; he lay in the arms of a woman in a vast white bed…”

Of course, the space of our perceptions is too large to be encoded by individual punches – unless the code is vastly combinatorial.

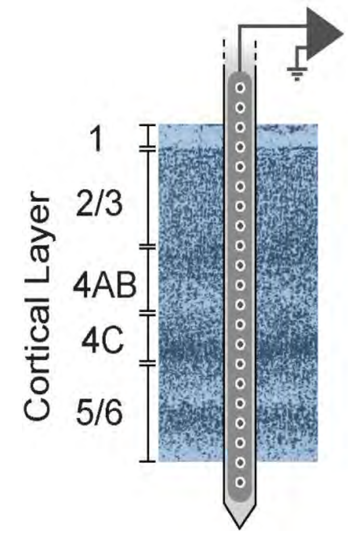
From single neurons to population recordings



Utah arrays

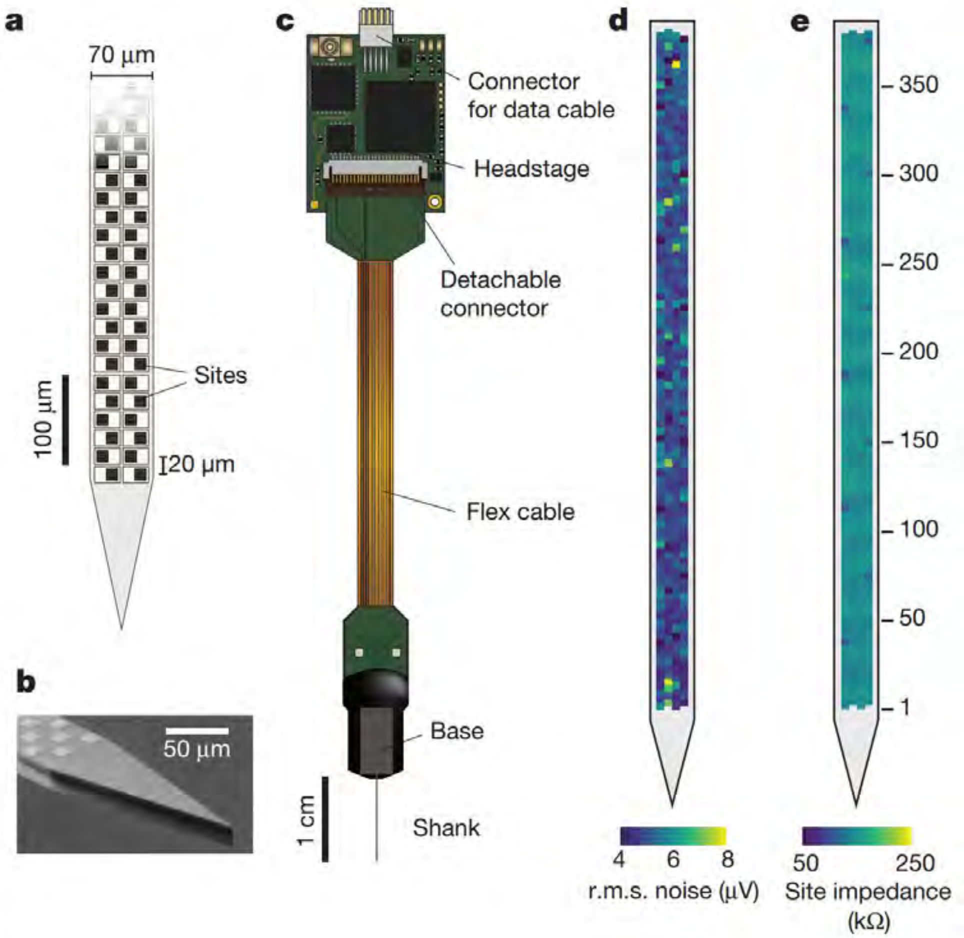


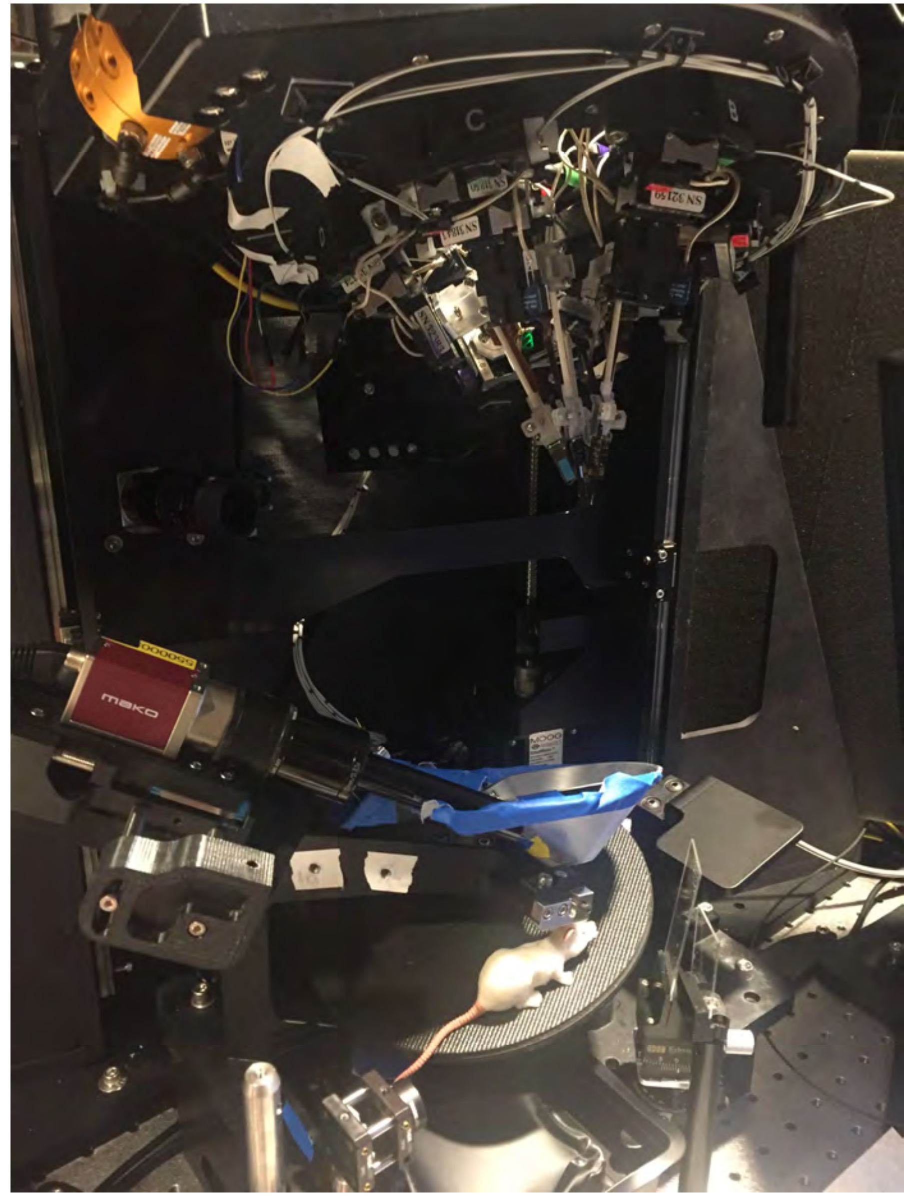
Laminar electrodes



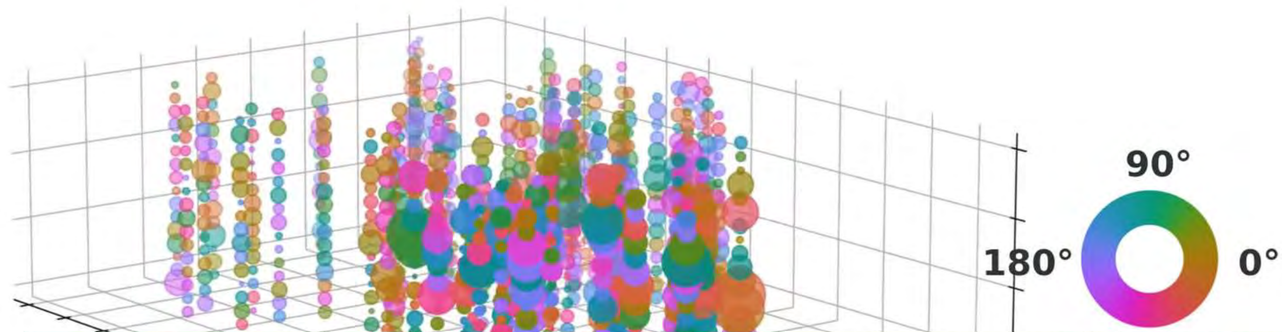
The revolution of NeuroPixels

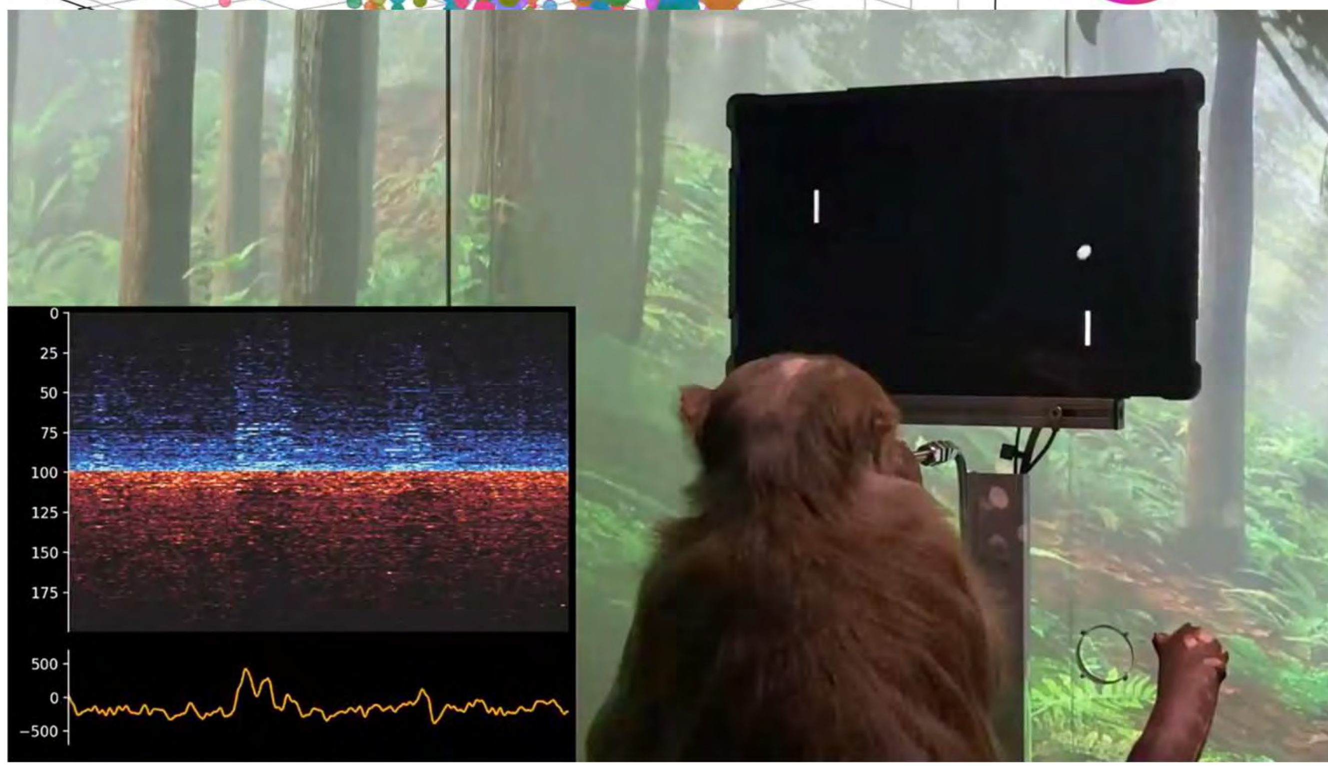
Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., … Harris, T. D. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature*, *551*(7679), 232–236.





NeuraLink : towards an industrialization of the brain-computer link





The neuralink chip is able to record wirelessly from 1024 electrodes, implanted in the motor cortex, and sealed, with very low power.

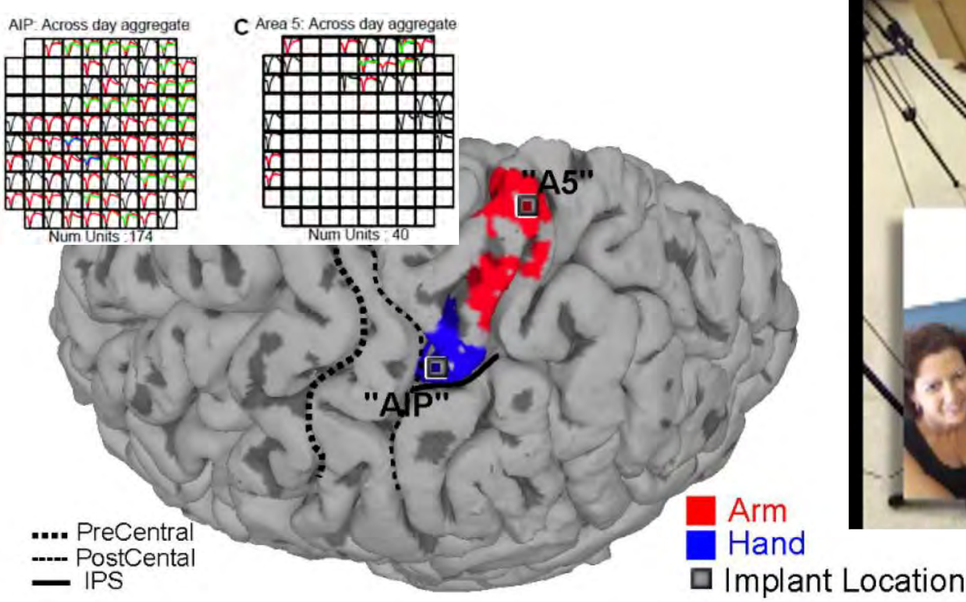
Here, they implanted two chips, one in each motor cortex, and used them to decode motor intentions (more on this later!) and control various computer games.

The rise of brain-computer interfaces for paralyzed patients

Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., … Andersen, R. A. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science (New York, N.Y.)*, *348*(6237), 906–910. https://doi.org/10.1126/science.aaa5417

A 32-year-old tetraplegic patient was implanted with two arrays, each comprising 100 microelectrodes.

Those electrodes can be used to decode various parameters of the intended movement, but also to give the patient some tactile feedback.

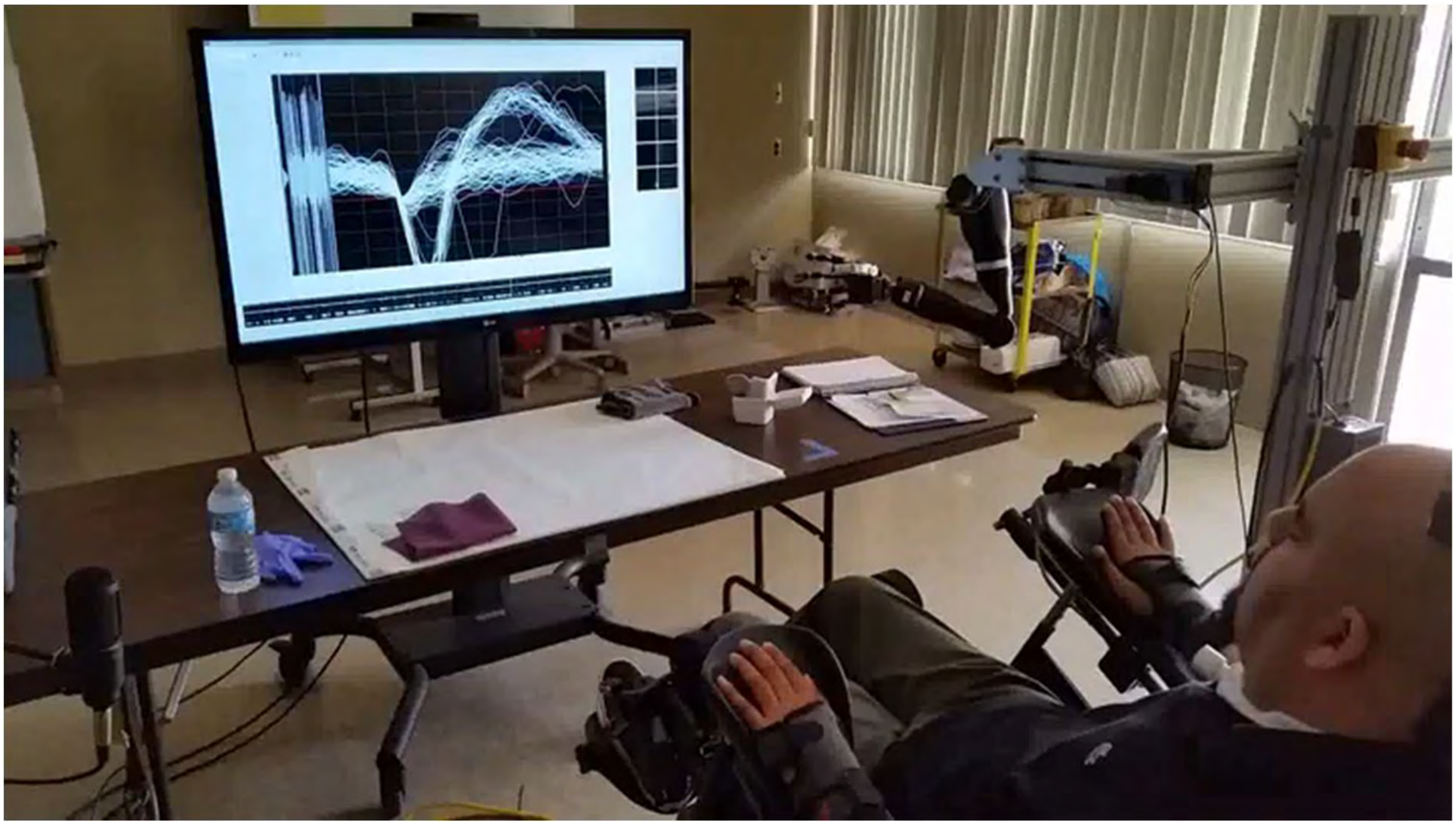


Imagination allows the patient to take control over a single neuron

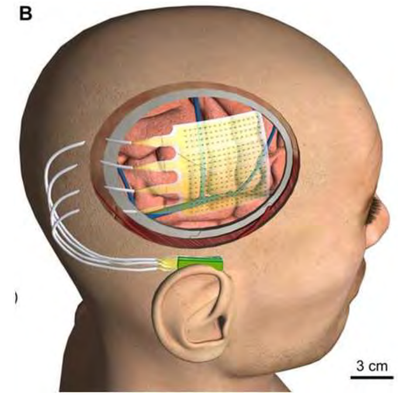
Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., … Andersen, R. A. (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. *Science (New York, N.Y.)*, *348*(6237), 906–910. https://doi.org/10.1126/science.aaa5417

The patient reports being able to control this particular neuron by imagining different activities:

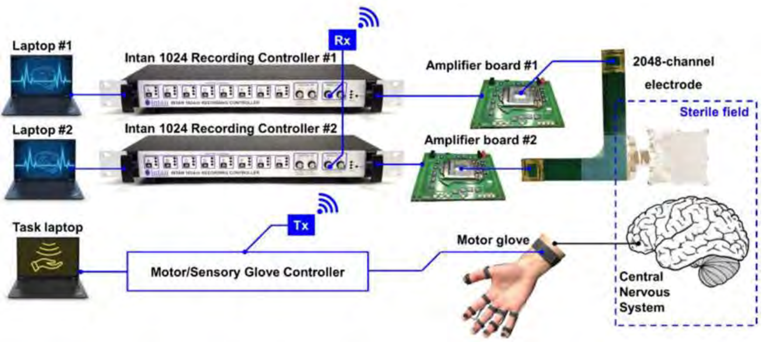
- Activating = rotating his shoulder
- Inhibiting = touching his nose



Thousands of electrophysiology channels in the human brain



Tchoe, Y., Bourhis, A. M., Cleary, D. R., Stedelin, B., Lee, J., Tonsfeldt, K. J., Brown, E. C., Siler, D. A., Paulk, A. C., Yang, J. C., Oh, H., Ro, Y. G., Lee, K., Russman, S. M., Ganji, M., Galton, I., Ben-Haim, S., Raslan, A. M., & Dayeh, S. A. (2022). Human brain mapping with multithousand-channel PtNRGrids resolves spatiotemporal dynamics. *Science Translational Medicine*. https://doi.org/10.1126/scitranslmed.abj1441



The vector field and streamlines represent the propagating **beta** waves, and green scattered dots represent the **high gamma** activity.

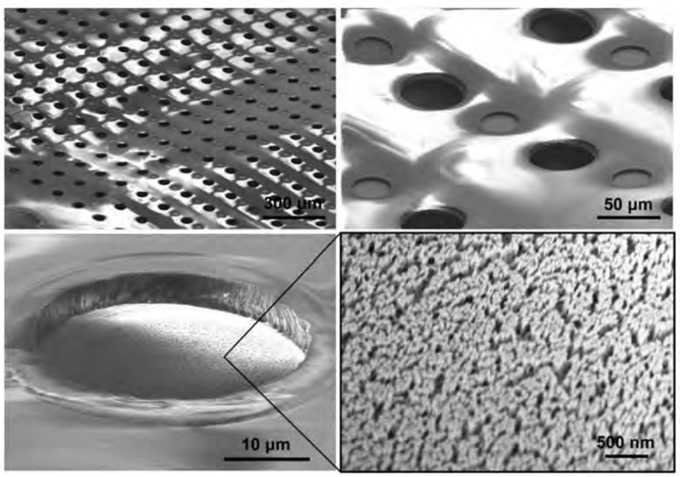
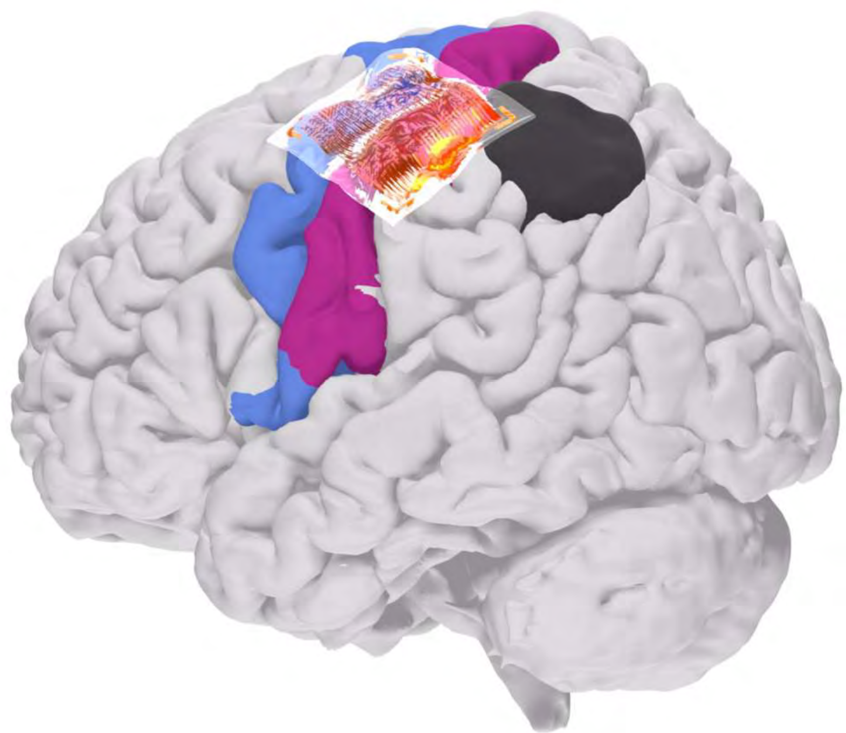
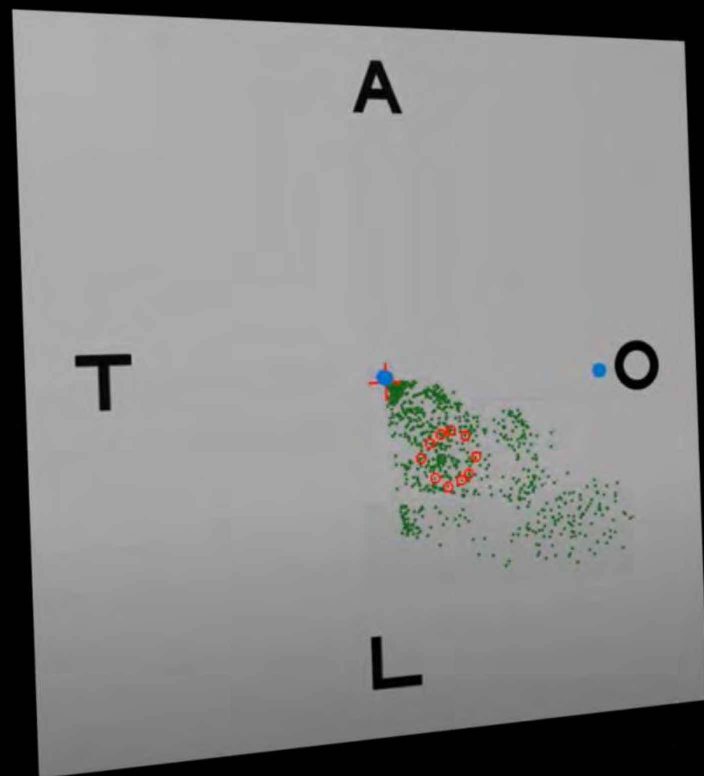
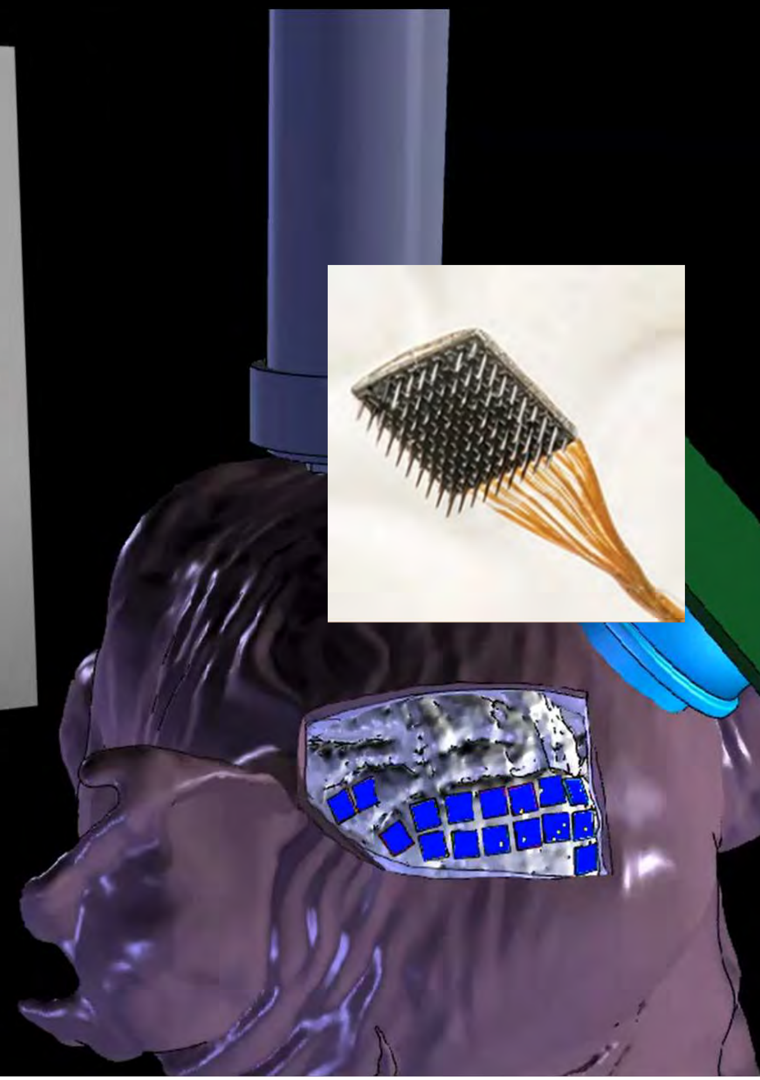


Figure S2. SEM images at different magnifications of the PtNRGrid and contacts.



Towards a restoration of fine vision through brain-computer interfaces





Roelfsema, P. R., Denys, D., & Klink, P. C. (2018). Mind Reading and Writing: The Future of Neurotechnology. *Trends in Cognitive Sciences*, *22*(7), 598–610. https://doi.org/10.1016/j.tics.2018.04.001
Chen, X., Wang, F., Fernandez, E., & Roelfsema, P. R. (2020). Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. Science, 370(6521), 1191-1196. https://doi.org/10.1126/science.abd7435

Optical imaging: recording from thousands of identified cells

Xie, Y., Hu, P., Li, J., Chen, J., Song, W., Wang, X.-J., Yang, T., Dehaene, S., Tang, S., Min, B., & Wang, L. (2022). Geometry of sequence working memory in macaque prefrontal cortex. Science, 375(6581), 632-639.

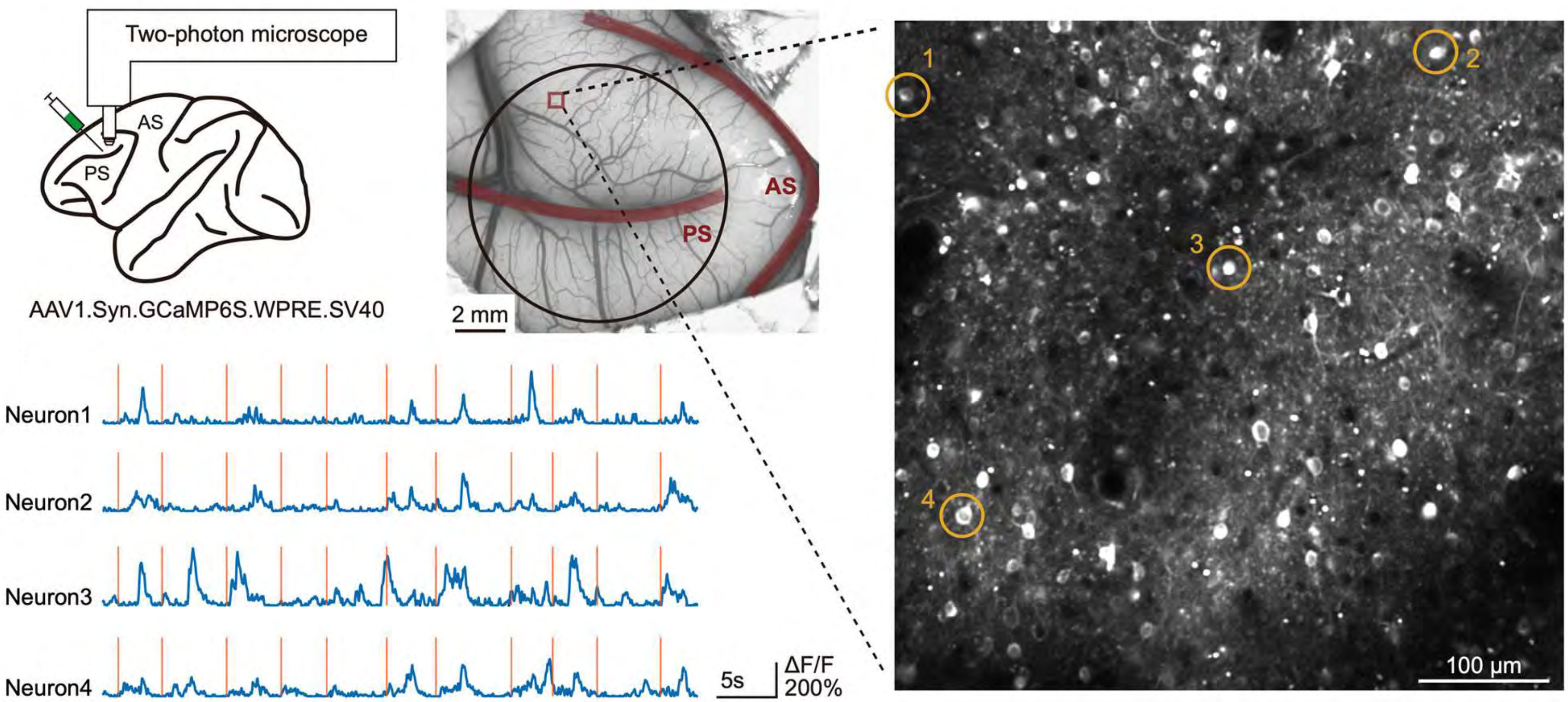
2- or even 3-photon imaging, combined with genetically encoded Calcium fluorescent indicators (usually GCaMP), allows to **visualize** (and not just record) hundreds or even thousands of neurons, in awake behaving monkeys, and to capture their spikes with a reasonable time resolution.
Here: over several days, total of 5325 neurons in 2 monkeys !


Yang Xie


Peiyao Hu


Bin Min

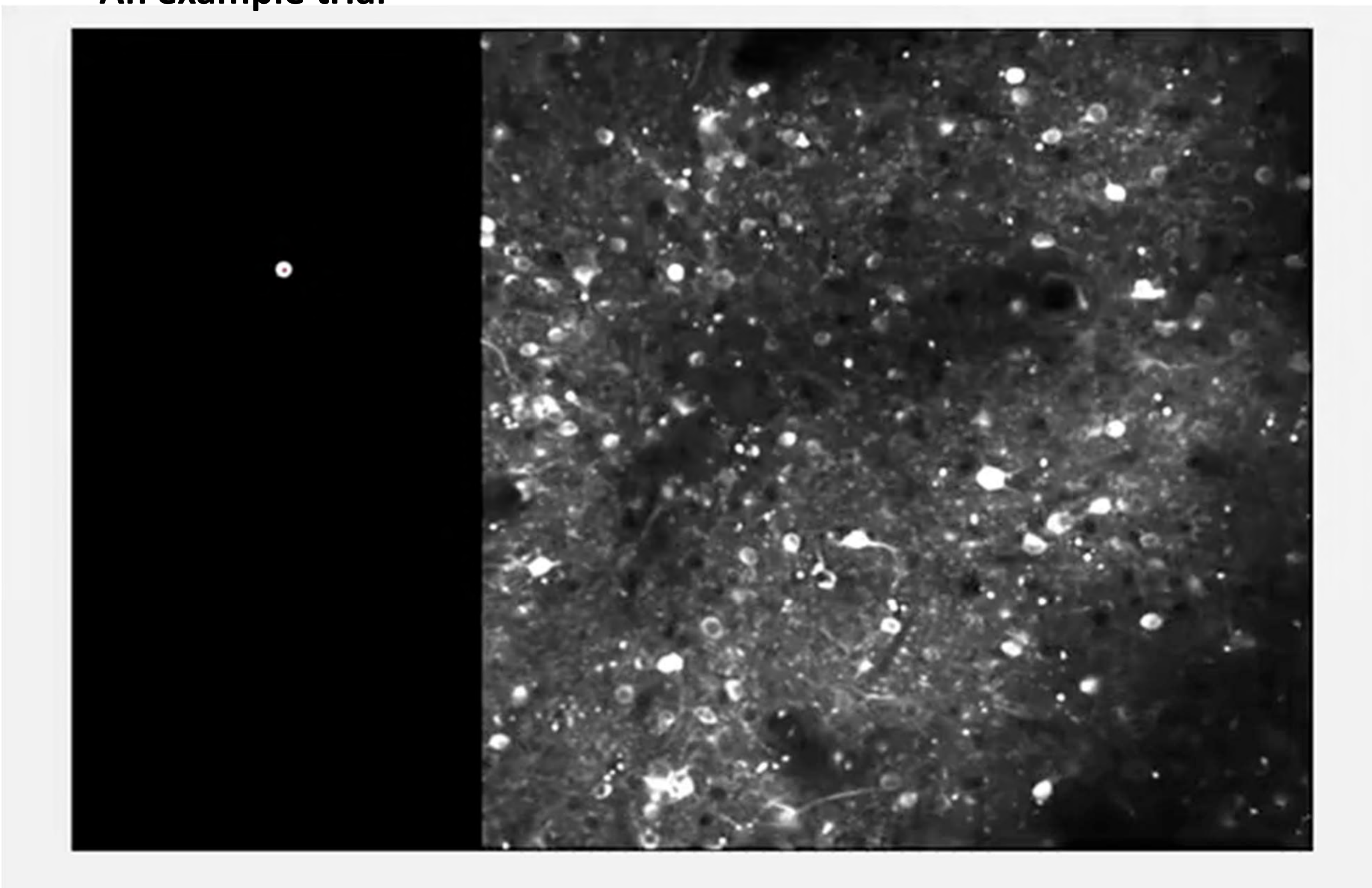

Shiming Tang (PKU)




Liping Wang

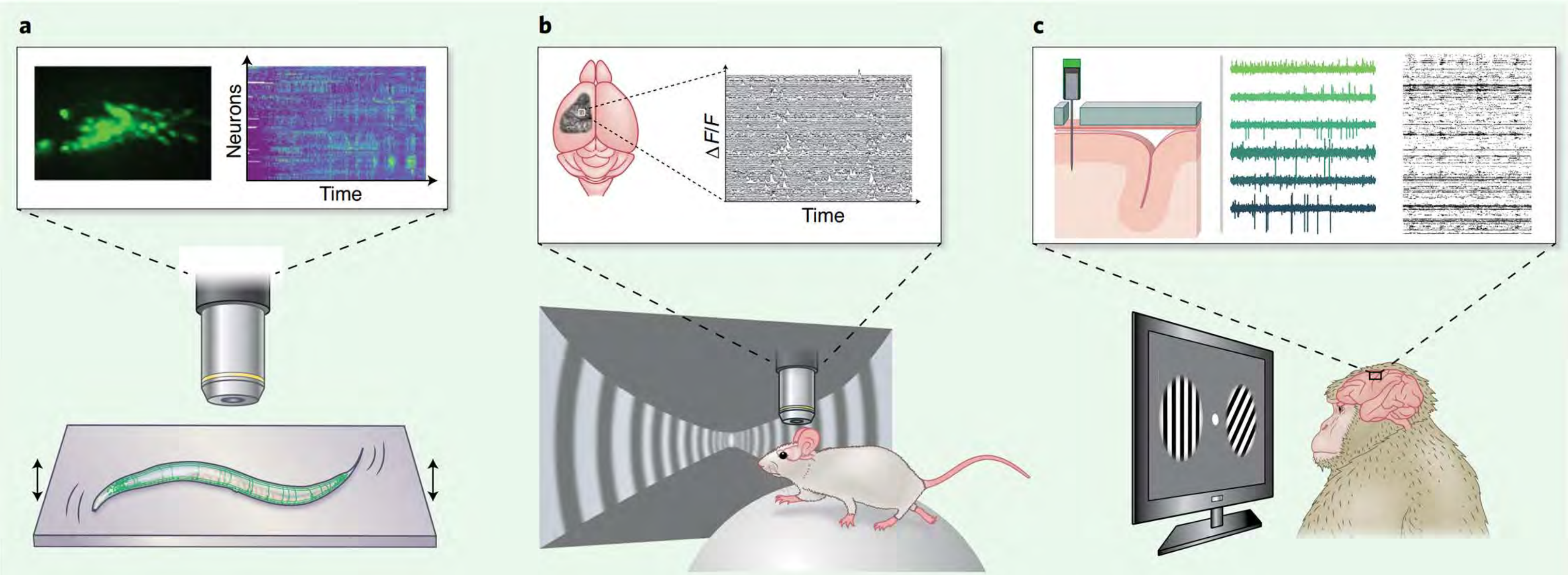
DLPFC, GCaMP6s, Field of view of 0.5X0.5mm, 32f/s

An example trial



Large-scale neurophysiology is compatible with awake, behaving animals.

Urai, A. E., Doiron, B., Leifer, A. M., & Churchland, A. K. (2022). Large-scale neural recordings call for new insights to link brain and behavior. *Nature Neuroscience*, 1-9. https://doi.org/10.1038/s41593-021-00980-9



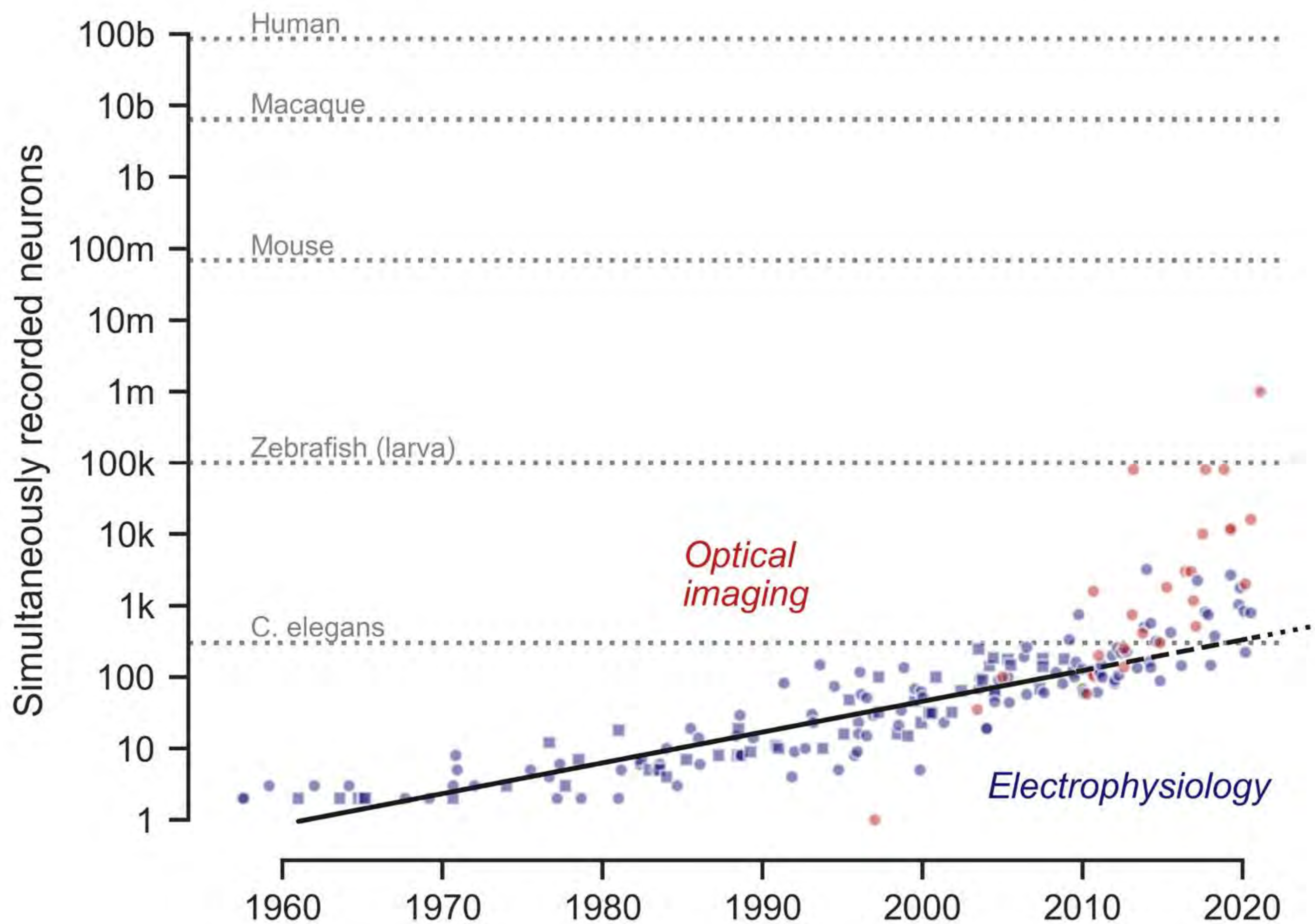
Large-scale neural recordings in behaving animals. **a**, Confocal microscopy of all neurons in *C. elegans* as it freely moves on an adjustable platform. Adapted from ref. [123] under a CC-BY license. **b**, Mesoscope two-photon imaging of the cortical surface while a mouse moves through a virtual reality by running on a ball. Adapted from ref. [124] under a CC-BY license. **c**, High-density electrophysiological recordings using Neuropixels probes, while a monkey performs a psychophysical decision-making task.

An exponential growth in the scale of neurophysiological recordings

Urai, A. E., Doiron, B., Leifer, A. M., & Churchland, A. K. (2022). Large-scale neural recordings call for new insights to link brain and behavior. *Nature Neuroscience*, 1-9. https://doi.org/10.1038/s41593-021-00980-9

Such large-recordings "call for new insights":

- Analysis tools for dimension reduction
- Visualization: new graphics
- Theoretical tools to characterize how a population of neurons responds and how it can encode features and concepts.



Two « doctrines » for neuroscience :
Single neurons versus population coding

- Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. *Neuron*. https://doi.org/10.1016/j.neuron.2021.07.011
- Chung, S., & Abbott, L. F. (2021). Neural population geometry : An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70, 137-144. https://doi.org/10.1016/j.conb.2021.10.010

But also

- Hebb, D. O. (1949). The organization of behavior. Wiley.
- Braitenberg, V. (1978). Cell assemblies in the cerebral cortex. In G. Palm (Éd.), Theoretical approaches to complex systems. (p. 171-188). Springer.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. PNAS, 79(8), 2554-2558.
- Amit, D. (1989). Modeling brain function : The world of attractor neural networks. Cambridge University Press.

and many others!

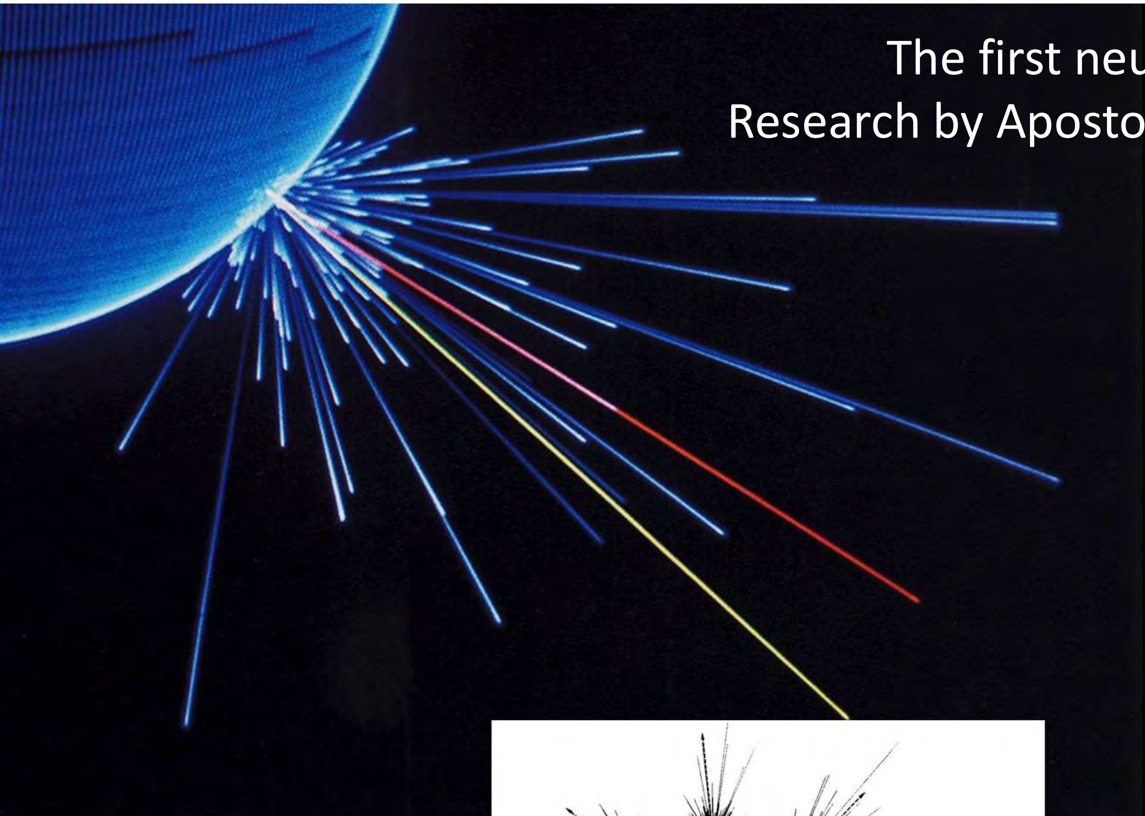
Five key concepts (from Ebitz & Hayden)
(1) the neural states that provide a snapshot of a pattern of activity across the population
(2) the manifold that encompasses the neural states that are possible (Manifold) or at least observed (manifold)
(3) the coding dimensions
(4) the subspaces that link neural states to behavior and cognition
(5) the dynamics that map activity from neural state to neural state, guiding how trajectories evolve through time and across the state space.

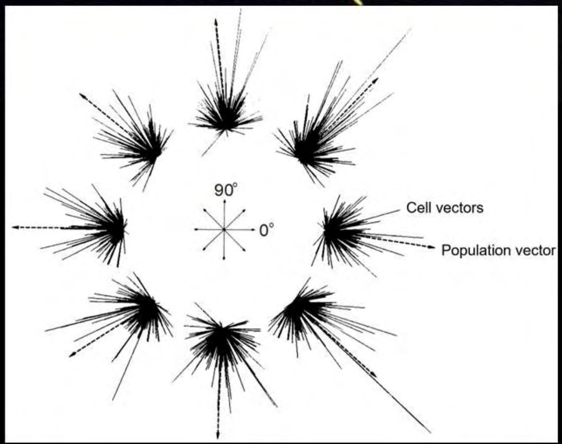
Consequence of this framework:
A “zoo” of individual cells with highly variable properties gets reinterpreted as
a **vector** with a dynamic trajectory

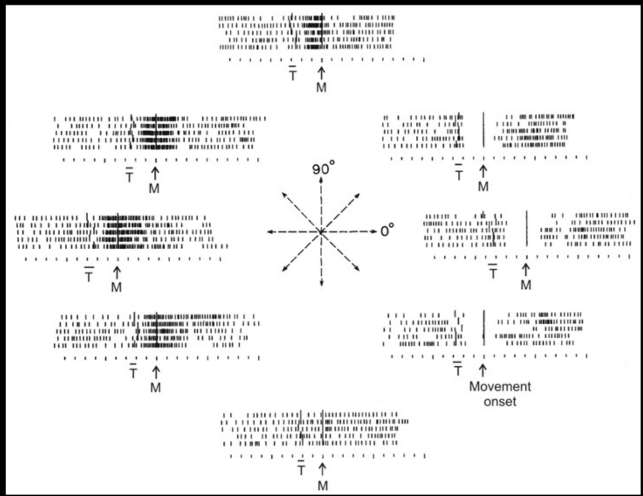
Main differences between the single-cell and the population view

Single-cell	Neural population
Each cell is dedicated to a **single feature or concept**.	Each cell can participate in **multiple cell assemblies**. Each cell can have distinct **weights**, corresponding to its variable (up or down) activation in response to various concepts.
Composite states can only be encoded by co-activating several cells (e.g. for color, motion, and location).	Composite states can be encoded by **superposition of orthogonal vectors**. A very large number of vectors can be superimposed, thus yielding a **factorial code**.
Learning requires the dedication of specific, previous unused cells, or the splitting of an existing population.	**Learning** requires either - the allocation of a new, possibly random coding vector, orthogonal to previous ones (those are in very large supply). - Or (at a slower time scale?) the reorganization of the entire manifold.

The first neuronal vector code:
Research by Apostolos Georgopoulos (1980's)

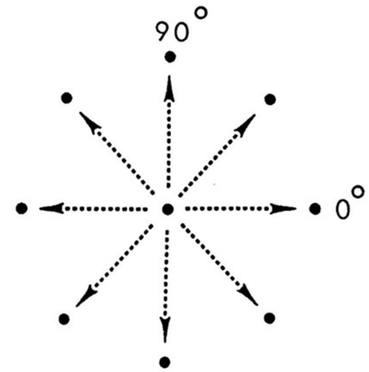


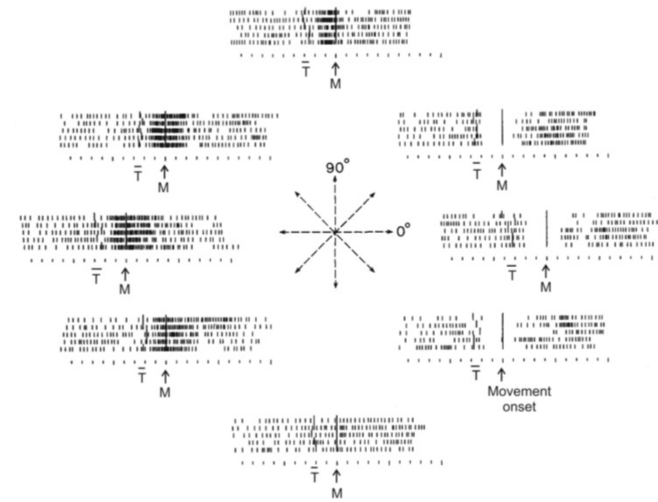




Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*(4771), 1416-1419. https://doi.org/10.1126/science.3749885

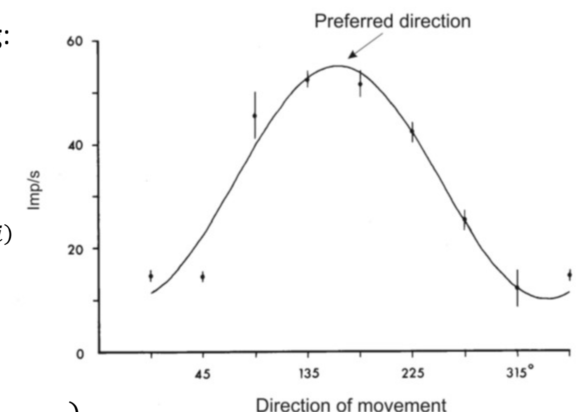
Macaque monkeys are asked to make movements in each of 8 possible directions, while single neurons are recorded in motor cortex.





Each neuron exhibits **directional tuning**: a sinusoidal tuning curve with a preferred direction of upcoming movement of the hand.

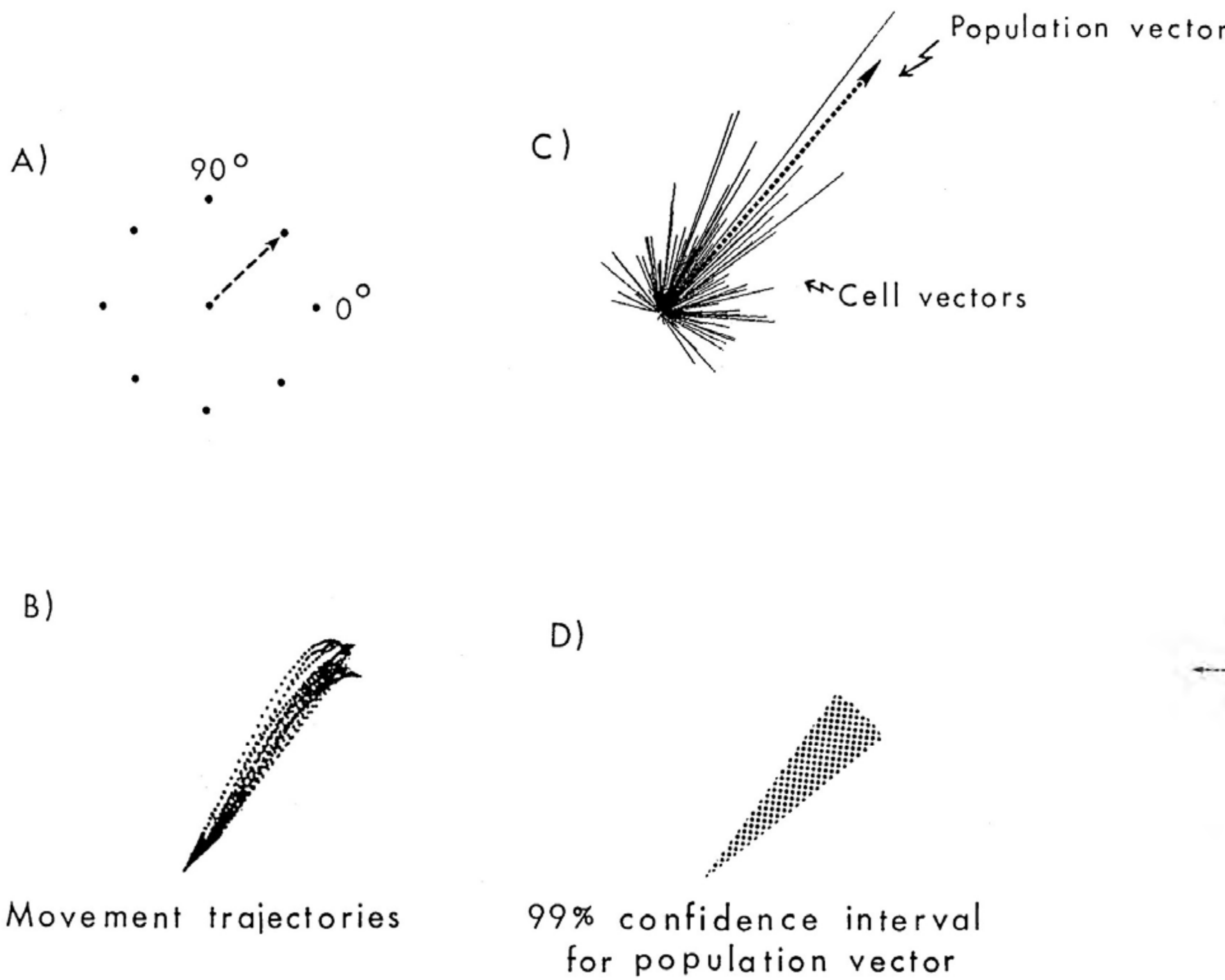
$$Firing\ (neuron\ i) = f_i$$
$$= k_i\, \vec{V}_{preferred(i)} \cdot \vec{V}_{movement(i)}$$
$$= k_i\ \cos(\theta_{movement} - \theta_i)$$



or equivalently :

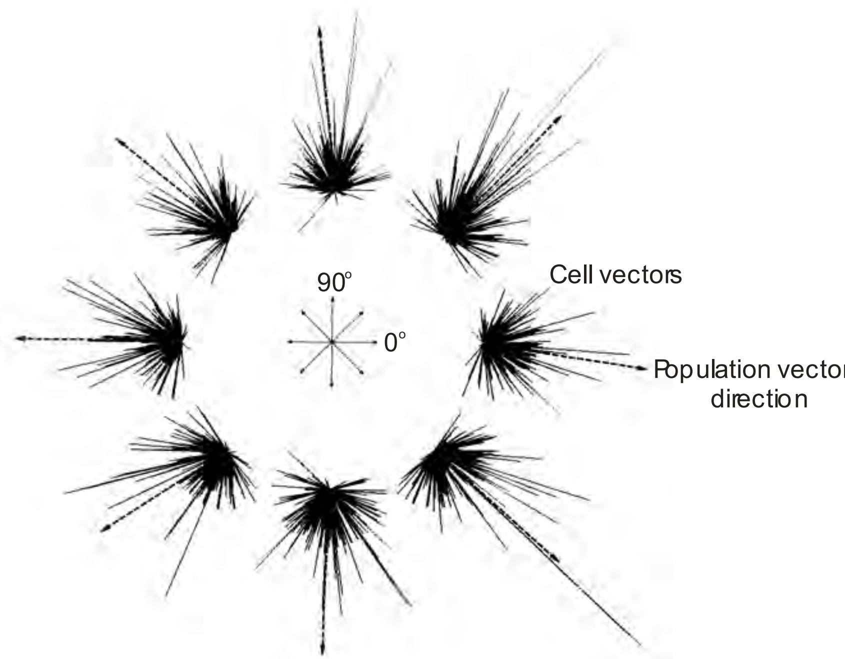
$$Firing\ (neuron\ i) = f_i$$
$$= a_i\ \cos(\theta_{movement}) + b_i\ \sin(\theta_{movement})$$
$$= a_i\, m_x + b_i\, m_y$$

Notion of "population vector"



Each neuron is ambiguous and noisy.
The direction of movement can be reconstituted by **pooling** over multiple neurons, each **voting** for a certain direction.
Each cell can be said to "point" in its preferred direction, its preferred **vector.**
The population vector is the vector sum (or average) of each neuron's preference.
Its components are simply given by

$V_x = \sum_i a_i f_i$ and $V_y = \sum_i b_i f_i$



The population vector for movements in 3 dimensions

The idea can be generalized to movements in 3D: each neuron possesses a “weight” for its preference along the x, y and z directions of movement.

$$Firing\ (neuron\ i) = f_i$$

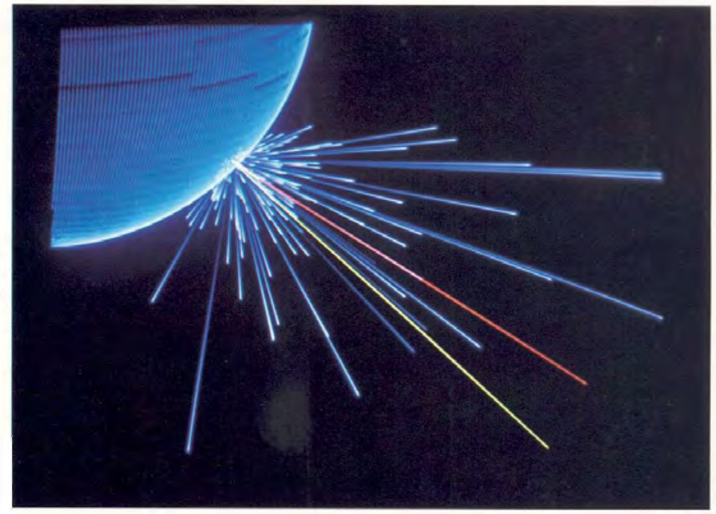
$$= a_i\, m_x + b_i\, m_y + c_i\, m_z$$

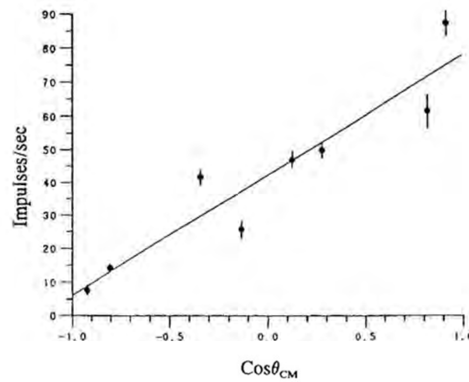
$$= k_i\ \cos(\theta_{movement,preferred})$$

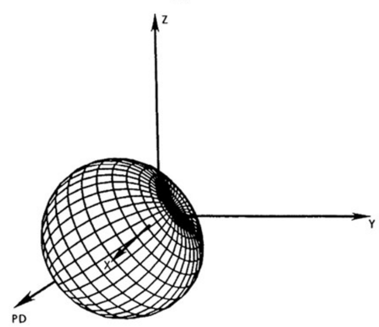
And the population vector can again be reconstituted by averaging the coefficients, each weighted by the neuron’s firing:

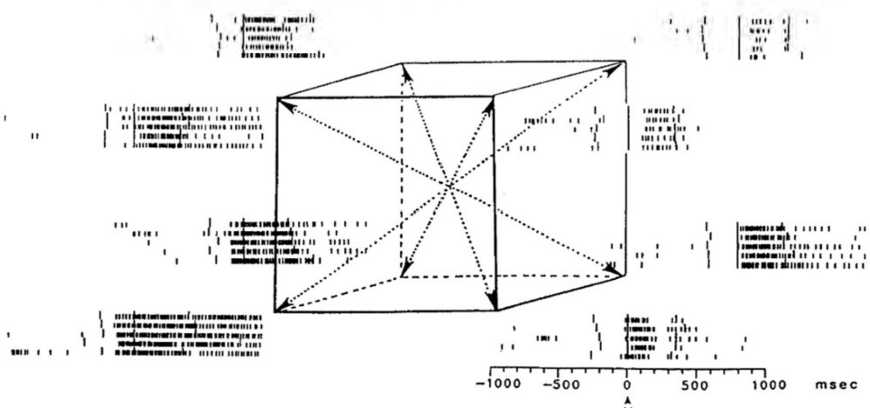
$$V_x = \sum_i a_i f_i\ , V_y = \sum_i b_i f_i\ , V_z = \sum_i c_i f_i$$

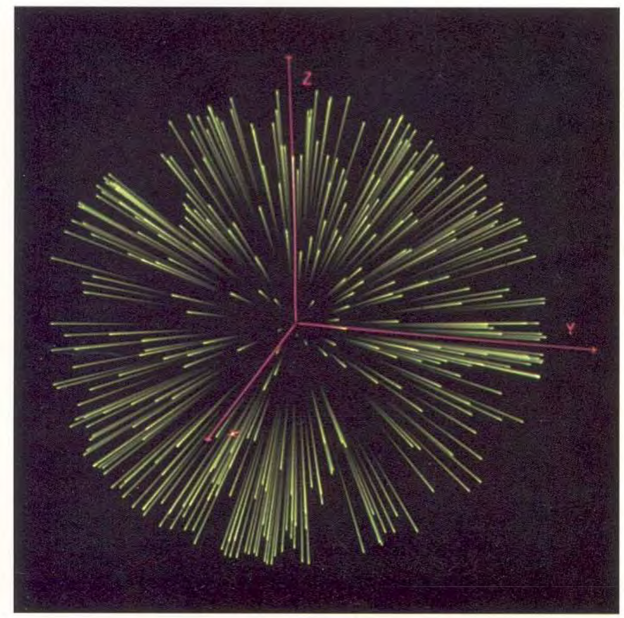
This simple idea works well because the preferred directions of the cells are roughly distributed equally across all directions.





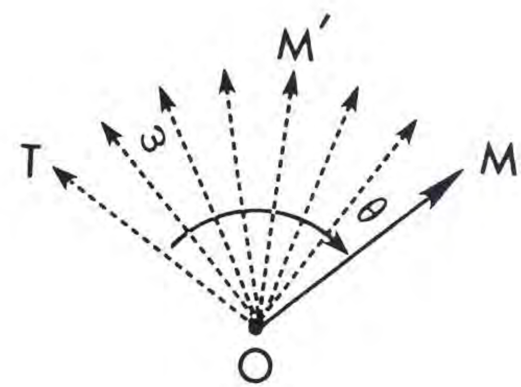


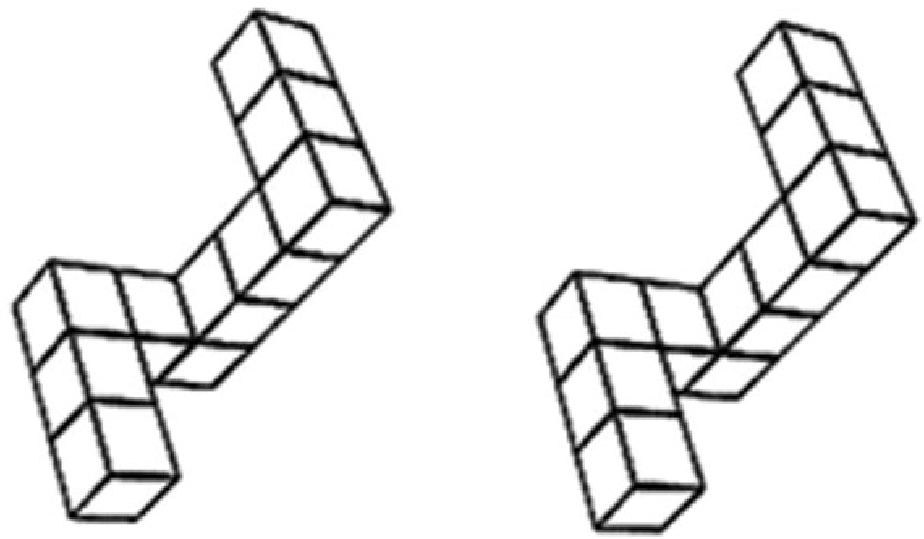


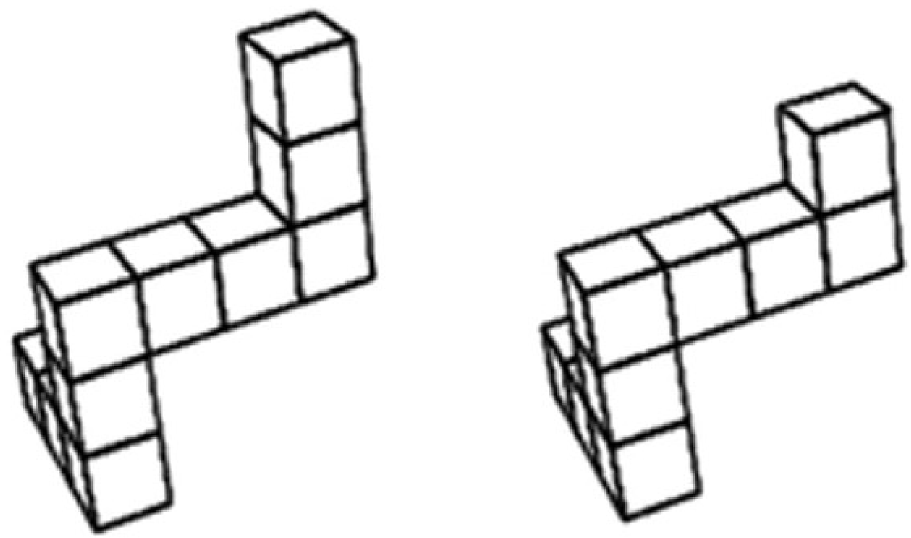


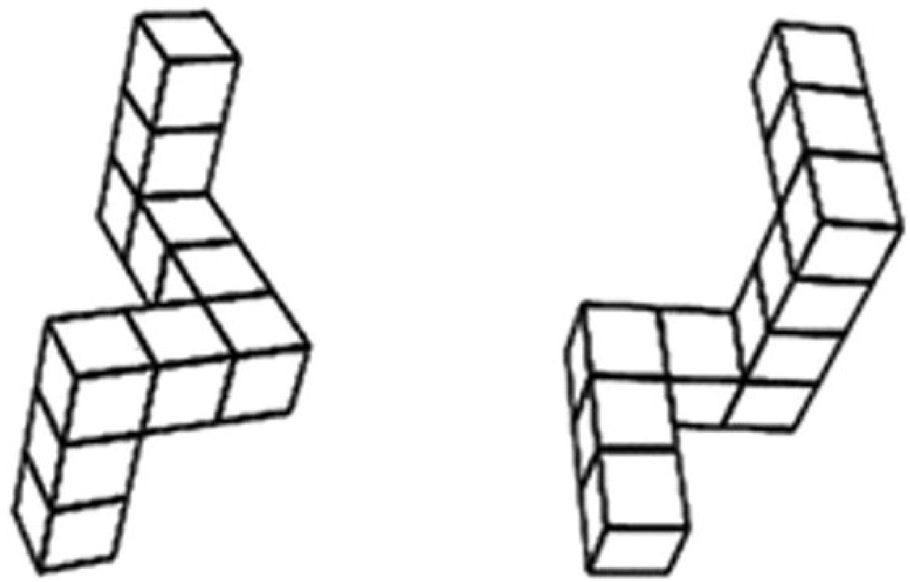
Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental Rotation of the Neuronal Population Vector. *Science*, *243*(4888), 234-236. https://doi.org/10.1126/science.2911737

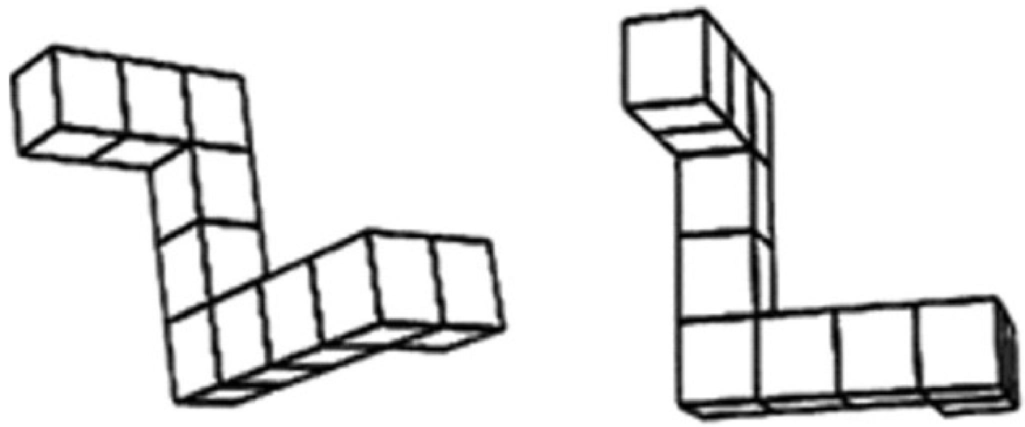
The mental rotation hypothesis:
If the monkey changes its mind and decides to move in another direction, the neural population vector should reflect this internal change by **rotating** internally.

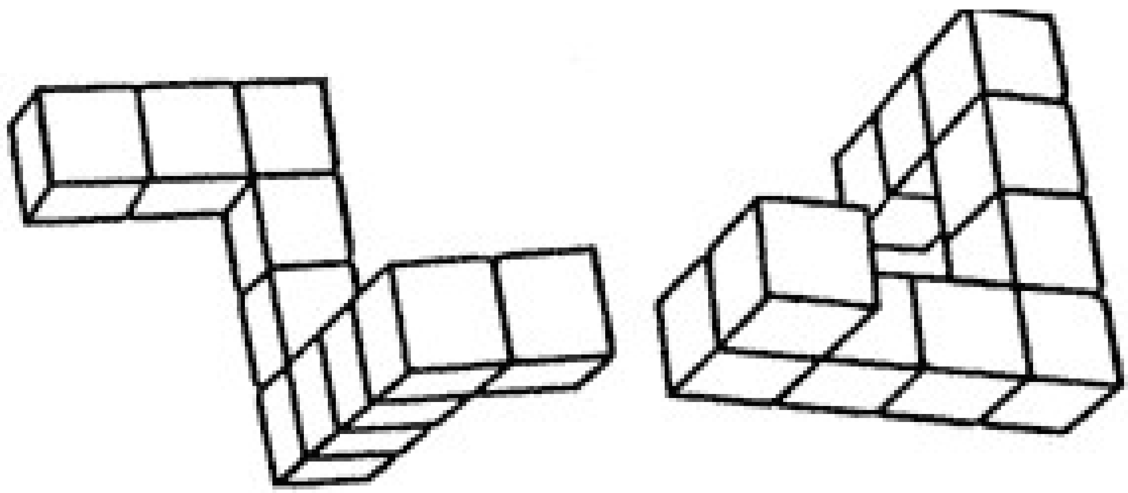




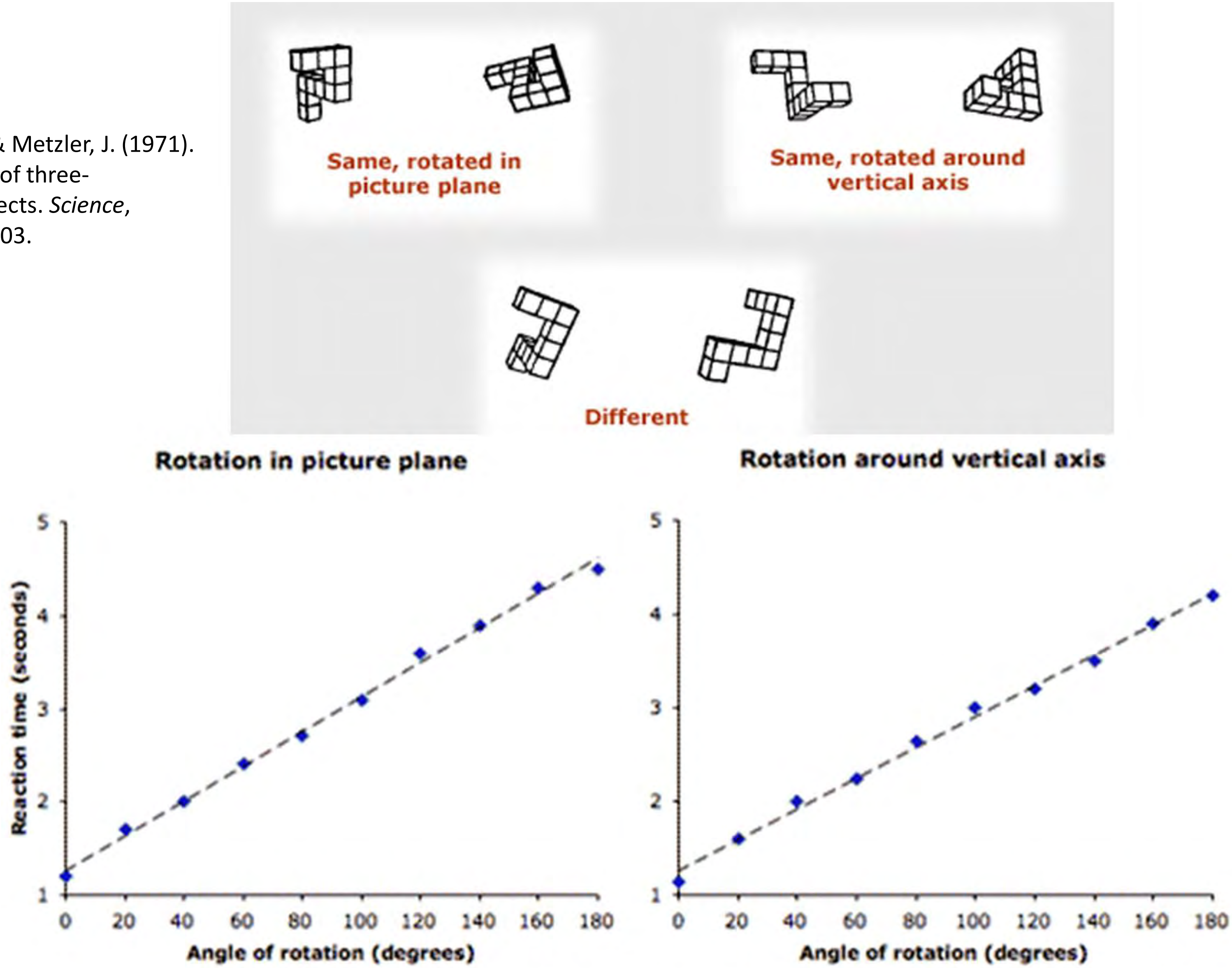






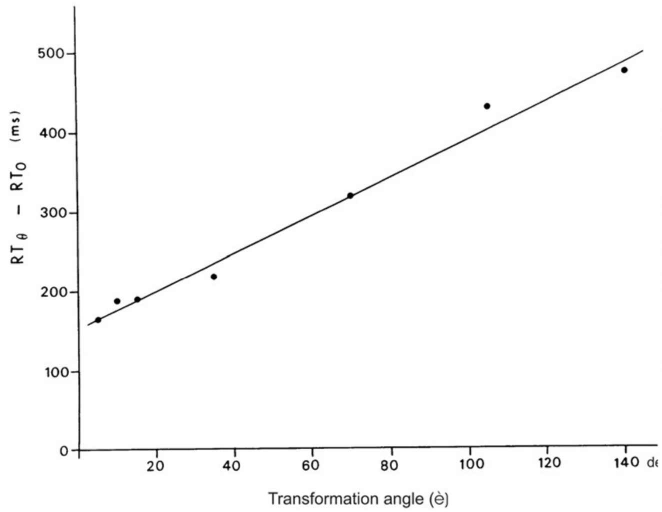


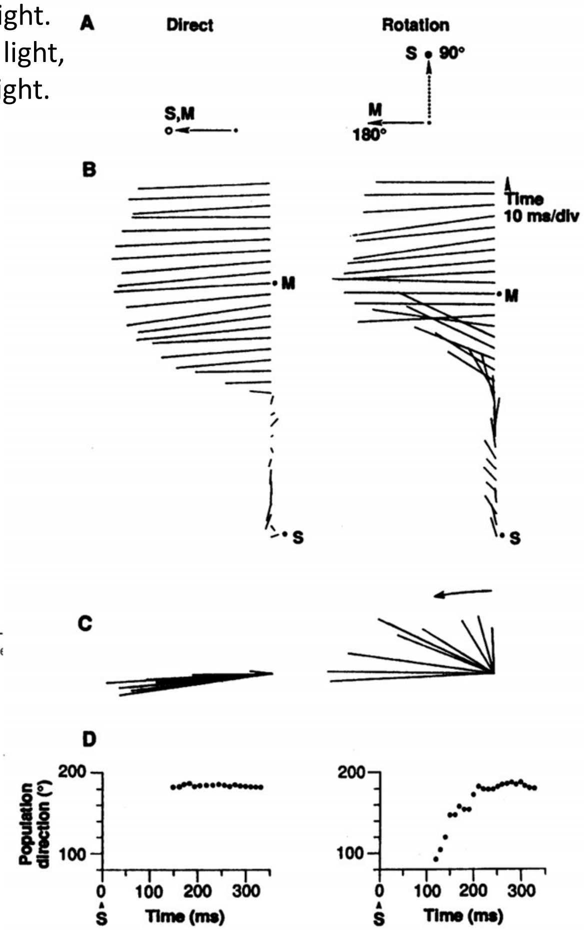
Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(972), 701–703.



Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental Rotation of the Neuronal Population Vector. *Science*, *243*(4888), 234-236. https://doi.org/10.1126/science.2911737

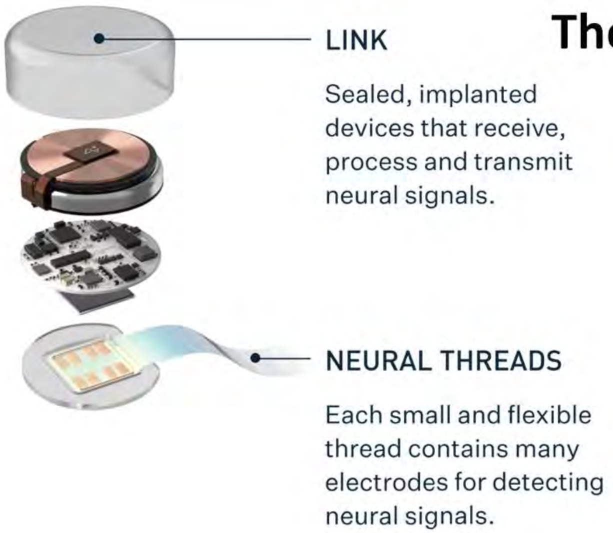
Monkeys are trained to move to the direction of a light. On some trials, depending on the brightness of the light, they have to make a movement at an angle to the light. Response time is proportional to angle:

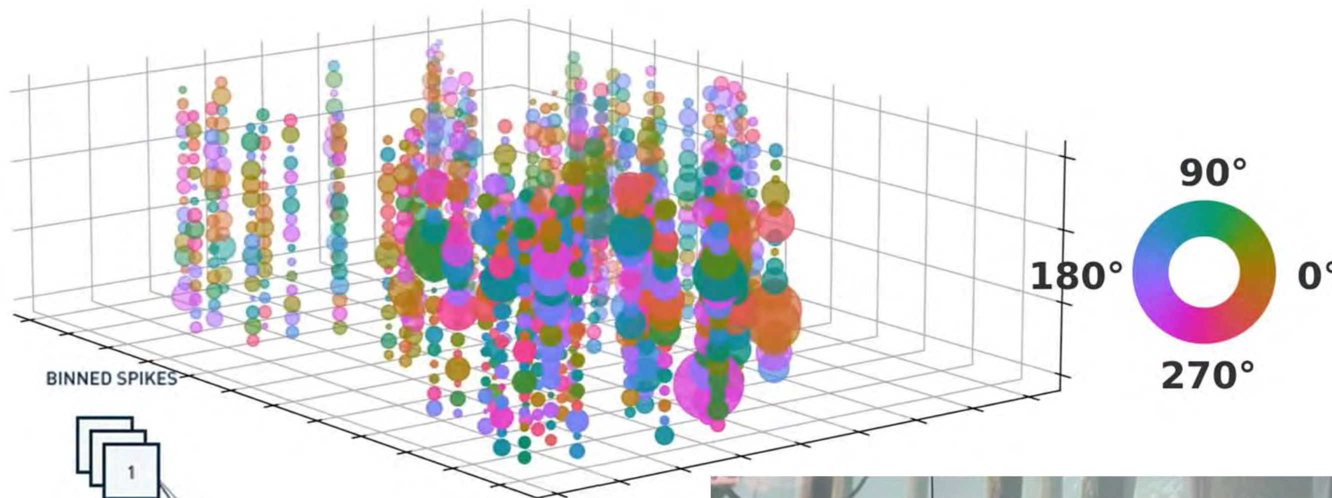




The population vector first indicates the direction of the light, and then turns towards the final movement direction

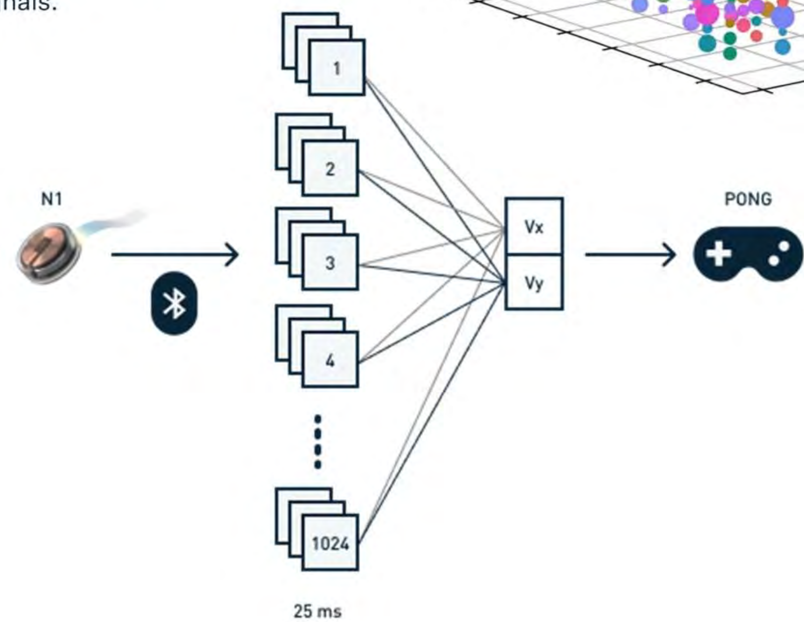
The NeuraLink decoder uses some variant of the population vector





The neuralink blog suggests that they use a simple weighted-sum mechanism (with time delays and integration over multiple windows of 25 ms) to predict the intended speed on the x and y directions.

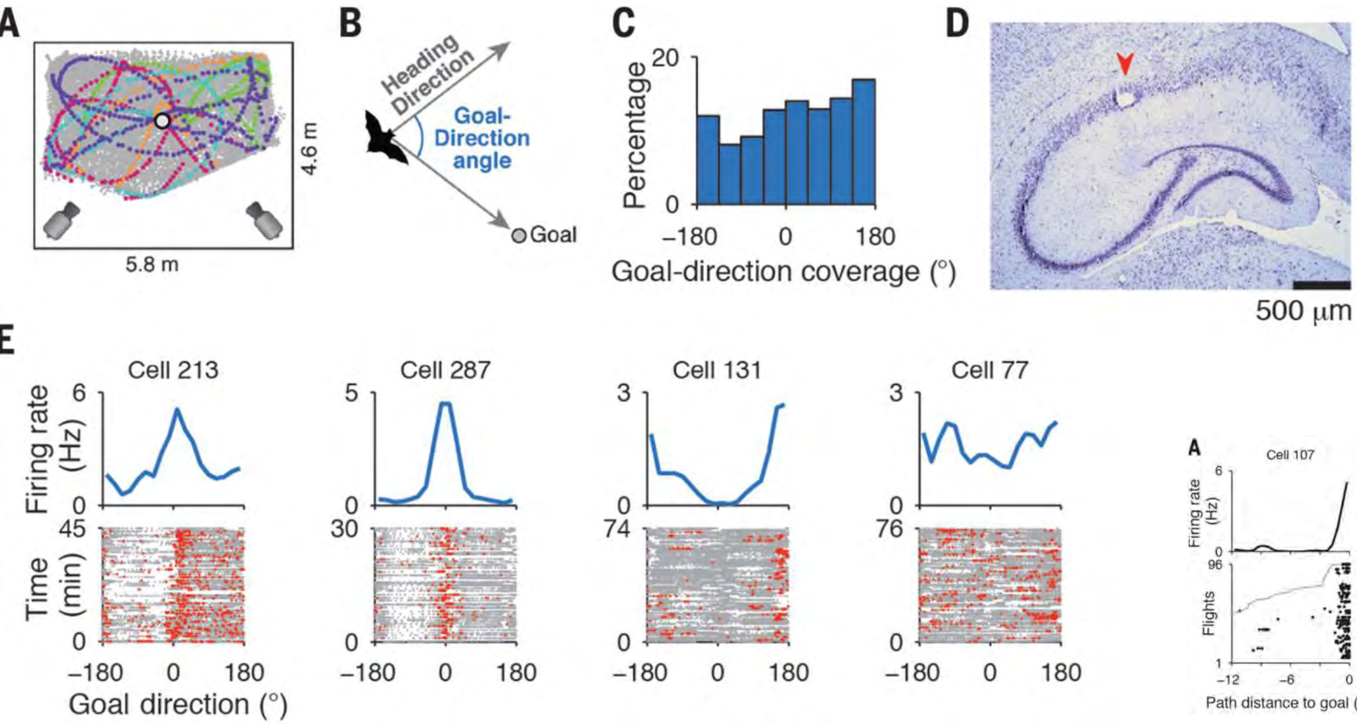
The video shows the weights assigned to each electrode where spikes are recorded.



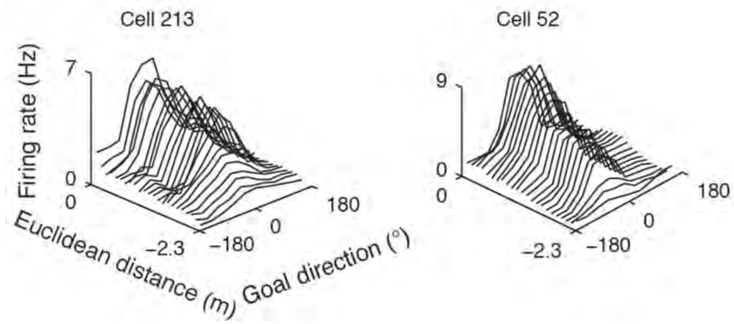
Vector coding of goal direction in the bat

Sarel, A., Finkelstein, A., Las, L., & Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. *Science*, *355*(6321), 176-180. https://doi.org/10.1126/science.aak9589

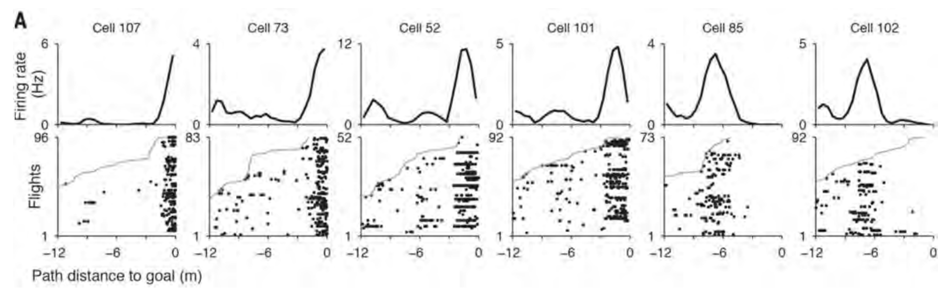
Vector coding by a “bump” of activity over a bank of neurons appears to be a very general strategy in many brains.
Here, bats were trained to land on a platform. in CA1 of the bat hippocampus, many neurons are tuned to the direction of the goal relative to the heading direction.



Distance is also encoded, often simply by increasing or decreasing the firing of the same neurons as a function of Euclidean distance.

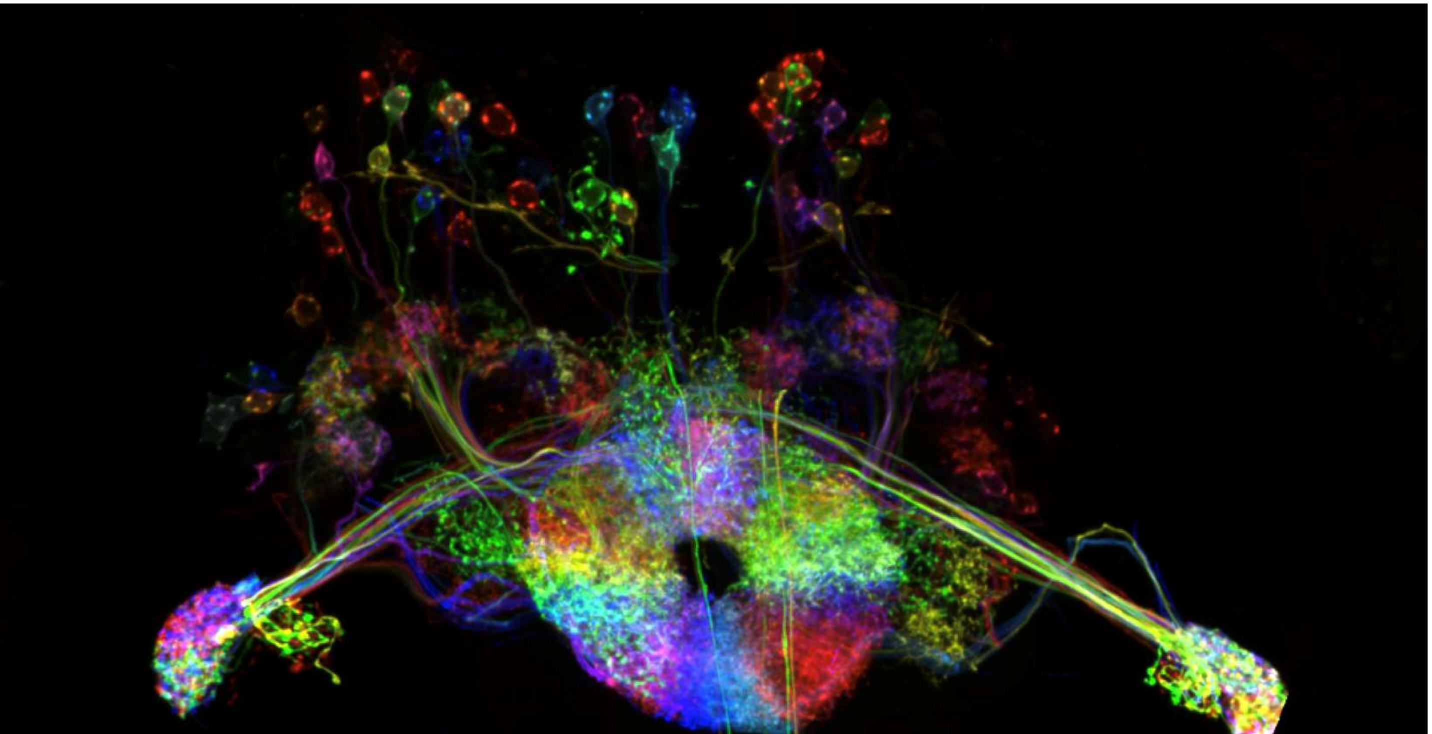


… although a few neurons are tuned to intermediate distances as well:

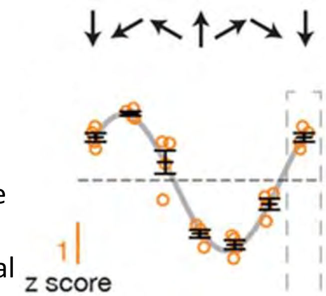


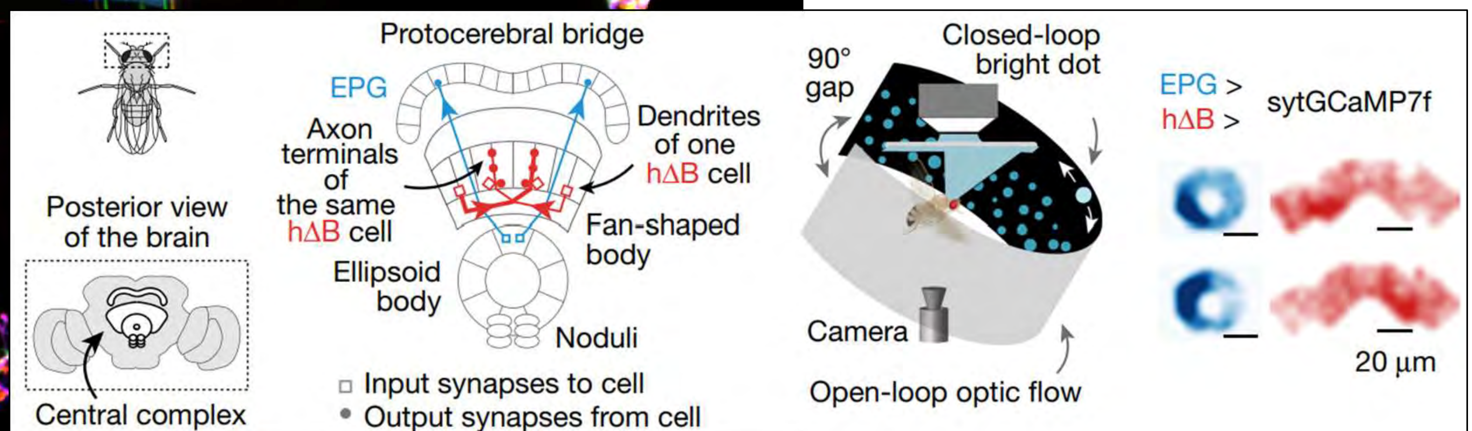
Similar encoding of vector direction in the fly brain!

Lyu, C., Abbott, L. F., & Maimon, G. (2022). Building an allocentric travelling direction signal via vector computation. *Nature*, *601*(7891), 92-97. https://doi.org/10.1038/s41586-021-04067-0



In drosophila, the complete connectome is available (EM).
Activity can be recorded using calcium reporters and 2P imaging, while moving in virtual reality!
Bumps of activity encode movement direction.
Cell activity is a sinusoidal function of direction.





Vector coding and vector arithmetic in the fly

Successive stages of the fly's brain represent stimuli (e.g. optic flow, target location) in egocentric coordinates, then world-centered, then back to egocentric motor commands.

The main principles of vector coding and arithmetic are well understood:

1. A **vector** is coded by a "**phasor**", a sinusoidal distribution of activity over a bank of neurons, whose amplitude represents vector size, and phase represents vector direction.
2. Vector **rotation** can be implemented by shifting the bump of activity left or right on this neural population
3. Vector **addition** is as simple as adding the activity of the two neural "phasor" populations
4. Vector **projection** can be encoded by non-linearities in the neural response to a combination of two inputs.

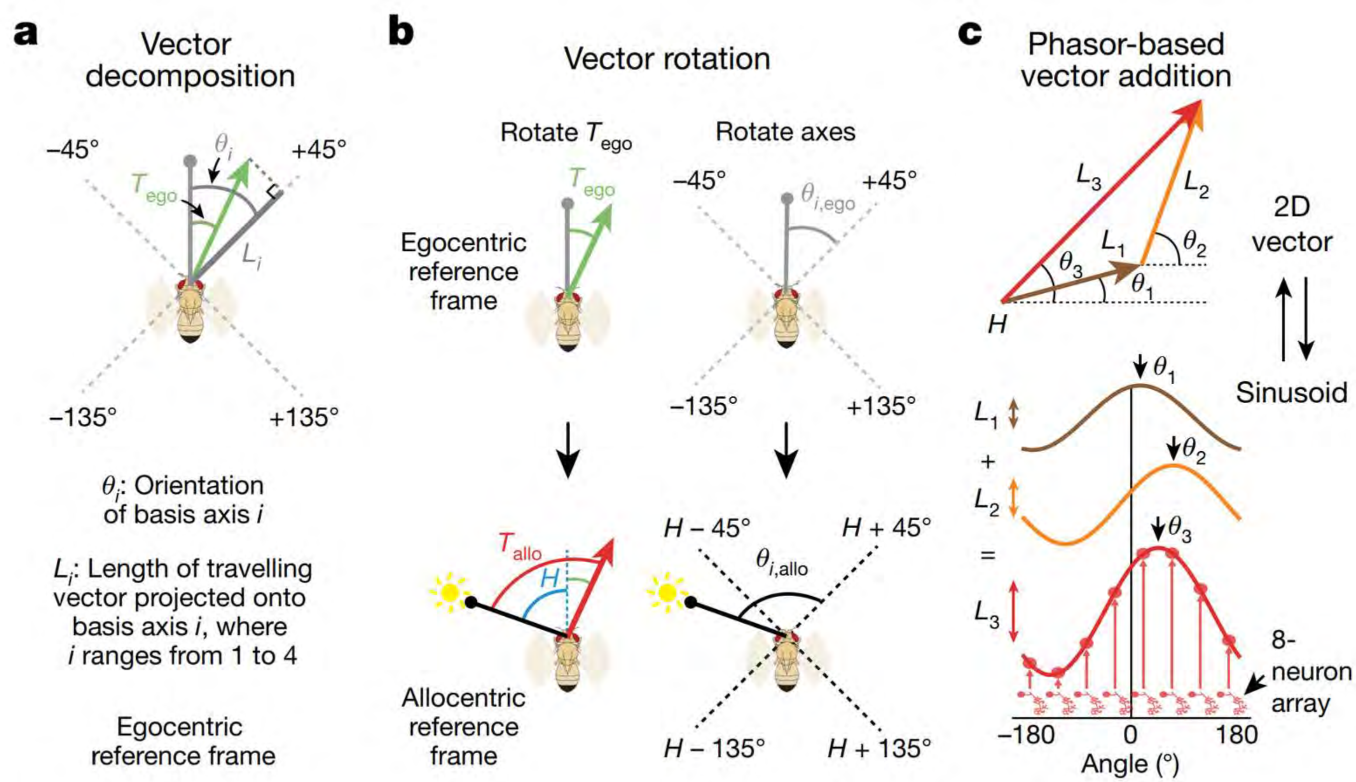


Fig. 2 | The allocentric travelling direction can be computed by vector rotation and summation, which can be implemented by phasors. a, The

Lyu, C., Abbott, L. F., & Maimon, G. (2022). Building an allocentric travelling direction signal via vector computation. *Nature*, *601*(7891), 92-97. https://doi.org/10.1038/s41586-021-04067-0

Conclusions

The concept of **neuronal population vector** is extremely useful:

- To summarize hundreds of recordings
- To pool over the variability and peculiarities of individual neurons
- Above all, to track **covert cognitive computations** (mental rotation)

Multiple questions remain open, which will be addressed in the next courses:

- How can vectors encode non-geometrical objects, such as a **face**?
- Do neurons make use of the full power of **vector spaces**: subspaces, null spaces, vector addition, etc
- Can vectors be used to track the **dynamics** of brain activity for instance during **decision making**?
- Can vectors encode **concepts**?
- Can vectors encode **sentences and syntax**?

Vendredi 6 Janvier

COURS : **Vecteurs neuronaux ou cellules grand-mère : les représentations mentales sont-elles localisées ou distribuées ?**

SÉMINAIRE : **L'intelligence artificielle peut-elle modéliser le langage mathématique ?** — François Charton (FAIR Paris)

Vendredi 13 Janvier

COURS : **Géométrie des représentations visuelles : chaque visage est un vecteur**

SÉMINAIRE : **Commonsense Physical Reasoning in man and machine** — Ernest Davis (NYU, par zoom)

Vendredi 20 Janvier

COURS: **Exploiter la factorisation et les sous-espaces vectoriels pour coder l'information et communiquer entre aires cérébrales**

SÉMINAIRE : **Number symbols in the brain and mind** — Daniel Ansari (University of Ontario)

Vendredi 27 Janvier

COURS : **Comment prendre une décision ou faire des calculs avec des vecteurs dynamiques?**

SÉMINAIRE : **Comment se développent les réseaux cérébraux associés aux concepts mathématiques ?** — Marie Amalric (Université de Trento, Italie)

Vendredi 3 Février

COURS : **La représentation vectorielle des mots et des concepts**

SÉMINAIRE : **Les succès et les nouveaux défis de l'intelligence artificielle en mathématiques** — Léon Bottou (FAIR, New York)

Vendredi 10 Février

COURS : **La représentation vectorielle du langage : Comment représenter une phrase ?**

SÉMINAIRE : **Intuitions of mathematics and their refinement with age and education** — Manuela Piazza (Université de Trento, Italie)