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°5

La représentation vectorielle des mots et des concepts

Course 5

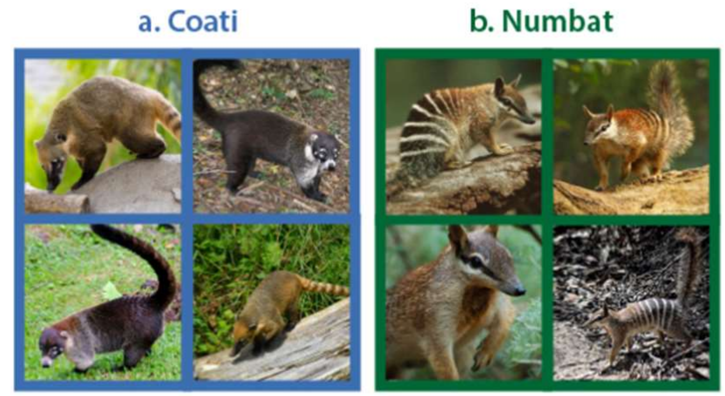
Vector representations of words and concepts

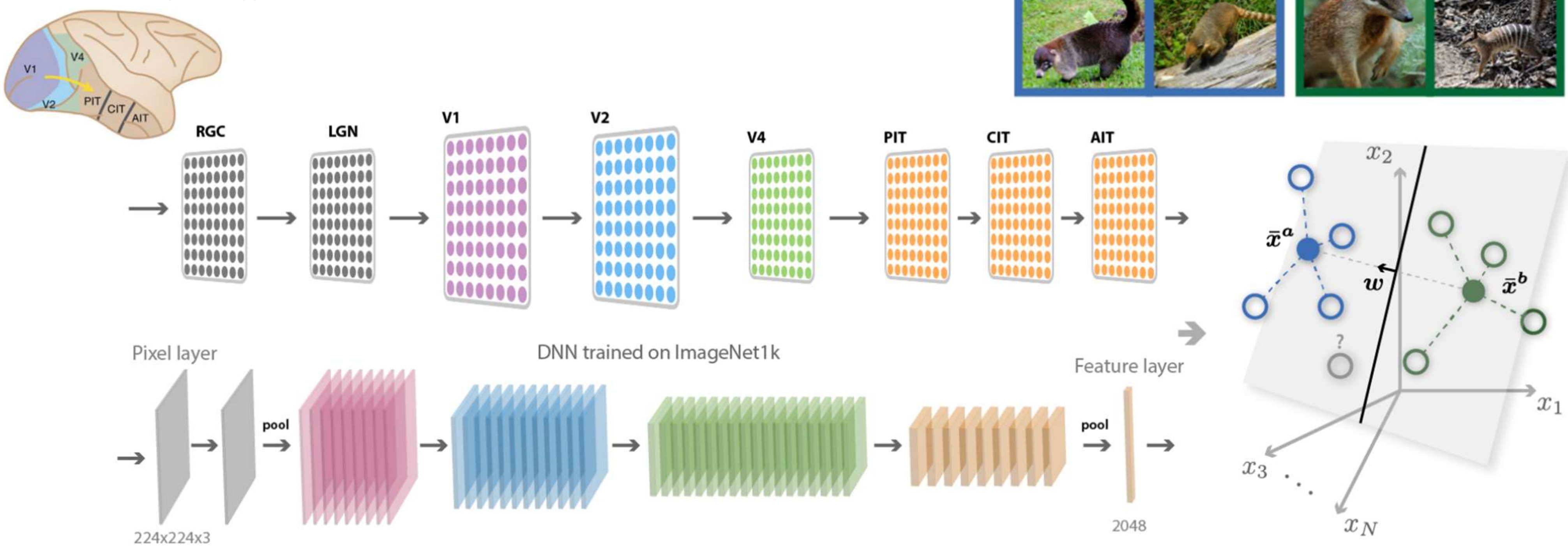
A theory of dimensionality and concept learning

Sorscher, B., Ganguli, S., & Sompolinsky, H. (2022). Neural representational geometry underlies few-shot concept learning. PNAS, 119(43), e2200800119.

Here, the authors propose a general theory of “few shot learning” for image recognition.

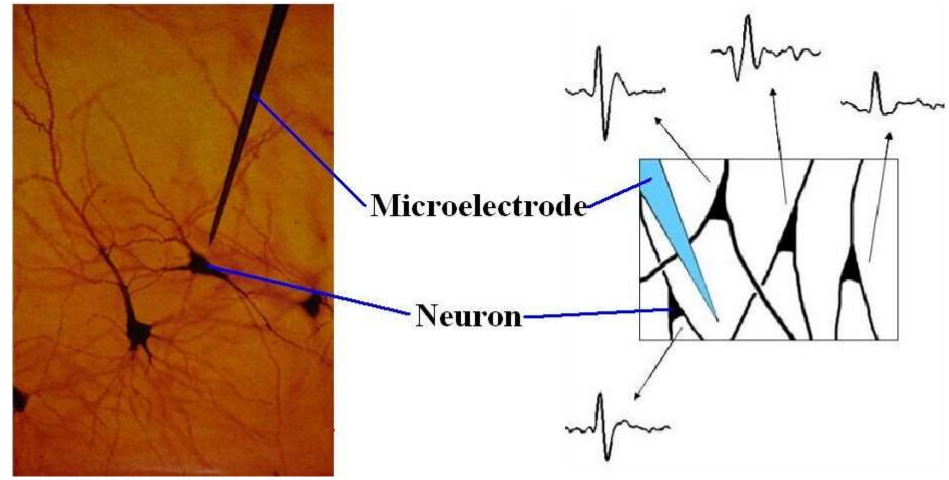
1. Prior training has resulted in a tuned high-dimensional vector space for images, which can be used to perform one-shot or few-shot learning of new concepts.
2. Each example image (possibly 1) is encoded in this high-dimensional vector space
3. The **barycenter of examples** defines a **prototypical vector** for the new concept.
4. Classification of new images, or discrimination between two possibilities, is based on the **nearest prototype**



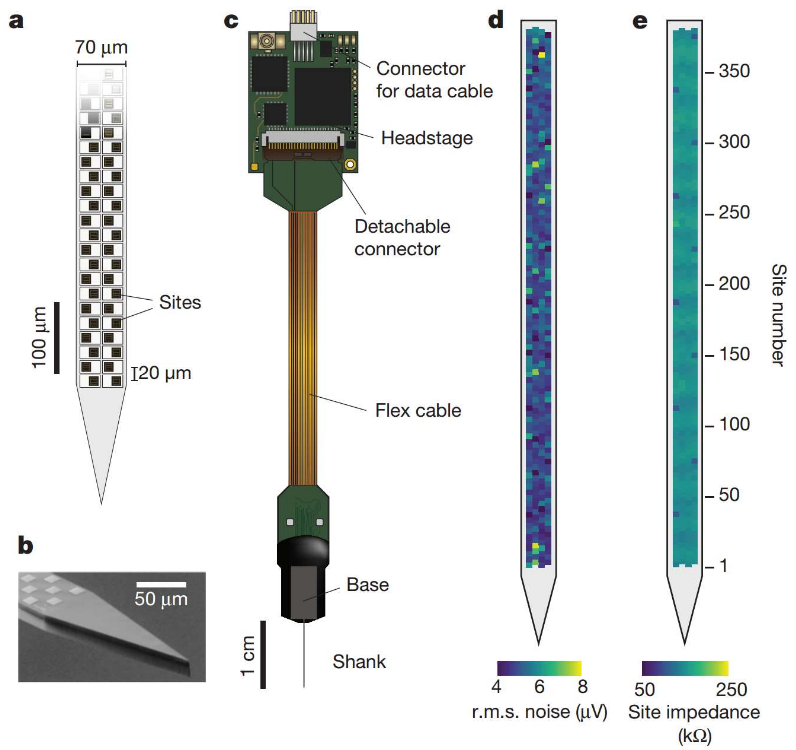


Which « neural vectors » are we talking about? A reminder of past courses

The hypothesis is that the “neural code” is supported specifically by the **discharges** of projecting nerve cells (neurons) and their joint activity. Neurons can be measured one by one, or (increasingly) in parallel across thousands of sites

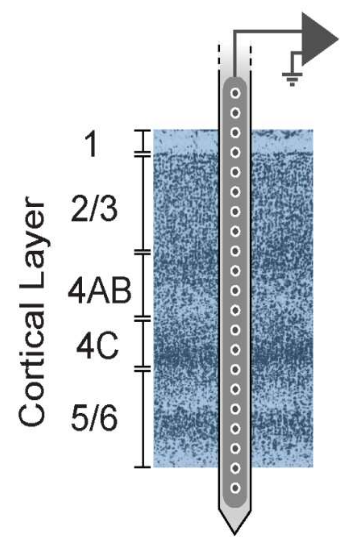


Neuropixels



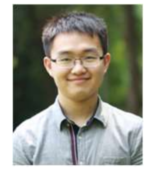
Utah arrays

Laminar electrodes



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.

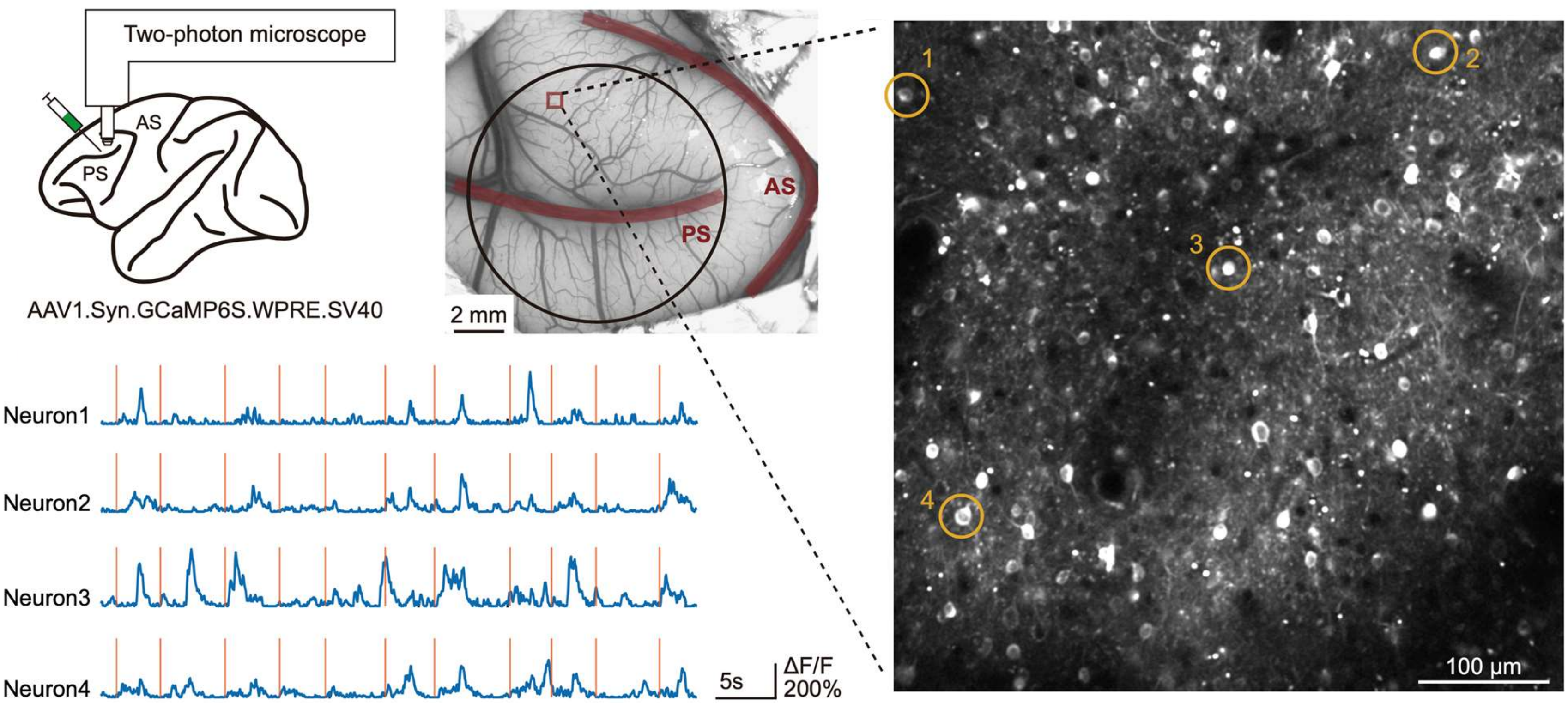

Yang Xie


Peiyao Hu

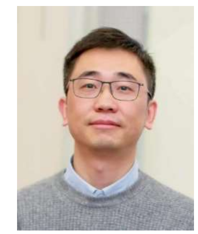

Bin Min


Shiming Tang (PKU)

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 !



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

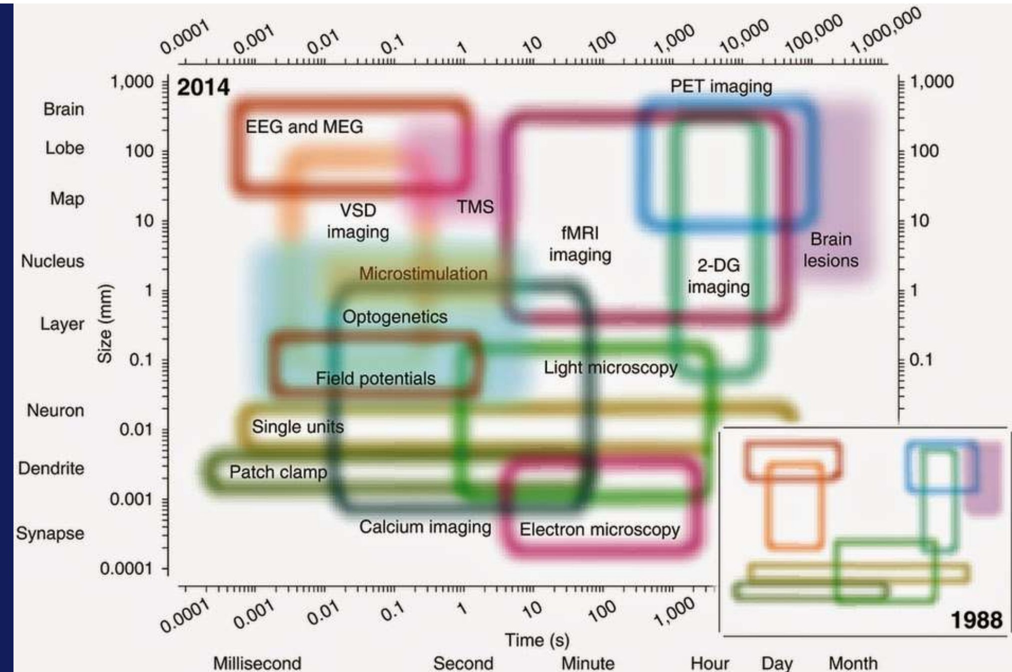

Liping Wang

In humans: methods for measuring neural vectors are much more limited

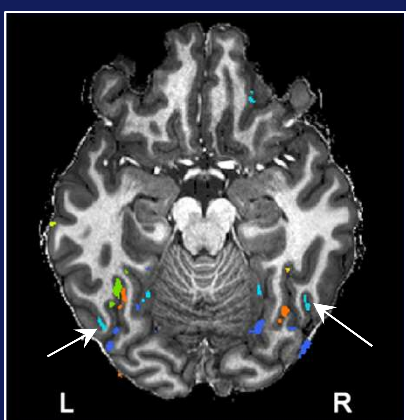
fMRI : thousands of voxels of one or a few millimeters aside (sometimes less).
1 mm² of cortical surface = ~92,000 neurons

MEG or EEG: a few hundreds of sensors.

For more details, please see my 2020 course on this topic:
https://www.college-de-france.fr/agenda/cours/progres-recents-en-imagerie-cerebrale-et-decodage-des-representations-mentales



Human MRI
at 3T, 7T, and 11.7 T



Electro- and
magneto-encephalography

EEG and 3T fMRI (soon 7T) are applicable to young children

Routing of neural information using neural subspaces

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space : Permitting preparation without movement. Nature Neuroscience, 17(3), 440-448. https://doi.org/10.1038/nn.3643
Semedo, J. D., Zandvakili, A., Machens, C. K., Yu, B. M., & Kohn, A. (2019). Cortical Areas Interact through a Communication Subspace. Neuron, 102(1), 249-259.e4. https://doi.org/10.1016/j.neuron.2019.01.026

Key proposal : **"The ability of a source area to communicate only certain activity patterns while keeping others ''private'' could be a means for the selective routing of signals between areas."**

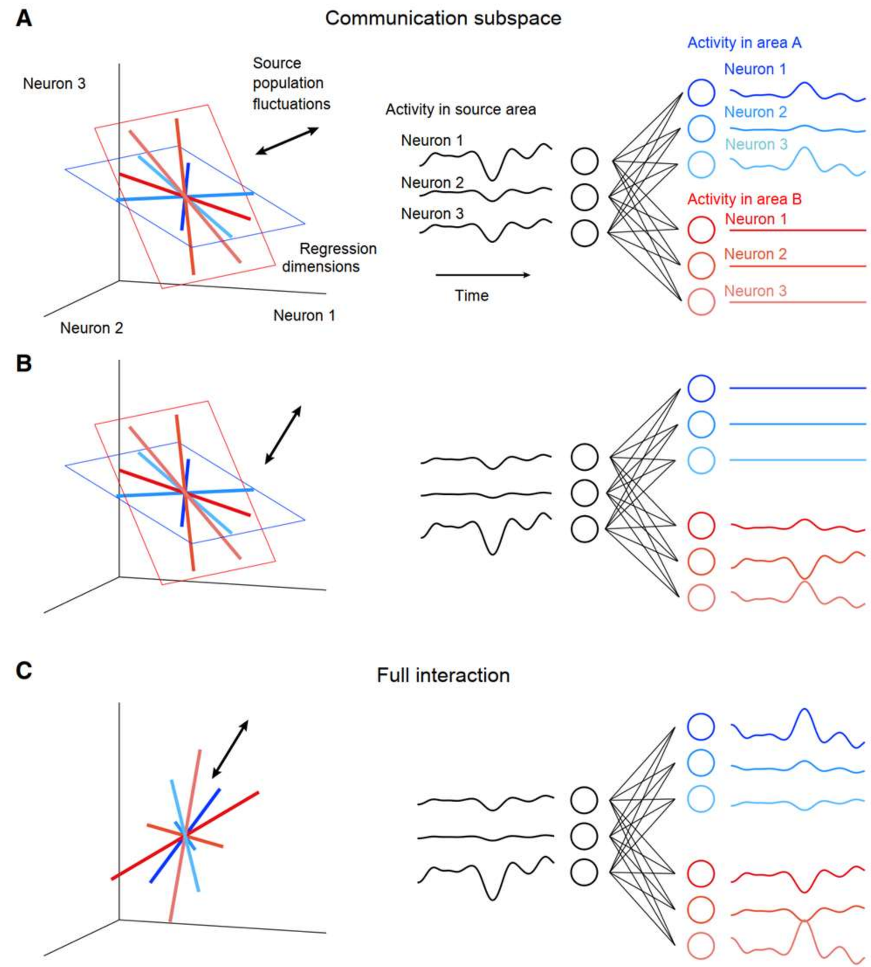
→ Completely different mechanism than the "communication through coherence" hypothesis (no need to set-up oscillations jointly in both areas)

→ More flexibility: different subspaces could be used to route information to different areas.

"The selective routing allowed by the communication subspace could be adjusted dynamically, allowing moment-to-moment modulation of interactions between cortical areas. Dynamic routing could be accomplished by altering the structure of population activity in a source area; it need not involve changing the communication subspace itself"

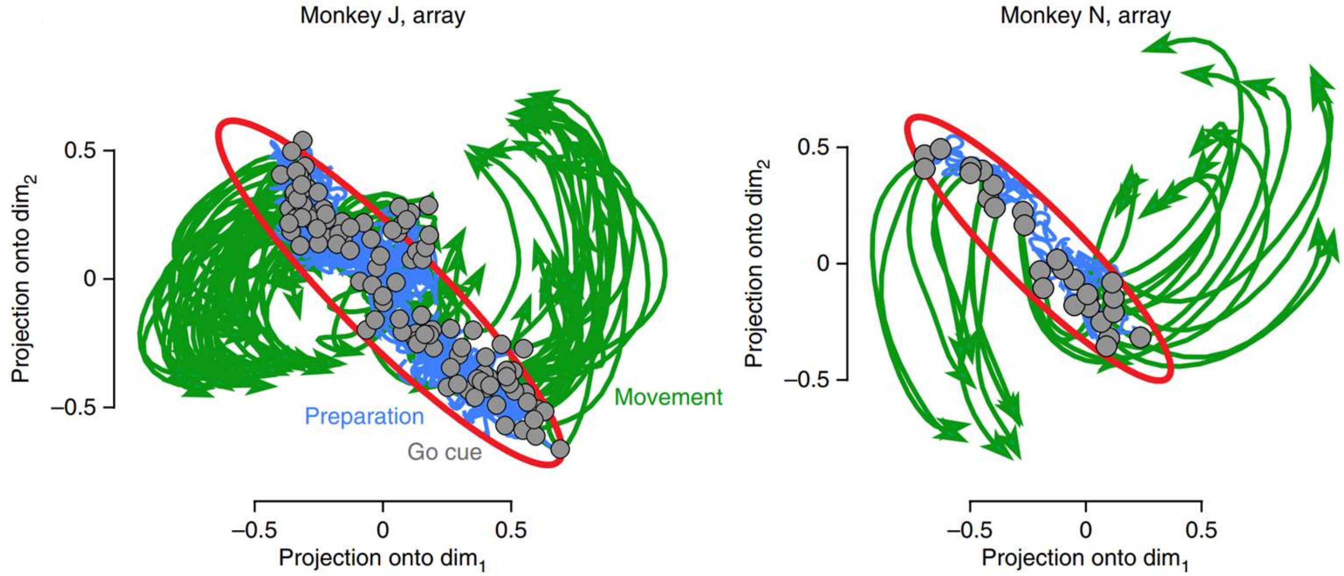
What the authors mean here is that a **projection** or **rotation** could be used to bring information to the appropriate subspace, thus opening or closing communication channels at will.

Meanwhile, the *private* dimensions (the "null space" or "kernel" for the projection to other areas) could be used to perform covert computations.



Movement preparation in the null space

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space : Permitting preparation without movement. Nature Neuroscience, 17(3), 440-448. https://doi.org/10.1038/



During movement preparation, the null space for muscle activity in M1 and premotor cortex seems to be used to prepare the upcoming movement.

For instance, in the graph at right, each gray dot represents one average condition of movement through the maze. Along the blue axis, all of their preparatory activity projects identically – but during the movement, their activity vectors rotate and predict movement.

Note that preparatory activity is far from random or noisy – it is predictive of the amplitude and orientation of the upcoming movement.

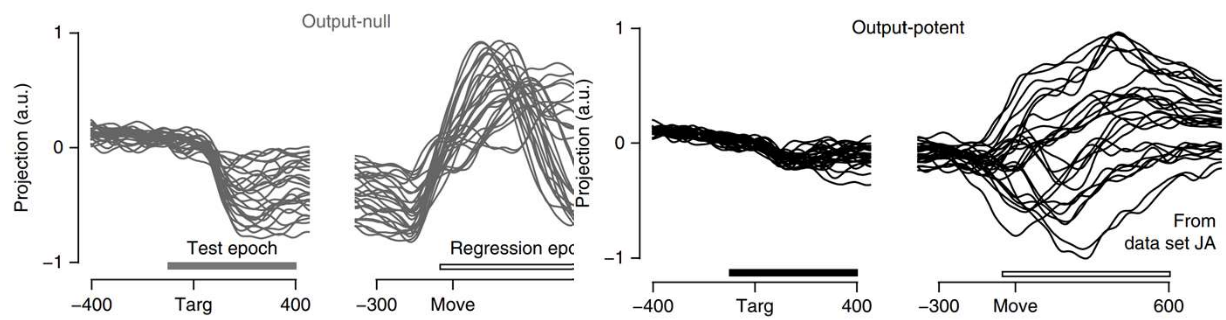
How can we identify the output space and null space?

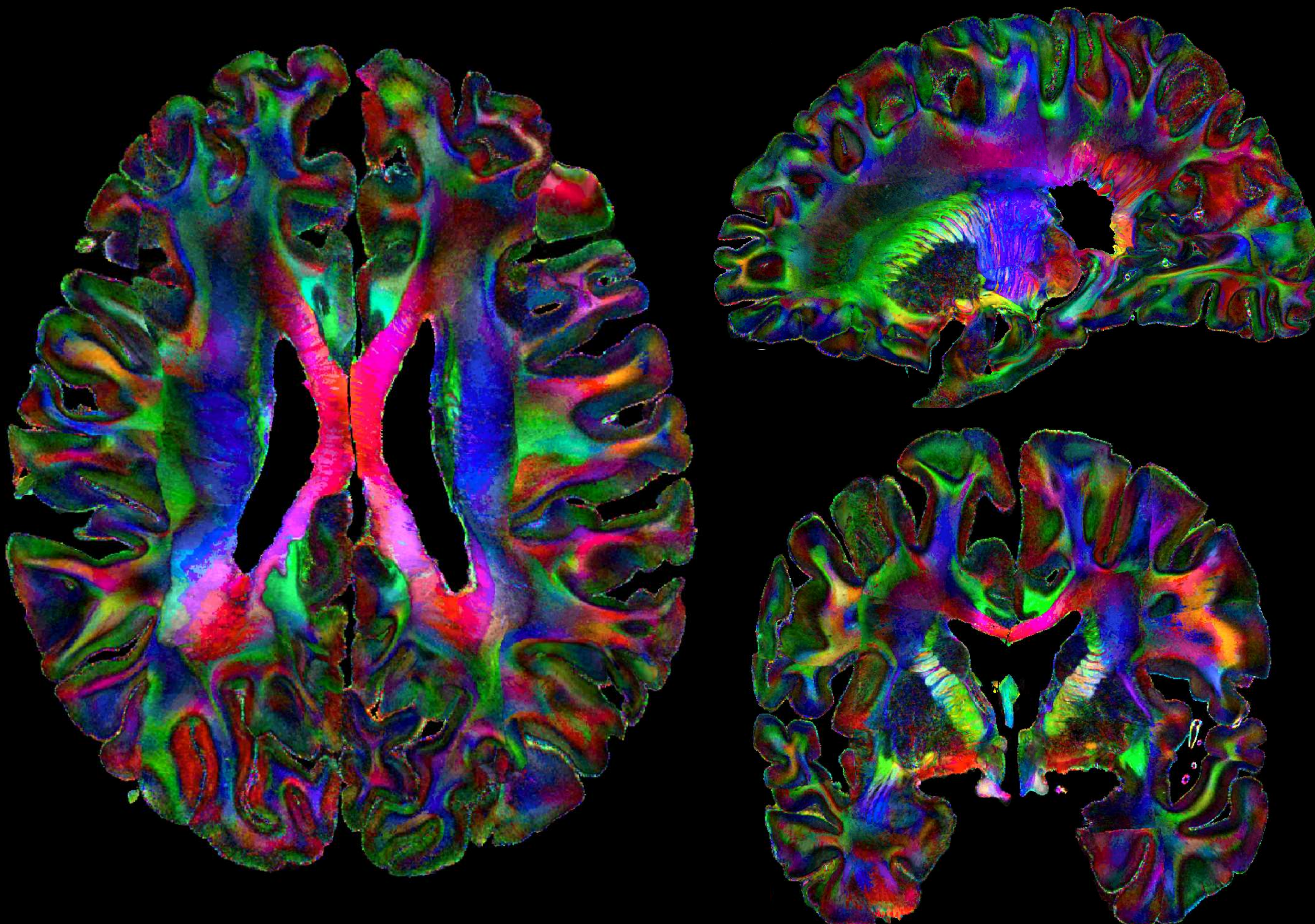
First, perform data reduction using principal component analysis, thus reducing muscle activity M to $\widetilde{M}$ (with 3 dimensions), and neural activity N to $\widetilde{N}$ (with 6 dimensions).

Then use regression to solve for $\widetilde{W}$ such that $\widetilde{M} = \widetilde{W}\widetilde{N}$

And finally project the entire trajectory onto (1) the null space of $\widetilde{W}$, and (2) the other orthogonal space.

Results: potent preparatory activity in the output-null dimensions of PMd and M1 (left), 3 to 8 times larger than in the output-potent dimensions (right).





Chenonceaux project
Image by Cyril Poupon

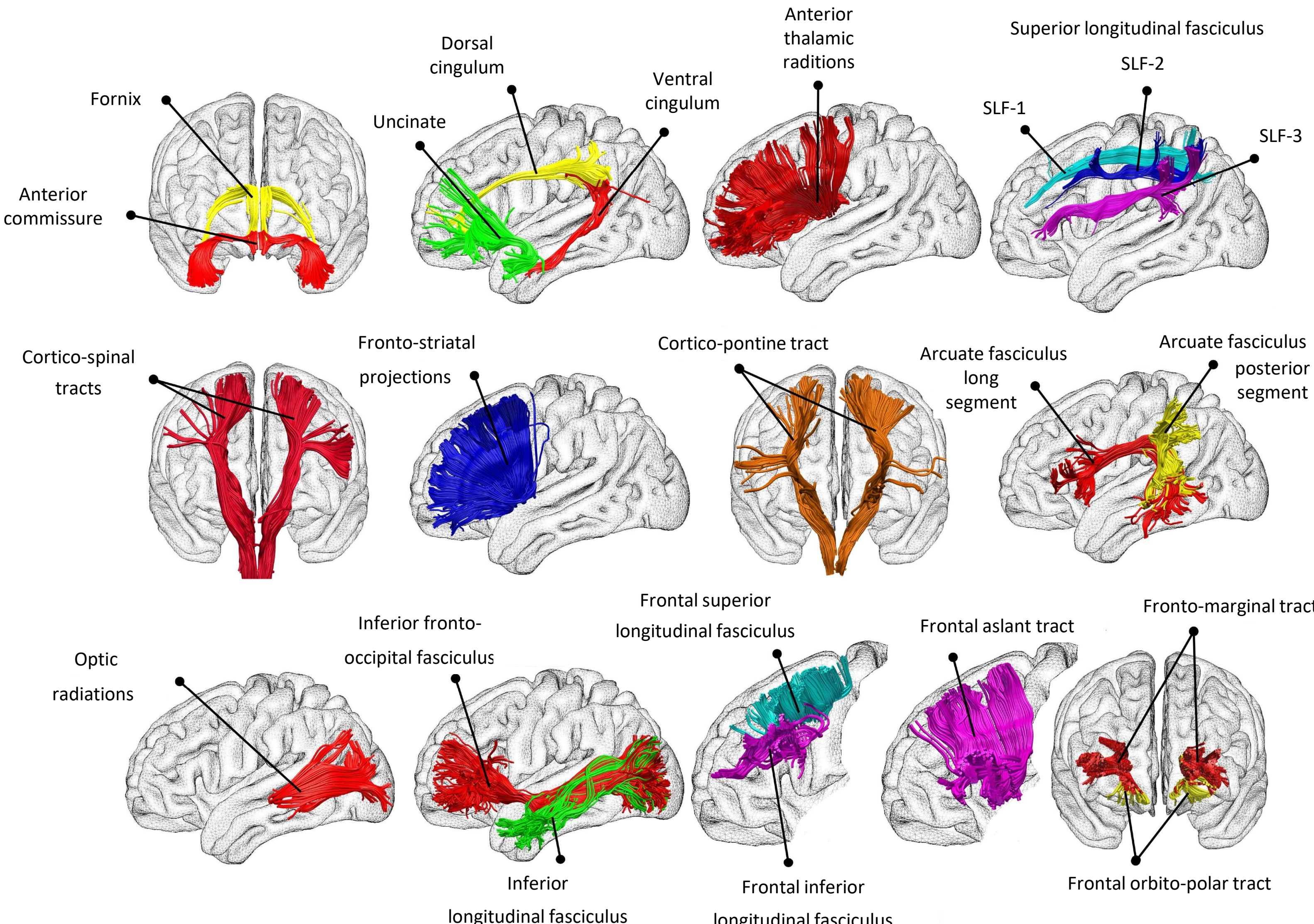
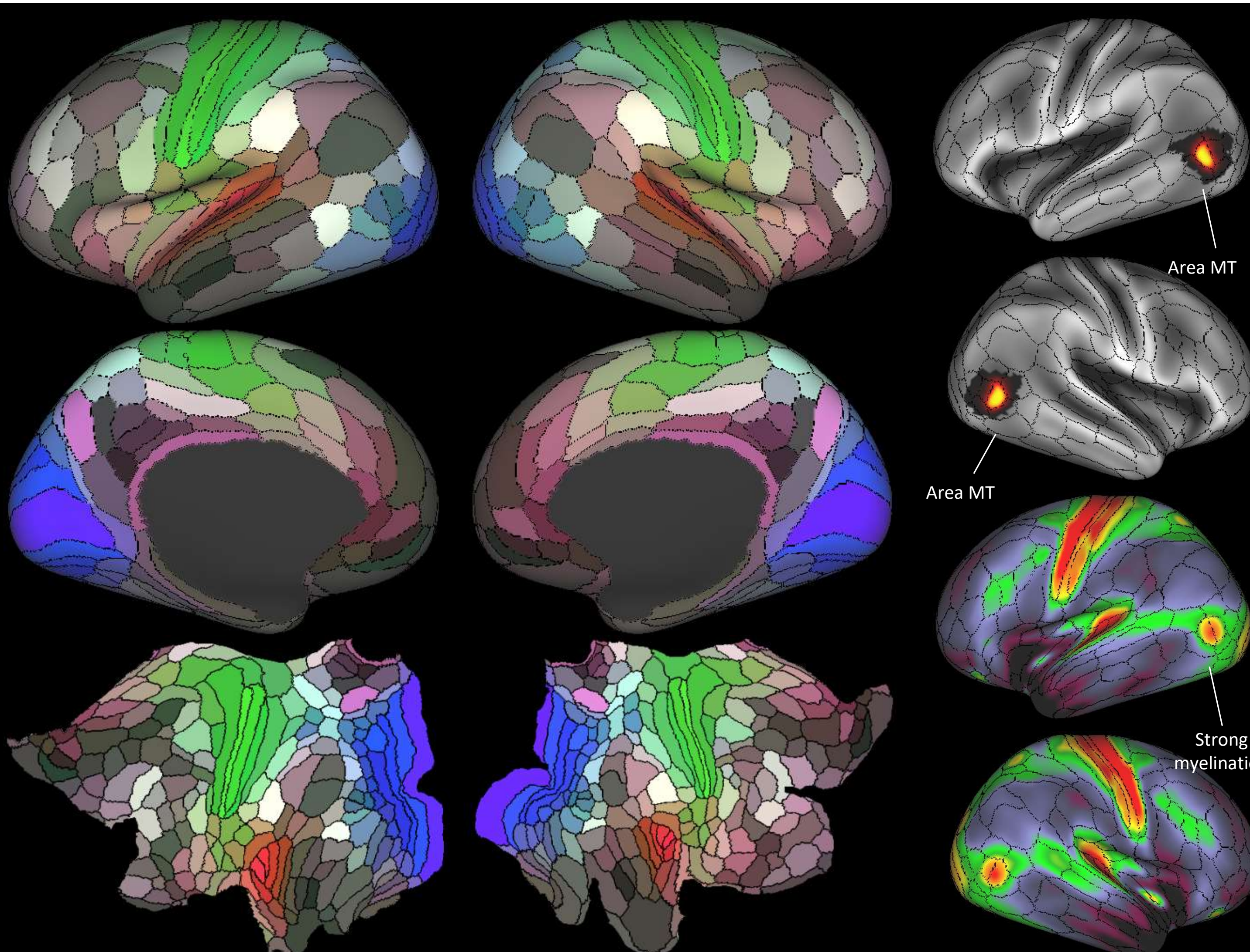


Image by Michel Thiebaut de Schotten



After Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. Nature, 536(7615), 171-178. https://doi.org/10.1038/nature18933

What is the size of the communication channel between human cortical areas ?

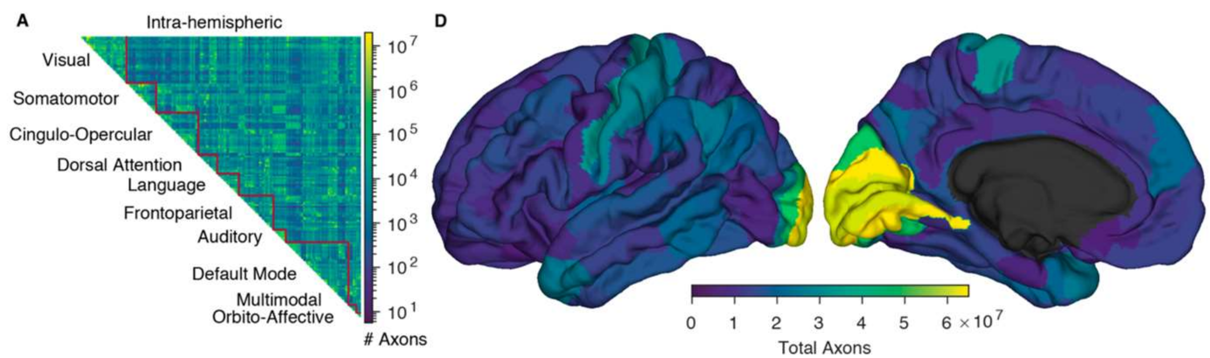
Rosen, B. Q., & Halgren, E. (2022). *Human cortical areas are sparsely connected : Combining histology with diffusion MRI to estimate the absolute number of axons* (p. 2021.06.07.447453). https://doi.org/10.1101/2021.06.07.447453

Context : in the human brain, there are ~16 billion cortical neurons (92,000 neurons per mm² of cortex) ; ~10,000 synapses per neuron.

« As the number of cortical neurons increases, maintaining the same probability of connectivity between neurons would require that axon number increase approximately with the square of neuron number, and this would require too much **volume**, impose an unsustainable **metabolic load** [6], and actually decrease computational power due to **conduction delays** [7]. The consequent imperative to minimize long distance cortico-cortical fibers has been posited to be reflected in **exponential decline in cortical connectivity with distance** [8], and to be partially compensated for with a **small-world graph architecture** [9], granting special properties to **rare long-distance fibers** in a **log-normal** neural physiology and anatomy [10].”

R&H estimate that there **370,000 axons per mm² of corpus callosum**, and use such estimates to calibrate diffusion tensor images.

The result is that each of the HCP parcels receives a very small number of axons from other areas:



For instance, “the connections between Wernicke’s and Broca’s areas are thought to integrate receptive and expressive aspects of language, but we estimate that there are only ~58,000 axons between the core cortical parcels in these regions”

This is “fewer than two for each word in an average university student’s vocabulary” ! [the comparison is absurd!]

Constraints on human cortical interactions

Rosen, B. Q., & Halgren, E. (2022). *Human cortical areas are sparsely connected : Combining histology with diffusion MRI to estimate the absolute number of axons* (p. 2021.06.07.447453). https://doi.org/10.1101/2021.06.07.447453

Conclusion: areas must talk to each other using a sparse set of axons, in the order of 100,000 output axons between any two cortical regions.

Only 100,000 neurons in a given area are **output cells** projecting onto a given target area → communication bottleneck.

Given ~180 areas in each hemisphere, each must contain ~50 million neurons, of which about 1/500 can communicate with another area.

Implications :

→ Actually, no difficulty for a distributed code :
  - A single-cell code would have a very low chance of getting transmitted outside the area (roughly 1/500)
  - But a distributed code would have a much larger chance of contacting at least one or several of those output cells.
  - The number of communicable codes remains huge: 2^50,000 binary states, or 50,000 superimposable (factorized) dimensions

→ Existence of a **vast null space** capable of processing information which is not transmitted to other areas.

The authors also draw the following, much more debatable conclusions :

→ Most of the computations must be done locally (?)

→ Information is heavily compressed during transmission (?) – or else it must be lost.

How many samples suffice to communicate a neural manifold ? Example of IT

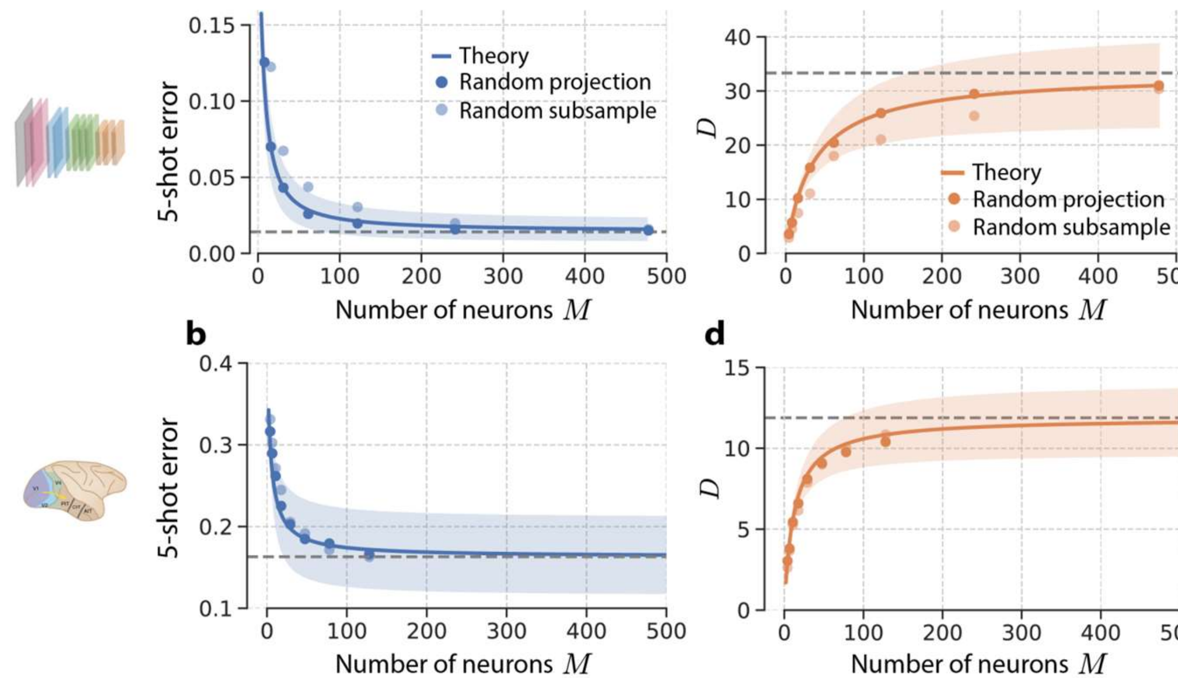
Sorscher, B., Ganguli, S., & Sompolinsky, H. (2022). Neural representational geometry underlies few-shot concept learning. PNAS, 119(43), e2200800119.

Using the theory of random projections, the authors manage to compute the effect of subsampling a small number of neurons (M).
“subsampling causes distortions in the manifold geometry that decrease both the SNR and the estimated dimensionality, as a function of the number of recorded neurons M”:

$$\mathrm{SNR}(M) = \frac{\mathrm{SNR}_\infty}{\sqrt{1 + D_\infty / M}}, \quad D^{-1}(M) = D_\infty^{-1} + M^{-1}$$

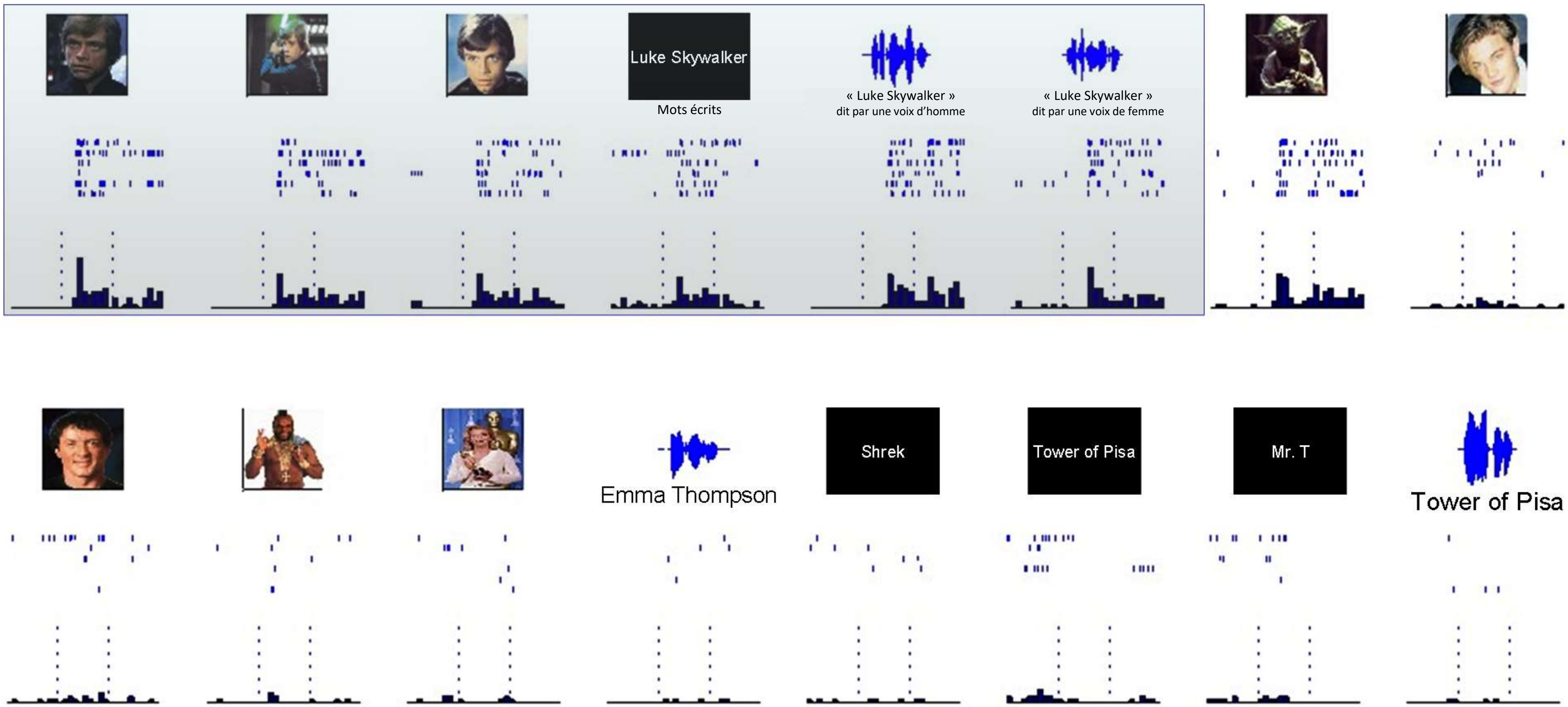
“These distortions are negligible when M is large compared to the asymptotic dimensionality D∞.
In both macaque IT and a trained DNN model (Fig. 6), a downstream neuron receiving inputs from only about 200 neurons performs essentially similarly to a downstream neuron receiving inputs from all available neurons.
With recordings of about 200 IT neurons, the estimated dimensionality approaches its asymptotic value (D∞ ≈ 35 in the trained DNN and D∞ ≈ 12 in macaque IT).”

These small dimensionalities are compatible with Doris Tsao’s work on faces, where D=50 suffices to encode any face.



And for word meanings? Is the neural vector hypothesis still appropriate?

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.



The geometry of ideas and concepts : an old idea in cognitive science

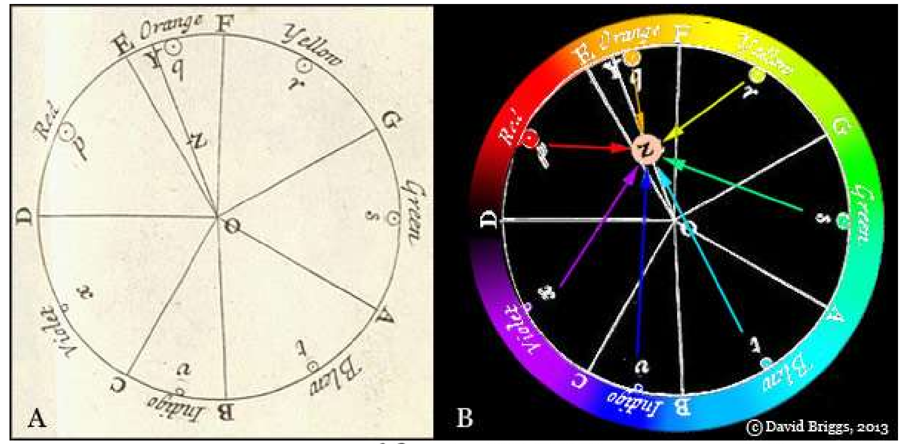
Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, *210*(4468), 390-398. https://doi.org/10.1126/science.210.4468.390

First insight: ideas and concepts are like points in an abstract space.

Isaac Newton (1704) already conceived of colors as a circle.

Shepard produced a variety of mathematical algorithms to objectivity find the optimal way to **plot concepts in space**, such that their proximities reflect psychological similarity.

Non-metric multidimensional scaling is an algorithm that starts with a similarity matrix between some items, and places them in space such that their similarities are a monotonic function of Euclidean distance.

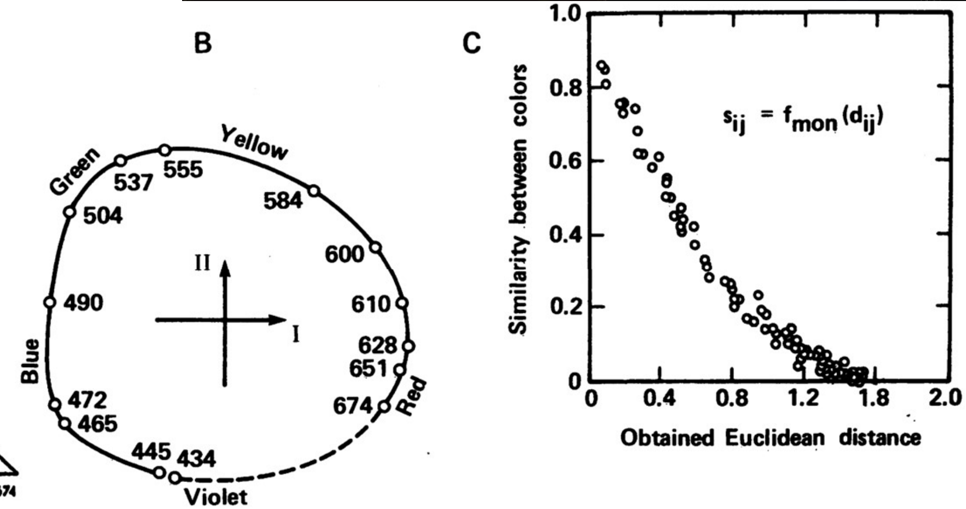


A

Judged similarities between colors (Ekman)

Wavelength (nanometers)	434	445	465	472	490	504	537	555	584	600	610	628	651
434													
445	.86												
465	.42	.50											
472	.42	.44	.81										
490	.18	.22	.47	.54									
504	.06	.09	.17	.25	.61								
537	.07	.07	.10	.10	.31	.62							
555	.04	.07	.08	.09	.26	.45	.73						
584	.02	.02	.02	.02	.07	.14	.22	.33					
600	.07	.04	.01	.01	.02	.08	.14	.19	.58				
610	.09	.07	.02	.00	.02	.02	.05	.04	.37	.74			
628	.12	.11	.01	.01	.01	.02	.02	.03	.27	.50	.76		
651	.13	.13	.05	.02	.02	.02	.02	.02	.20	.41	.62	.85	
674	.16	.14	.03	.04	.00	.01	.00	.02	.23	.28	.55	.68	.76

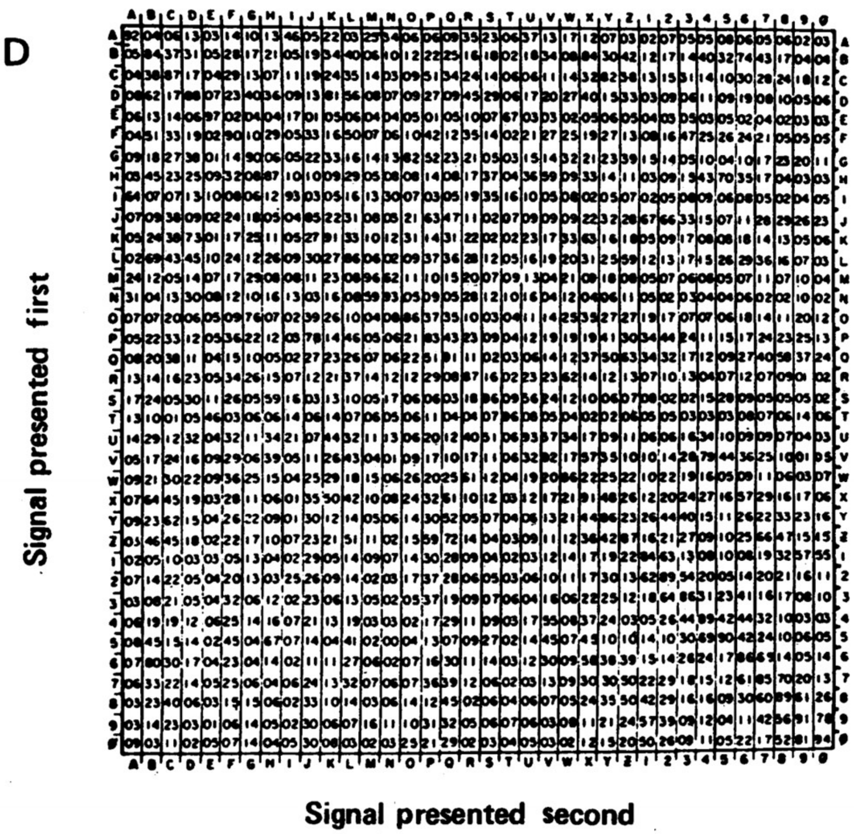
Wavelength (nanometers)

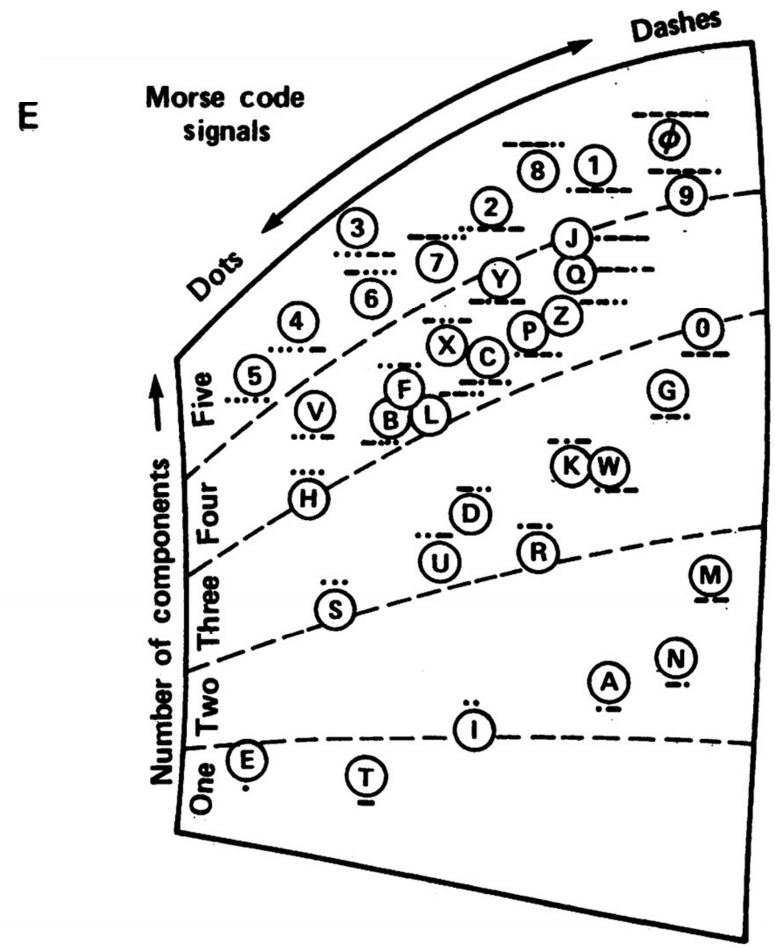


The geometry of ideas and concepts : an old idea in cognitive science

Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, *210*(4468), 390-398. https://doi.org/10.1126/science.210.4468.390

Non-metric multidimensional scaling of the Morse code

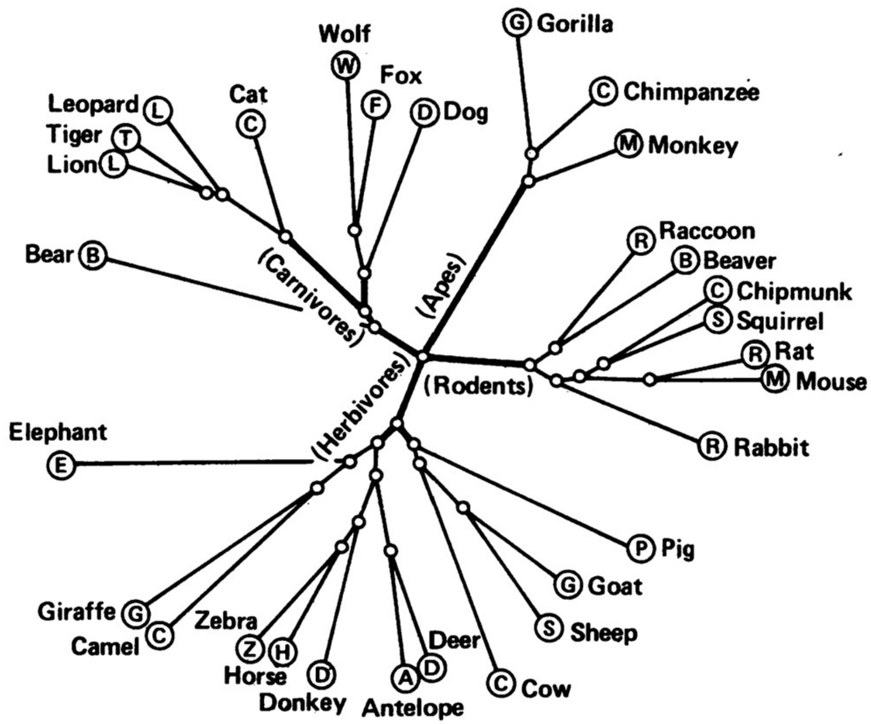




The geometry of ideas and concepts : an old idea in cognitive science

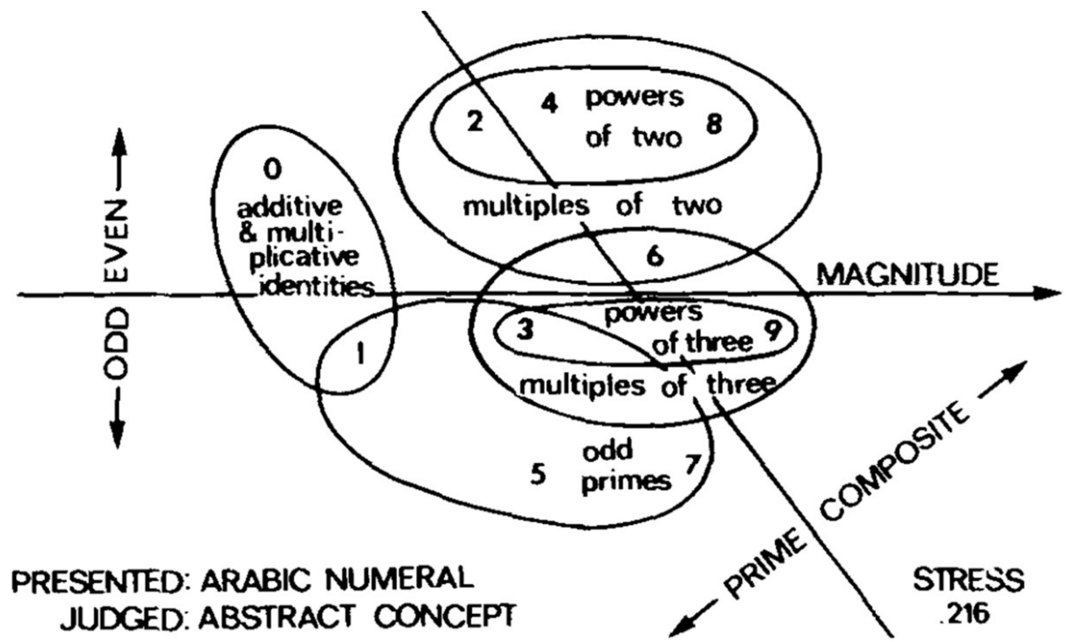
Shepard, R. N. (1980). Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, *210*(4468), 390-398. https://doi.org/10.1126/science.210.4468.390

Extension of the idea to tree structures:
automatic construction of a tree whose distances explain the similarities among 30 animal species



Application of multi-dimensional scaling to **number concepts**

Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. Cognitive Psychology, 7, 82-138.



Second insight: words can be *defined* by their neighboring words

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Multi-dimensional scaling (MDS) requires the dissimilarity between any two concepts.

→ we need a huge matrix of similarity ratings (n^2)

Others have tried to define concepts by a list of their features (e.g. « has feathers », « can be found in the kitchen »)

→ Linear in the number of concepts, but requires a list of human-generated features.

Latent semantic analysis (LSA) uses a different idea: « You shall know a word by the company it keeps » (John Rupert Firth, 1957)

→ words with a similar meaning tend to occur in similar contexts

In LSA, a word is represented as a semantic vector in a high-dimensional space, where similarity between word vectors reflects similarity of the contexts in which those words appear

Method = Count how often each given word occurs in a given context (here a given article of the Grollier encyclopedia)

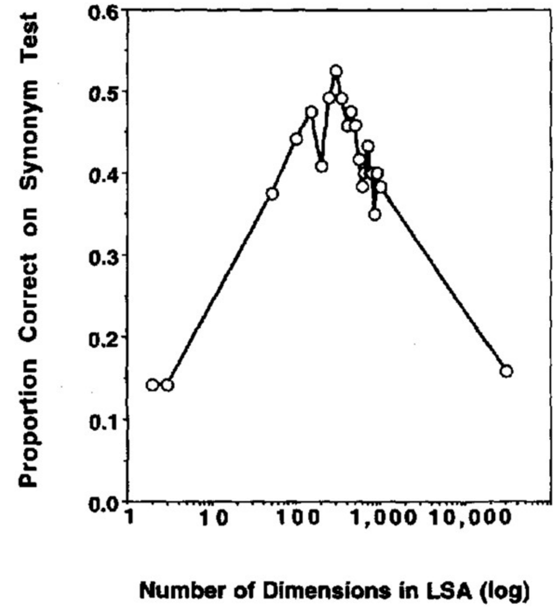
"The text data were cast into a matrix of 30,473 columns, each column representing one text sample, by 60,768 rows, each row representing a unique word type that appeared in at least two samples"

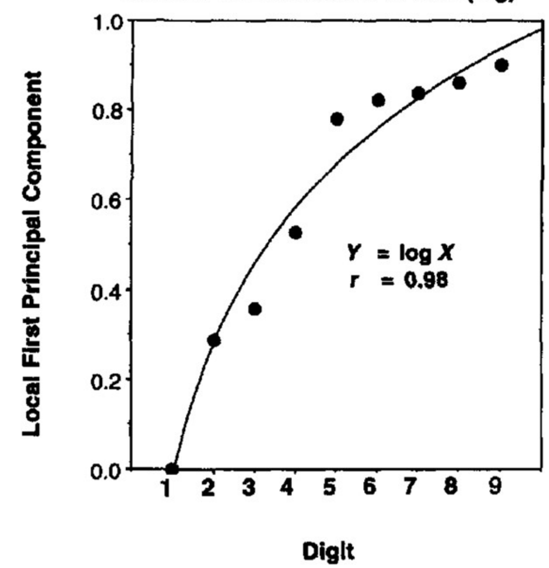
Then perform dimensionality reduction on that matrix, keeping e.g. only the first 300 dimensions.

TOEFL test: for each word, the model was given one synonym and 4 unrelated words.

→ Cosine similarity was able to pick the synonym ~50% of time.

Numbers: the cosine distances, once processed with MDS, yield a main dimension of magnitude.





Word2vec embedding

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic regularities in continuous space word representations. Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies, 746-751.

https://www.tensorflow.org/tutorials/text/word2vec?hl=en

https://web.stanford.edu/class/cs224n/readings/cs224n-2019-notes01-wordvecs1.pdf

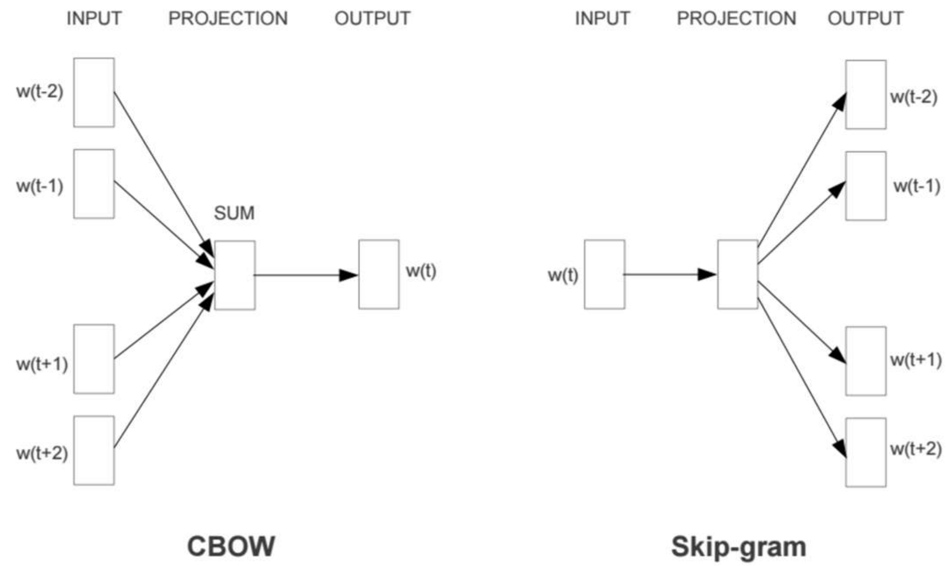
A whole family of methods whose main advantage is that we no longer need to compute with huge matrices.

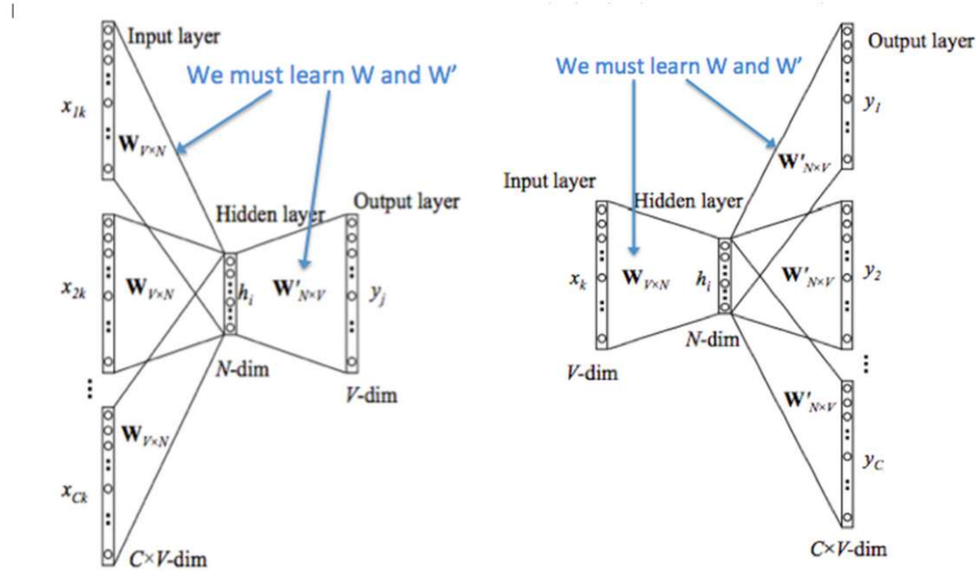
The algorithm finds a vector encoding for words by analyzing the other words that surround it in the corpus (within a certain window size).

e.g. “Je heure me bonne suis longtemps”
→ couché

There are two version: Continuous bag of words (CBOW) and Skip-Gram. Either the model uses the current word to try to predict its neighbors, or the converse.

Several variants of the training method (cost function) are available, including negative sampling (~estimate the probability that the sample came from the training set) and hierarchical softmax (first introduced by Morin and Bengio).





Implicit learning of many interesting relationships (vector subspaces)

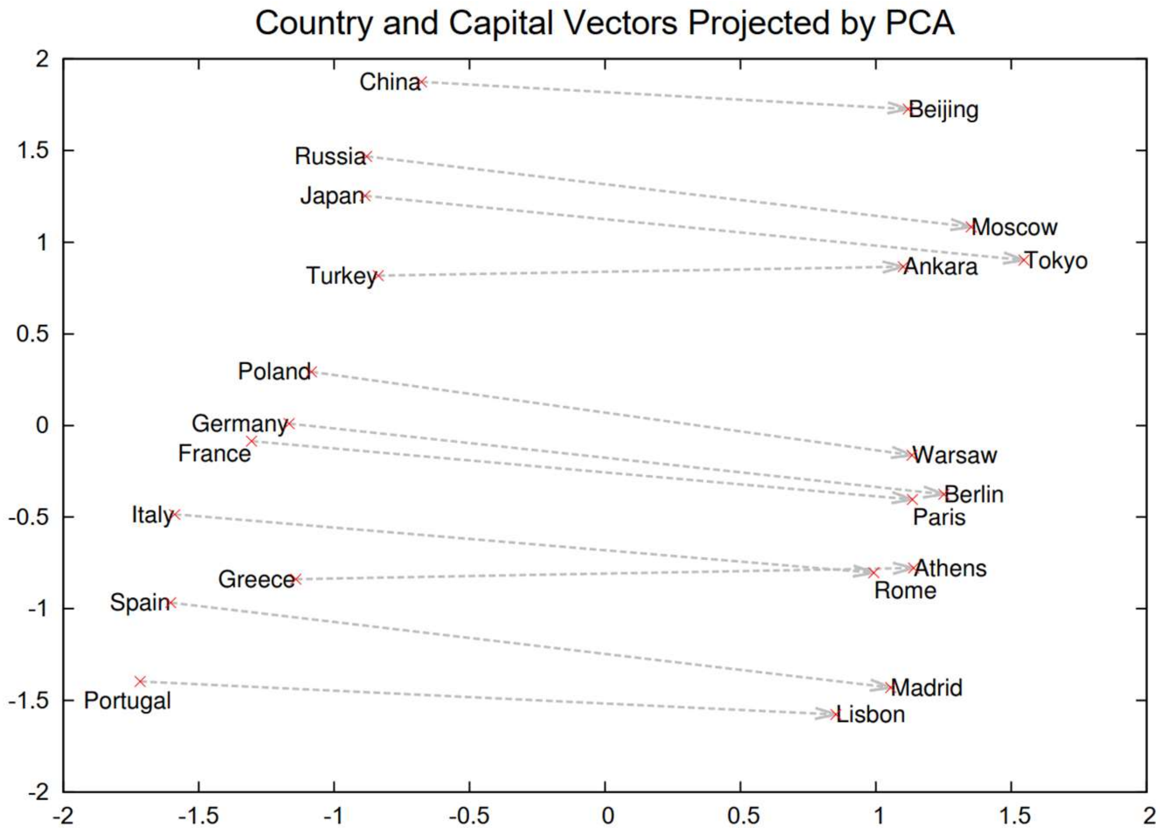
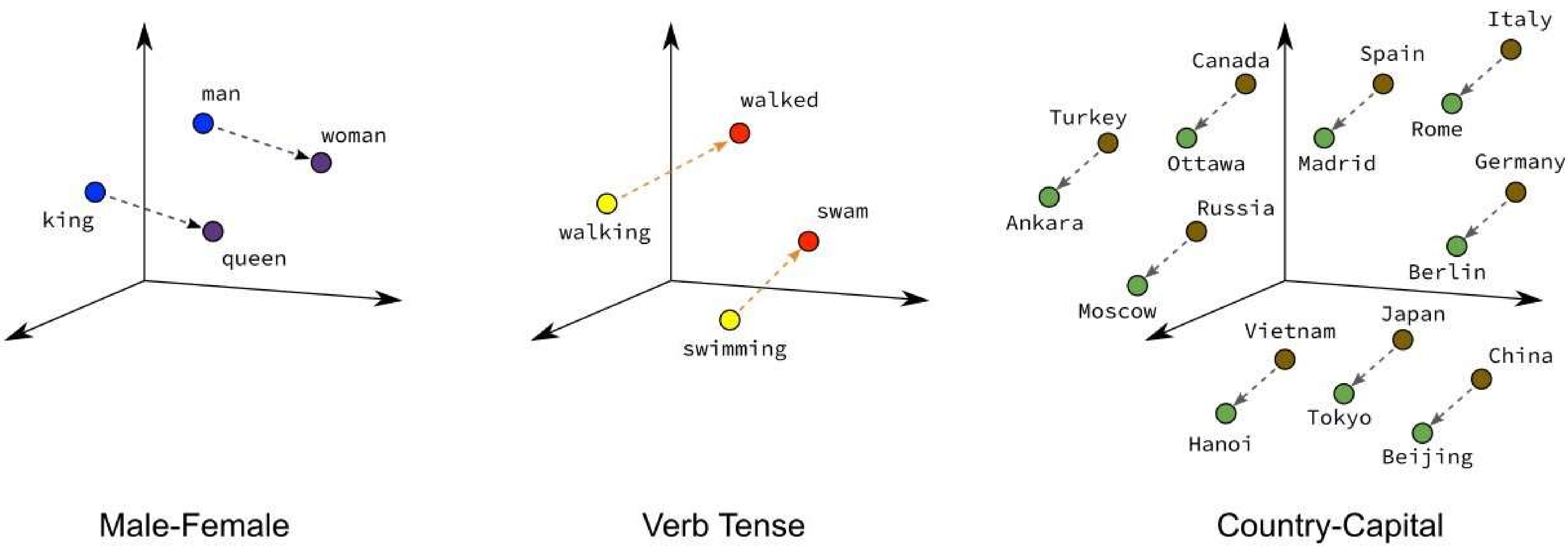


Figure 2: Two-dimensional PCA projection of the 1000-dimensional Skip-gram vectors of countries and their capital cities. The figure illustrates ability of the model to automatically organize concepts and learn implicitly the relationships between them, as during the training we did not provide any supervised information about what a capital city means.

Word2vec embeddings comprise lots of specific syntactic and semantic vectors

https://developers.google.com/machine-learning/crash-course/embeddings/translating-to-a-lower-dimensional-space



Glove embedding: a more principled, mathematical approach to word vectors

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe : Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. https://nlp.stanford.edu/projects/glove/

" The main intuition underlying the model is the simple observation that ratios of word-word co-occurrence probabilities have the potential for encoding some form of meaning. For example, consider the co-occurrence probabilities for target words *ice* and *steam* with various probe words from the vocabulary. Here are some actual probabilities from a 6 billion word corpus:

Probability and Ratio	$k = solid$	$k = gas$	$k = water$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}
$P(k\|ice)/P(k\|steam)$	8.9	8.5×10^{-2}	1.36

"As one might expect, *ice* co-occurs more frequently with *solid* than it does with *gas*, whereas *steam* co-occurs more frequently with *gas* than it does with *solid*. Both words co-occur with their shared property *water* frequently, and both co-occur with the unrelated word *fashion* infrequently. Only in the ratio of probabilities does noise from non-discriminative words like *water* and *fashion* cancel out, so that large values (much greater than 1) correlate well with properties specific to ice, and small values (much less than 1) correlate well with properties specific of steam. In this way, the ratio of probabilities encodes some crude form of meaning associated with the abstract concept of thermodynamic phase."

"The training objective of GloVe is to learn word vectors such that their dot product equals the logarithm of the words' probability of co-occurrence". More precisely:

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

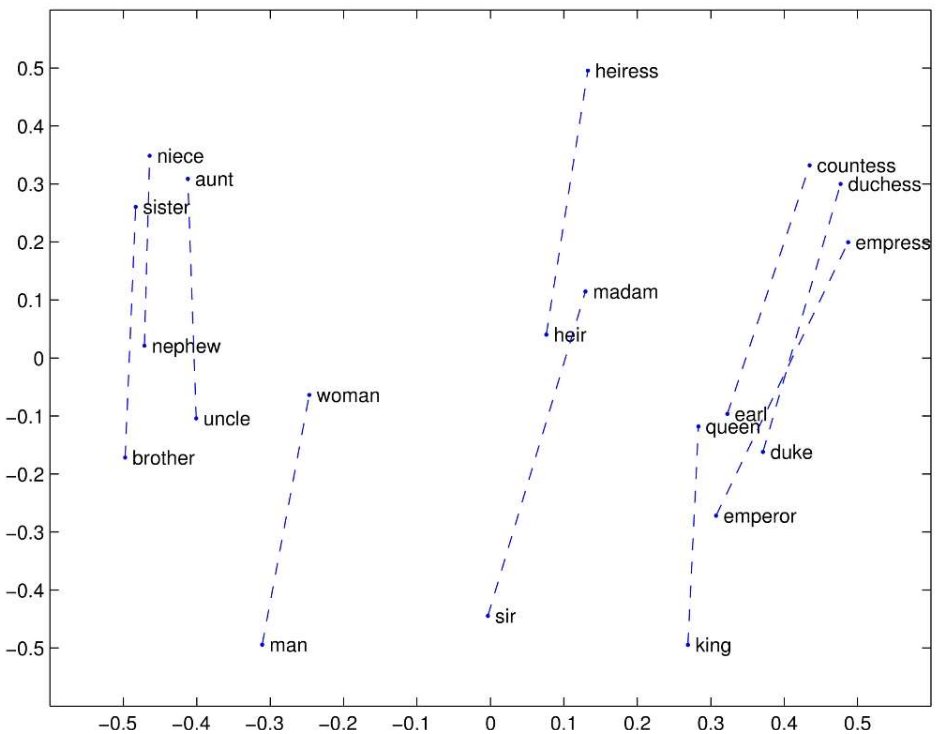
The solution to Eqn. (4) is $F = \exp$, or,

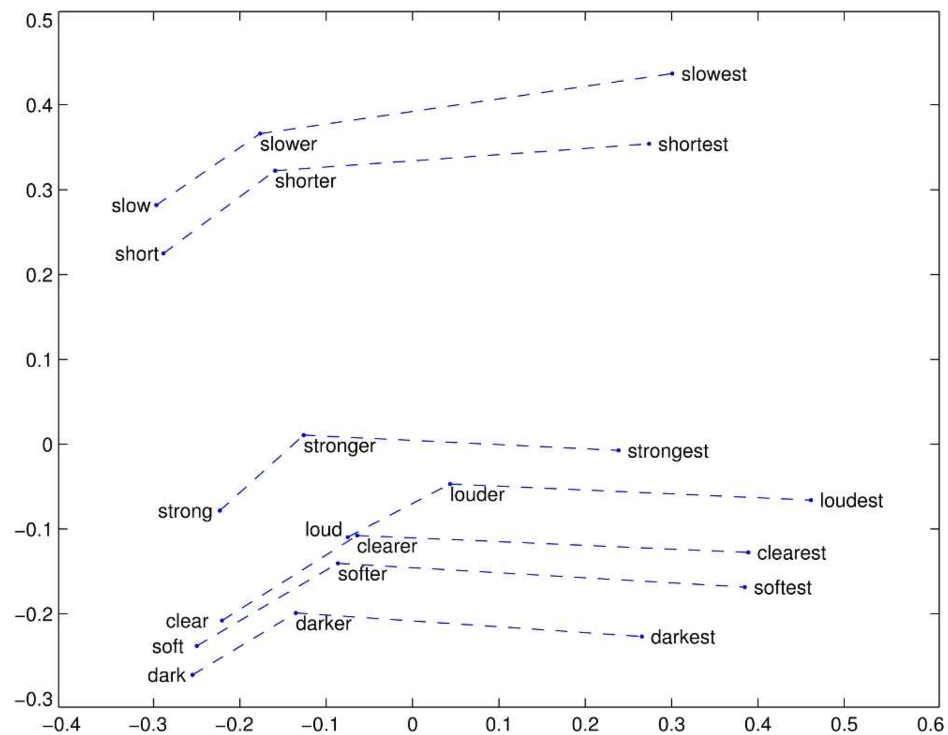
$$w_i^T \tilde{w}_k = \log(P_{ik}) = \log(X_{ik}) - \log(X_i)$$

Owing to the fact that the logarithm of a ratio equals the difference of logarithms, this objective associates (the logarithm of) ratios of co-occurrence probabilities with vector differences in the word vector space. Because these ratios can encode some form of meaning, this information gets encoded as vector differences as well."

Glove embeddings also discover syntactic and semantic relations among words

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe : Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing (EMNLP)*, 1532-1543. https://nlp.stanford.edu/projects/glove/





Can vector semantics account for human similarity judgments?

Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, *33*(3-4), 175-190. https://doi.org/10.1080/02643294.2016.1176907

Pereira et al. examine whether a variety of vector representations of word meaning can explain 7 databases on human word similarity judgments.

Glove and Word2vec representations achieve similarly high degrees of prediction of human judgments.

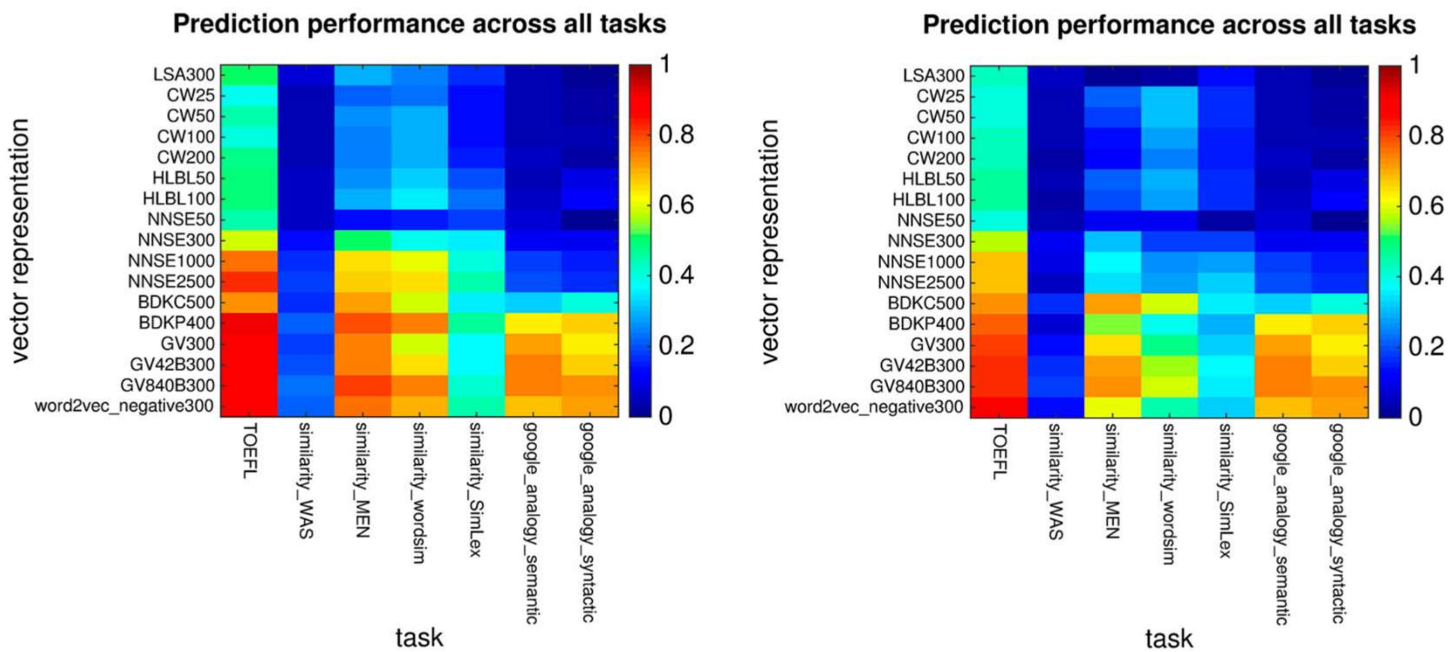


Figure 1. Performance of predictions generated from various vector representations across all tasks available, using cosine similarity (left) or euclidean distance (right).

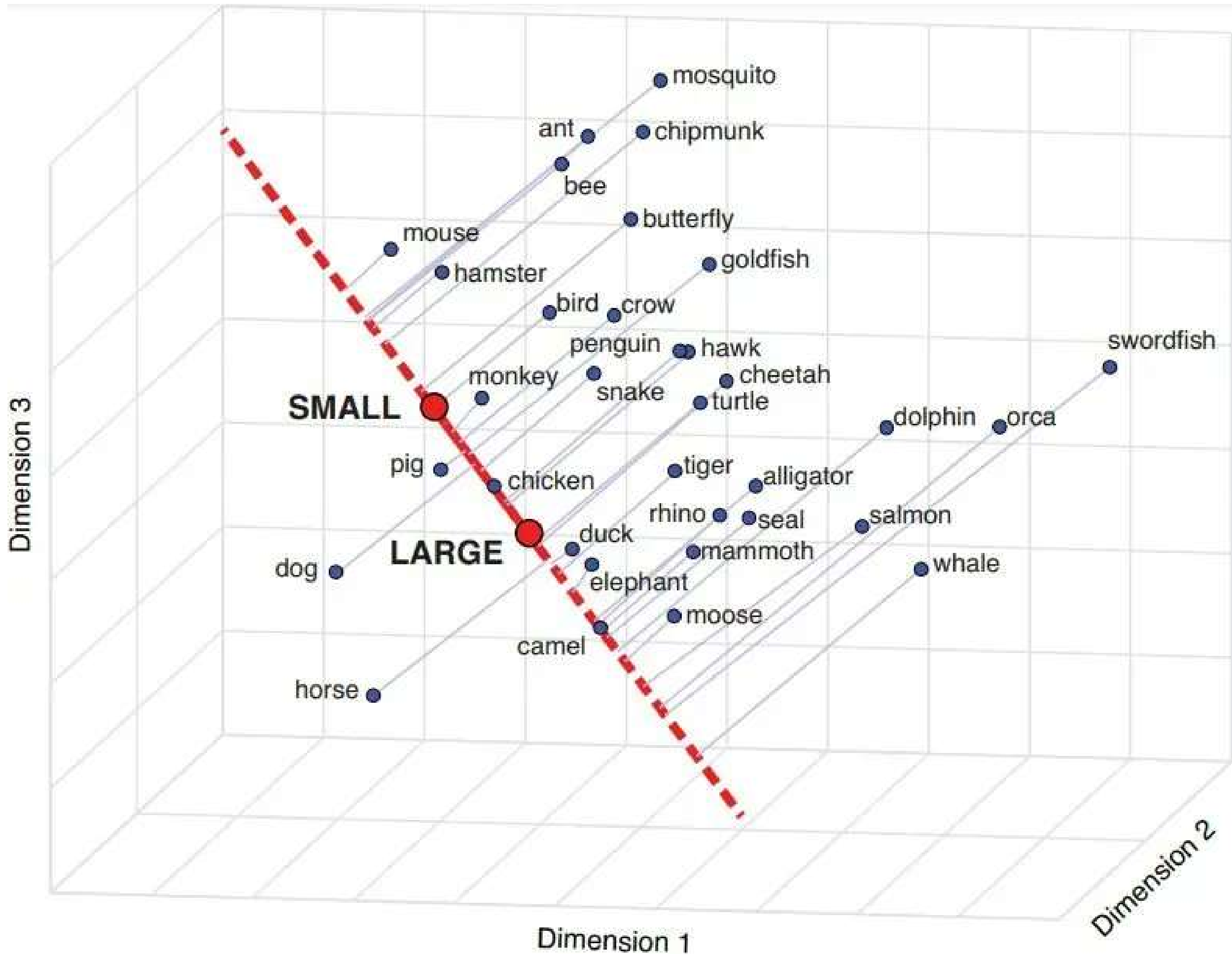
Recovery of linear dimensions by « semantic projection » from a large semantic space

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 1-13. https://doi.org/10.1038/s41562-022-01316-8

Can we recover multiple features of objects or concepts by projection on a linear dimension?

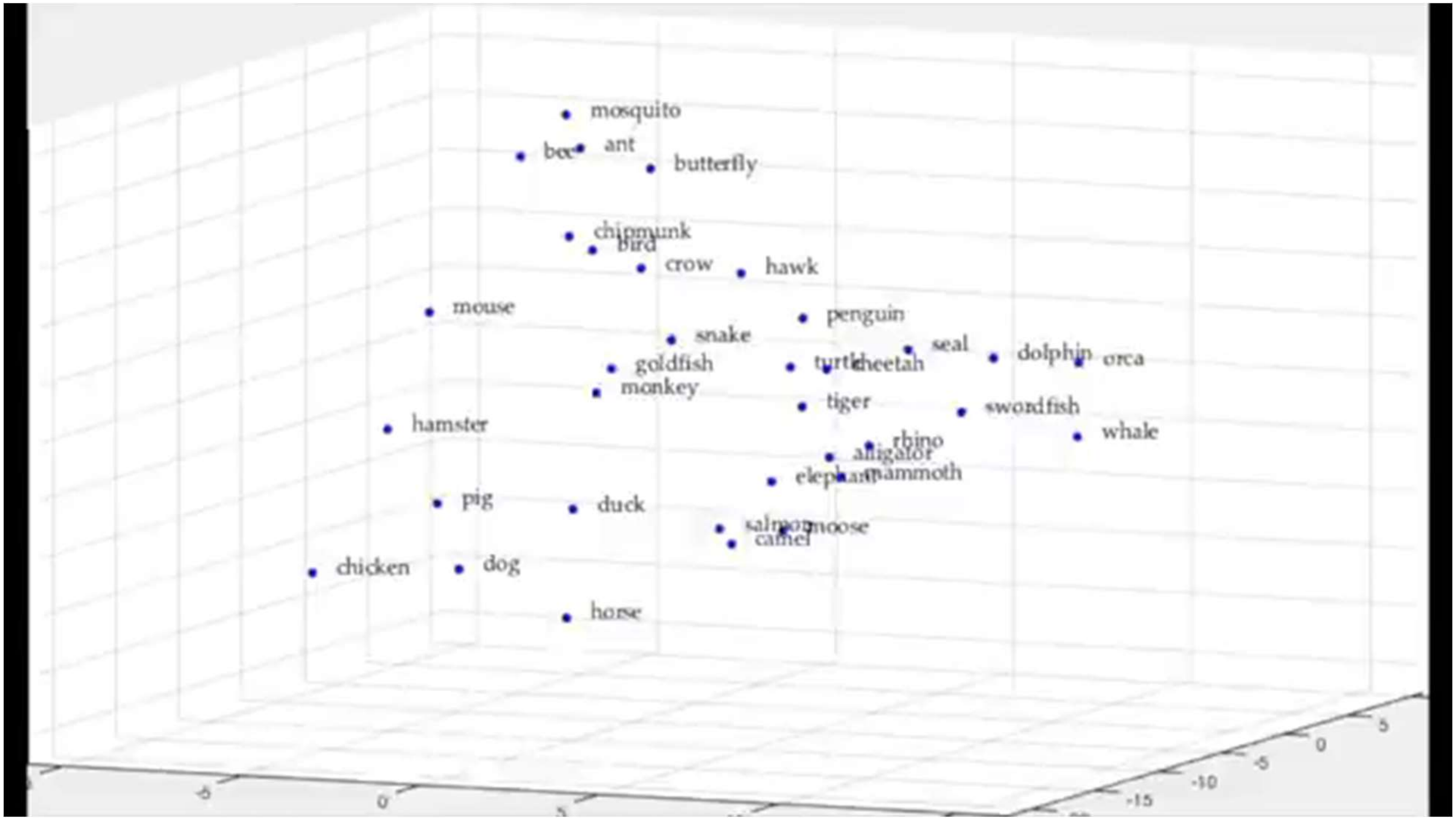
Here, no need to train a linear decoder – the authors merely « down-project the embedding space onto a line that corresponds to some psychologically-relevant semantic feature, like size" (the line connecting the vectors for "small" and "large")

Video : https://twitter.com/i/status/1514676729578725380



Recovery of linear dimensions by « semantic projection » from a large semantic space

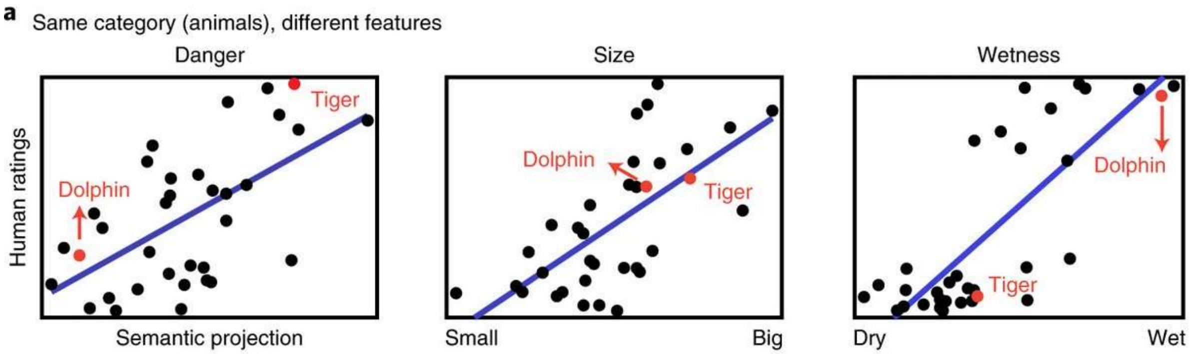
Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 1-13. https://doi.org/10.1038/s41562-022-01316-8



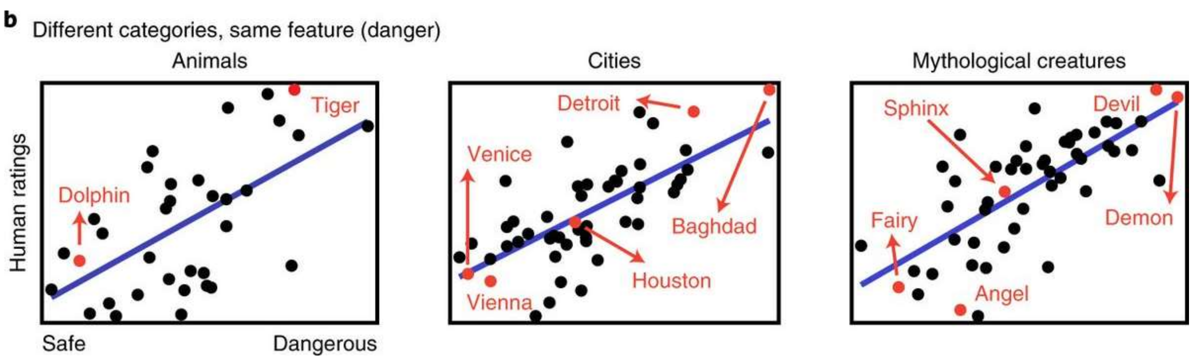
Recovery of linear dimensions by « semantic projection » from a large semantic space

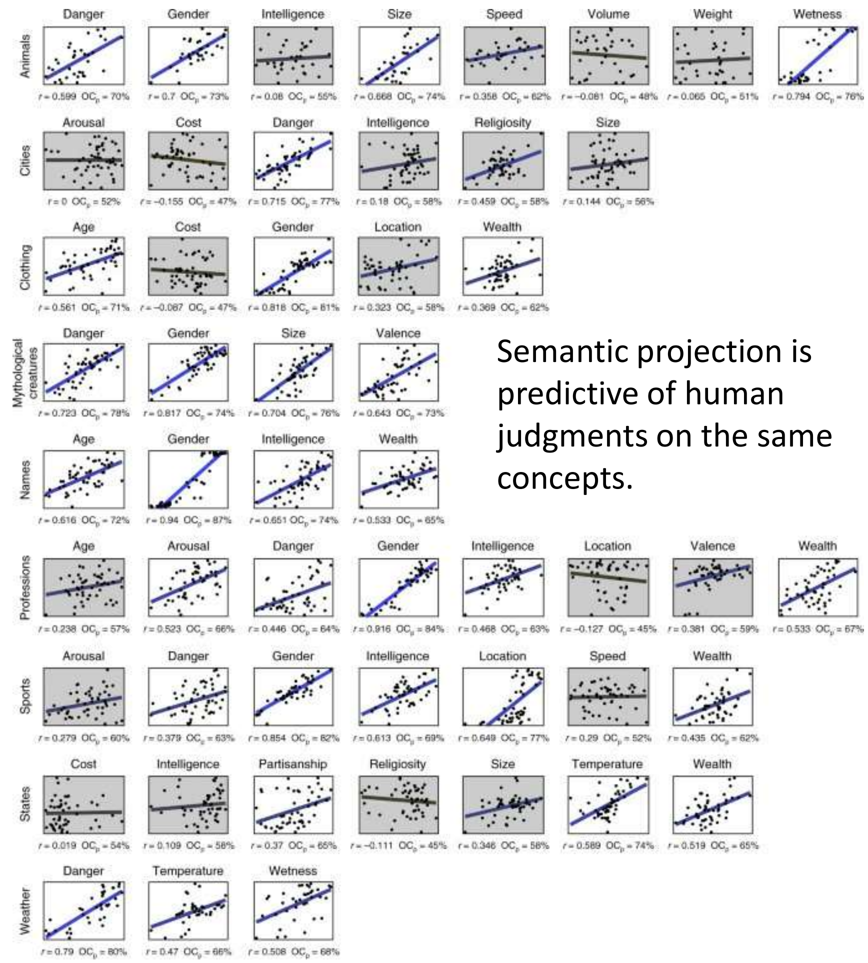
Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, 1-13. https://doi.org/10.1038/s41562-022-01316-8

Semantic projection recovers multiple dimensions for the same concept



Semantic projection recovers a given dimension across many distinct categories.





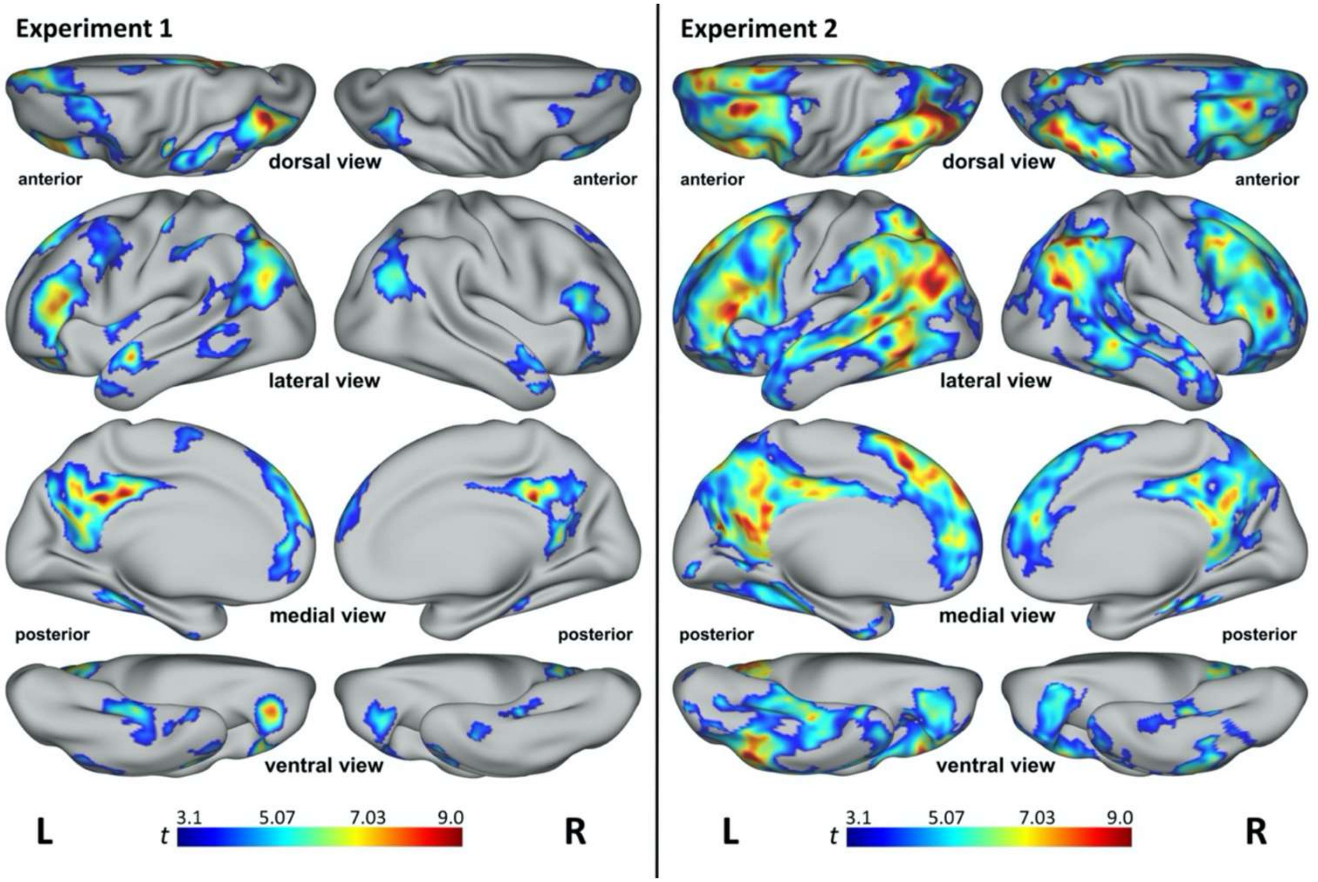
Semantic projection is predictive of human judgments on the same concepts.

Experiential and Word2vec embeddings predict the fMRI response to words

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A Distributed Network for Multimodal Experiential Representation of Concepts. Journal of Neuroscience, 42(37), 7121-7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

The authors measured the fMRI response to individual words (with a task of « rating how often they encountered the corresponding entity in their daily lives, on a scale from 1 to 3 »).
Experiment 1: 320 nouns (160 nouns of various categories [40 each of animals, foods, tools, and vehicles], 160 event nouns [social, verbal, non-verbal sound, and negative])
Experiment 2: 300 nouns (50 each of animals, body parts, food/plants, human traits, quantities, and tools).
Each word is presented 6 times, using 3T fMRI 2x2x2 mm voxels (3 sessions/subject) and surface extraction.
Each local patch of 5 mm radius is extracted, its representation similarity matrix is computed as the Pearson correlation of fMRI vectors.
Each **local brain similarity matrix** is then modelled by

- Word2Vec similarity
- 10 other potential confound matrices : number of letters, of phonemes, of syllables; mean bigram frequency, mean trigram frequency, orthographic neighborhood density, phonological neighborhood density, phonotactic probability for single phonemes, phonotactic probability for phoneme pairs, and word frequency

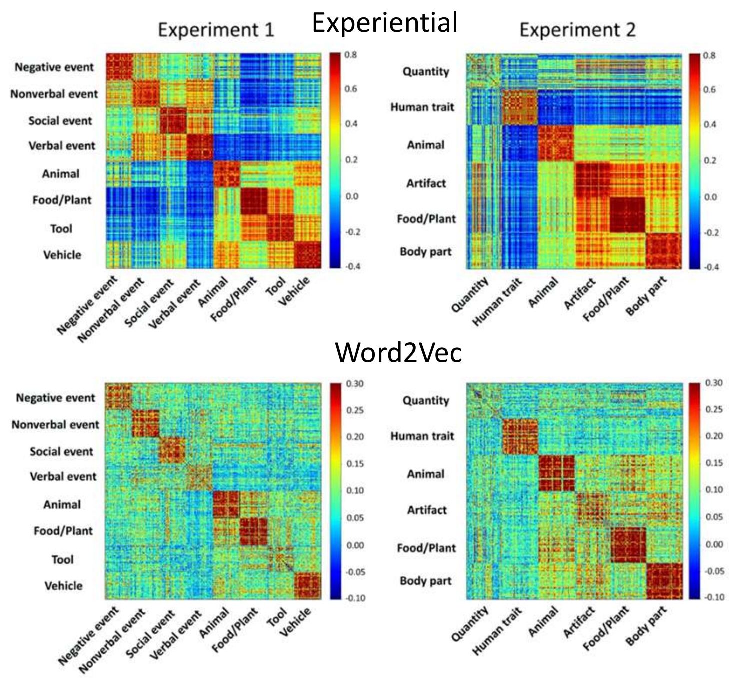


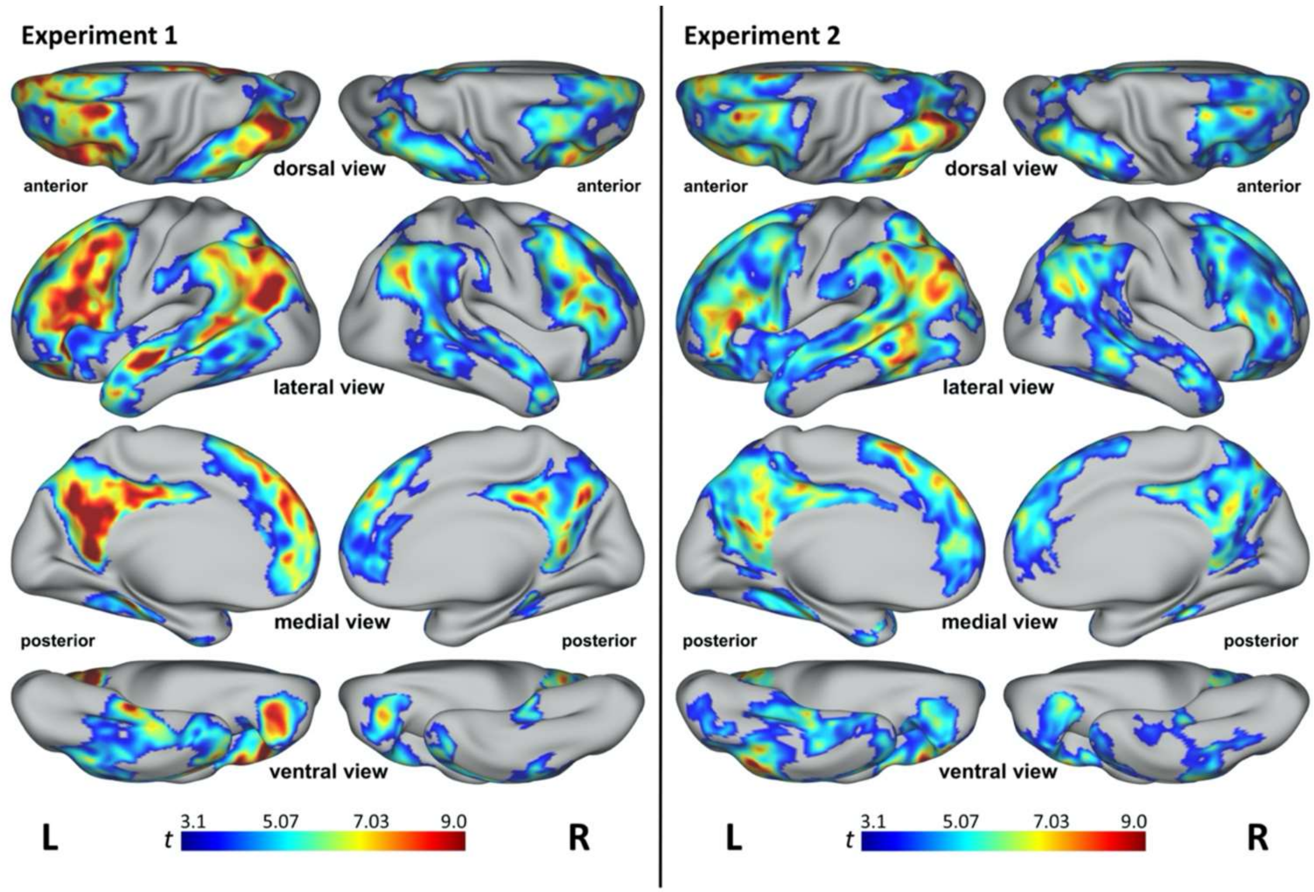
This is the cross-subject map for Word2vec similarity, after regressing the confounds.

Experiential and Word2vec embeddings predict the fMRI response to words

Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A Distributed Network for Multimodal Experiential Representation of Concepts. Journal of Neuroscience, 42(37), 7121-7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

The authors, however, test whether a better prediction can be obtained with a more subjective measure of semantic features, i.e.
Multimodal "**experiential semantic similarity"**: ratings of relevance for 65 sensory, motor, affective, and other experiential dimensions.
The matrices are indeed quite correlated:



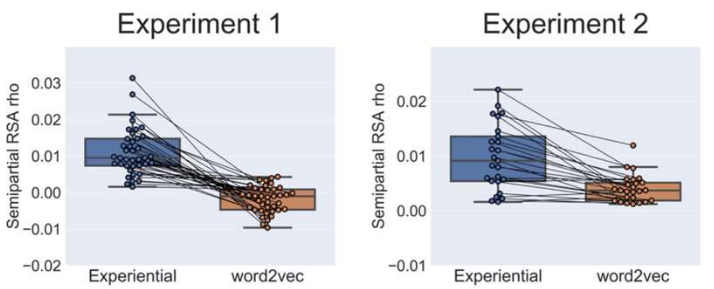


This is the cross-subject map for experiential similarity, after regressing the confounds.

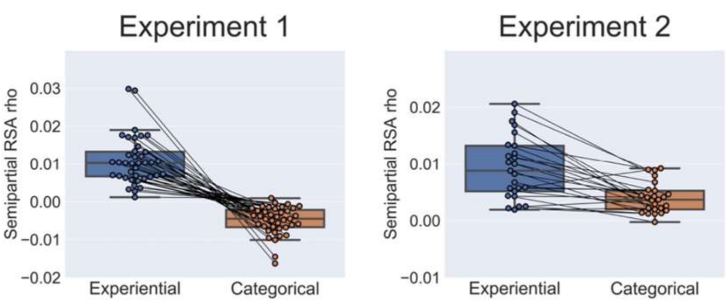
Experiential and Word2vec embeddings predict the fMRI response to words

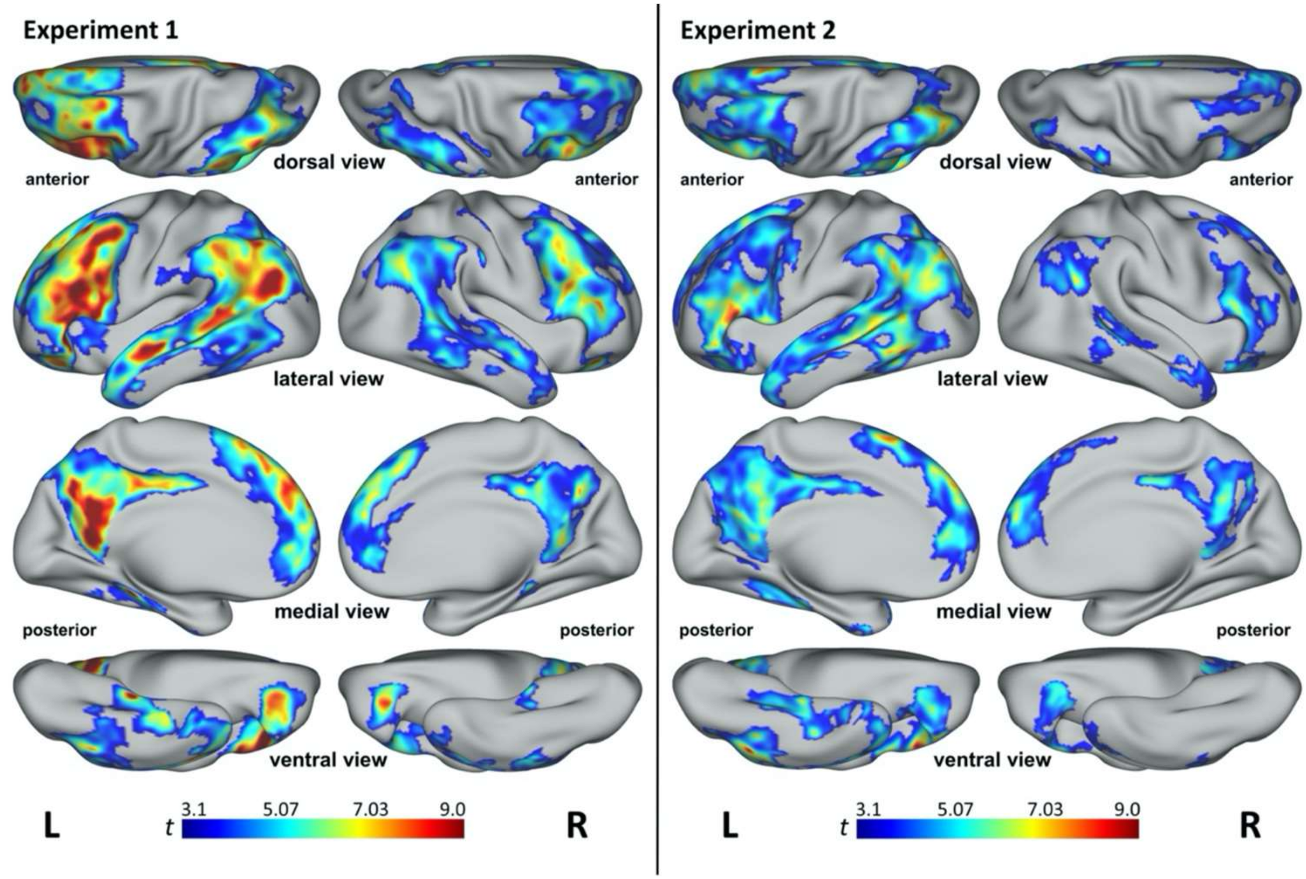
Tong, J., Binder, J. R., Humphries, C., Mazurchuk, S., Conant, L. L., & Fernandino, L. (2022). A Distributed Network for Multimodal Experiential Representation of Concepts. Journal of Neuroscience, 42(37), 7121-7130. https://doi.org/10.1523/JNEUROSCI.1243-21.2022

Although the word2vec model provides some degree of fit, there are many brain regions where the fit is better with the experiential model.



Categories alone (animals, tools, quantities, etc) do not suffice to explain the similarities)





Searchlight RSA results for the multimodal experiential model after controlling for word2vec similarity

Complete coverage of semantic space

Pereira, F., Lou, B., Pritchett, B., Ritter, S., Gershman, S. J., Kanwisher, N., … Fedorenko, E. (2018). Toward a universal decoder of linguistic meaning from brain activation. *Nature Communications*, *9*(1), 963. https://doi.org/10.1038/s41467-018-03068-4

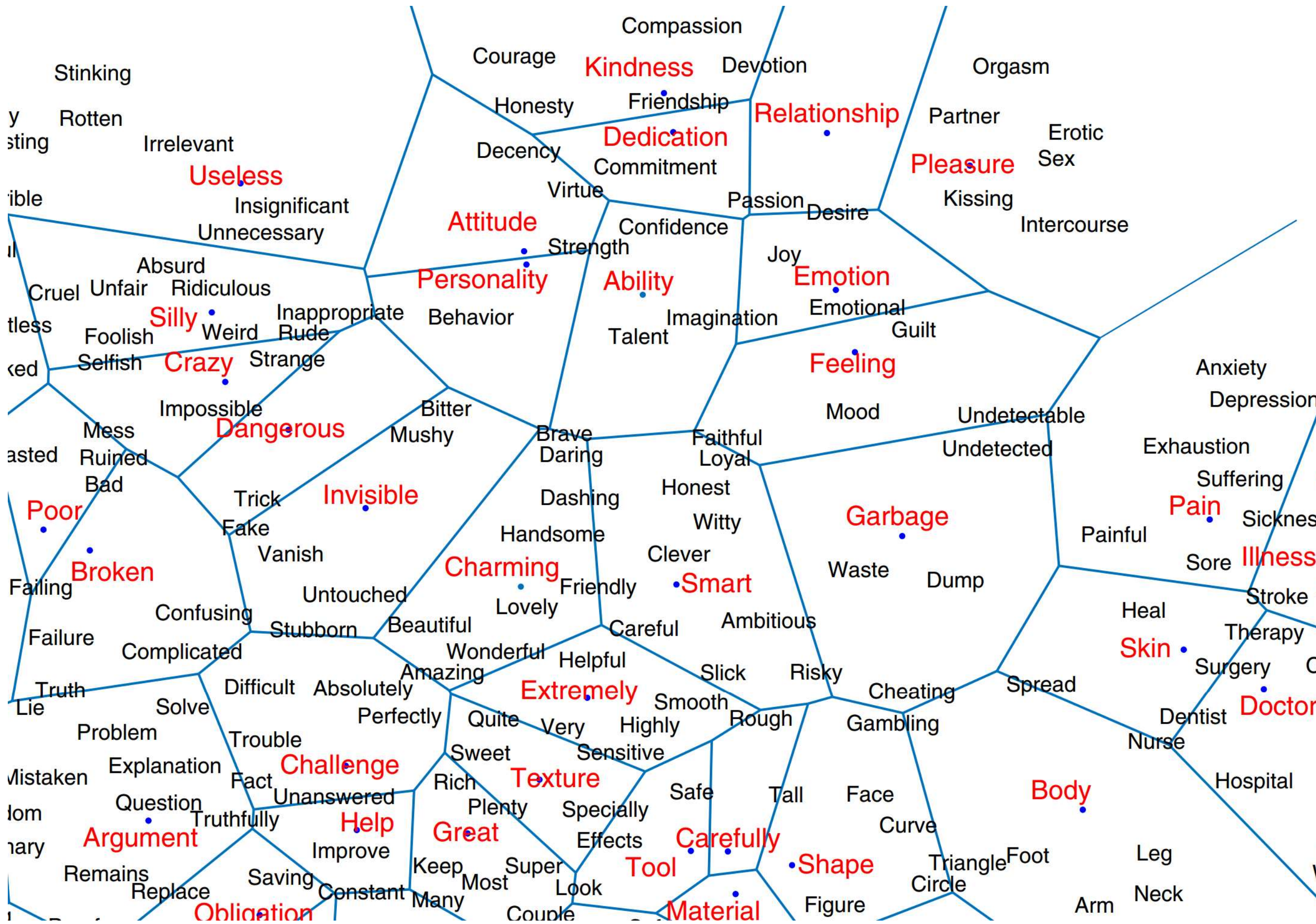
Motivation = Train a « brain decoder » with the brain activation evoked by individual words, and which could generalize to (1) novel words, based on the similarity of their semantic vectors (2) entire sentences or passages.

First challenge = achieve good coverage of semantic space (~30,000 words in a basic vocabulary) by presenting only a few hundred well-selected words.

Method =

- Get 300-d GloVe vectors for 30,000 words.
- Use Spectral clustering to group them into 200 regions.

All but 20 regions were semantically interpretable: locations, body parts, numbers
The authors then hand-selected 180 words representative of each of the 180 regions.



"We scanned 16 participants in three paradigms, all aimed at highlighting the relevant meaning of each of 180 words (128 nouns, 22 verbs, 29 adjectives and adverbs, and 1 function word).
In the first paradigm, the target word was presented in the context of a sentence that made the relevant meaning salient.
In the second, the target word was presented with a picture that depicted some aspect(s) of the relevant meaning.
In the third, the target word was presented in a "cloud", surrounded by five representative words from the cluster.
These paradigms were chosen over a simpler paradigm where the target word appears in isolation because words are highly ambiguous, especially outside the realm of concrete nouns.
These paradigms ensure that the subject is thinking about the relevant (intended) meaning of each word."
In Experiment 2, subjects read sentences that cover a subtopic of the main topic (e.g. musical instruments)

Experiment 1:

Bird

1. The bird flew around the cage.
2. The nest was just big enough for the bird.
3. The only bird she can see is the parrot.
4. The bird poked its head out of the hatch.
5. The bird holds the worm in its beak.
6. The bird preened itself for mating.

Nest
Flock
Bird
Beak
Mating
Winged

Wash

1. To make the counter sterile, wash it.
2. The dishwasher can wash all the dishes.
3. He likes to wash himself with bar soap.
4. She felt clean after she could wash herself.
5. You have to wash your laundry beforehand.
6. The maid was asked to wash the floor.

Clean
Shower
Sink
Wash
Soap
Laundry

Unaware

1. She was unaware of how oblivious he really was.
2. She was unaware of her status.
3. Unprejudiced and unaware, she went full throttle.
4. Unaware of current issues, he is a terrible candidate.
5. He was unaware of how uninterested she was.
6. He was unaware of the gravity of the situation.

Unprepared
Unprotected
Unaware
Unwilling
Inexperienced
Unconcerned

Experiment 2:

Musical instruments (clarinet)

A clarinet is a woodwind musical instrument. It is a long black tube with a flare at the bottom. The player chooses notes by pressing keys and holes. The clarinet is used both in jazz and classical music.

Musical instruments (accordion)

An accordion is a portable musical instrument with two keyboards. One keyboard is used for individual notes, the other for chords. Accordions produce sound with bellow that blow air through reeds. An accordionist plays both keyboards while opening and closing the bellows.

Musical instruments (piano)

The piano is a popular musical instrument played by means of a keyboard. Pressing a piano key causes a felt-tipped hammer to hit a vibrating steel string. The piano has an enormous note range, and pedals to change the sound quality. The piano repertoire is large, and famous pianists can give solo concerts.

Experiment 3:

Skiing (passage 1)

I hesitantly skied down the steep trail that my buddies convinced me to try. I made a bad turn, and I found myself tumbling down. I finally came to a stop at a flat part of the slope. My skis were nowhere to be found, and my poles were lodged in a snow drift up the hill.

Skiing (passage 2)

A major strength of professional skiers is how they use ski poles. Proper use of ski poles improves their balance and adds flair to their skiing. It minimizes the need for upper body movements to regain lost balance while skiing.

Skiing (passage 3)

New ski designs and stiffer boots let skiers turn more quickly. But faster and tighter turns increase the twisting force on the legs. This has led to more injuries, particularly to ligaments in the skier's knee.

Gambling (passage 1)

When I decided to start playing cards, things went from bad to worse. Gambling was something I had to do, and I had already spent close to $10,000 doing it. My friends were sick of watching me gamble my savings away. The hardest part was the horror of leaving a casino after losing money I did not have.

Gambling (passage 2)

Good data on the social and economic effects of legalized gambling are hard to come by. Some studies indicate that having a casino nearby makes gambling problems more likely. Gambling may also be associated with personal bankruptcies and marriage problems.

Gambling (passage 3)

Over the past generation, there has been a dramatic expansion of legalized gambling. Most states have instituted lotteries, and many have casinos as well. Gambling has become a very big but controversial business.

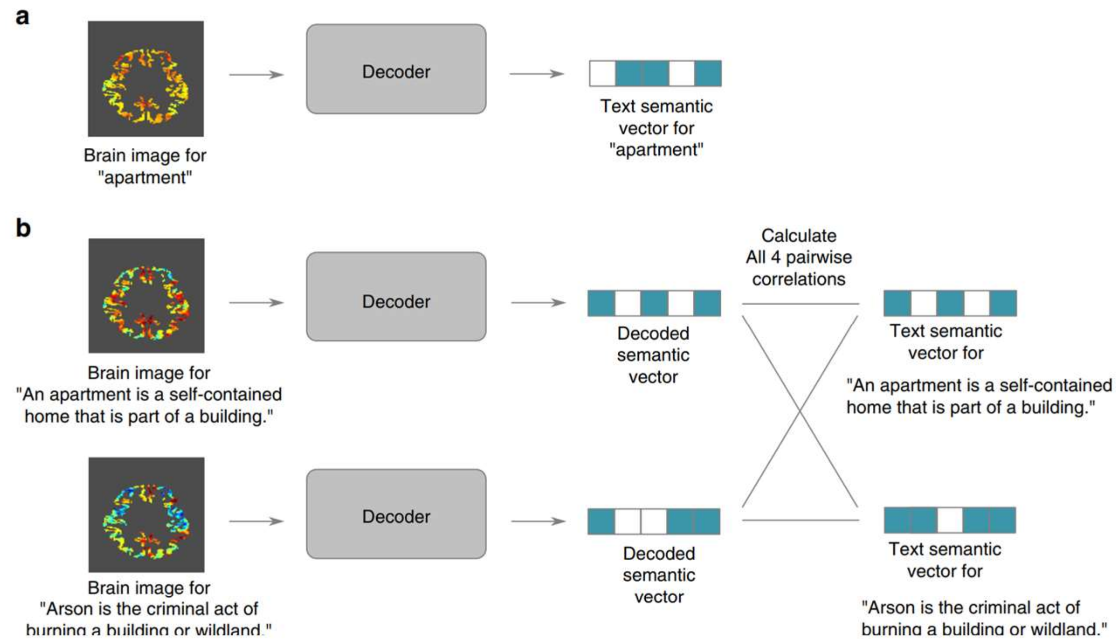
The authors use whole-brain activation pattern to
(1) Train a decoder, using 170 concepts, to recover the single-word semantic vector from the activation pattern
(2) Test generalization to the remaining 10 concepts and to new conditions.

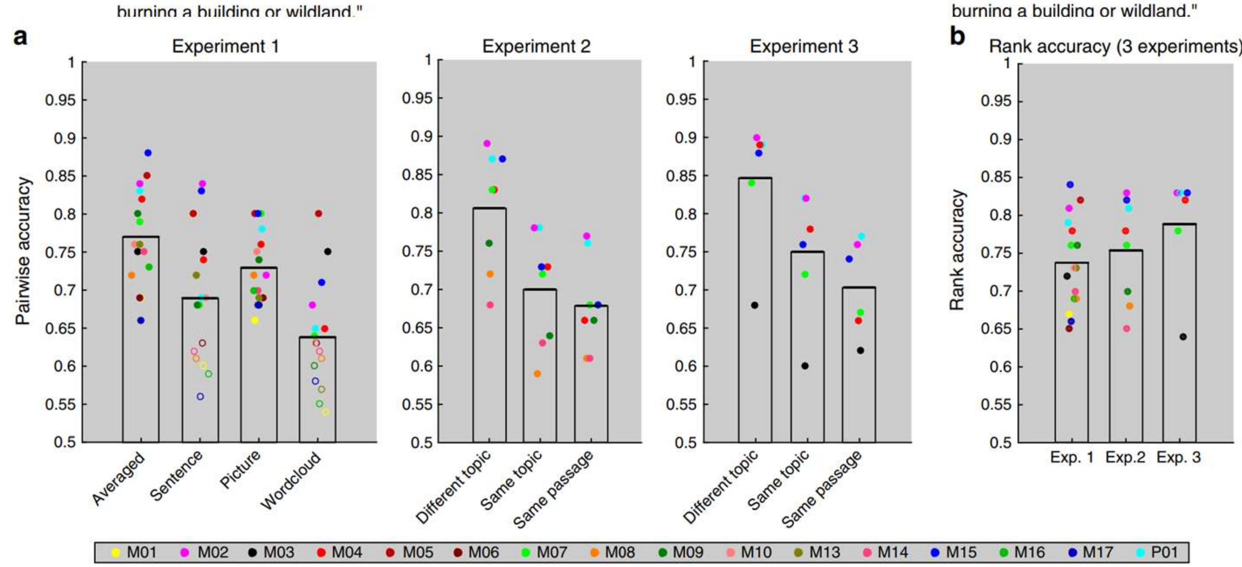
Chance level is 50 % -- the authors do better than chance, but this is more of a proof of concept.

In particular, sentences “hover” around the same semantic vector space as individual words.

For much better decoding results, including a reconstruction of sentences from MEG see e.g.
Défossez, A., Caucheteux, C., Rapin, J., Kabeli, O., & King, J.-R. (2022). Decoding speech from non-invasive brain recordings (arXiv:2208.12266). arXiv.
https://doi.org/10.48550/arXiv.2208.12266

https://twitter.com/JeanRemiKing/status/1564964019965927424

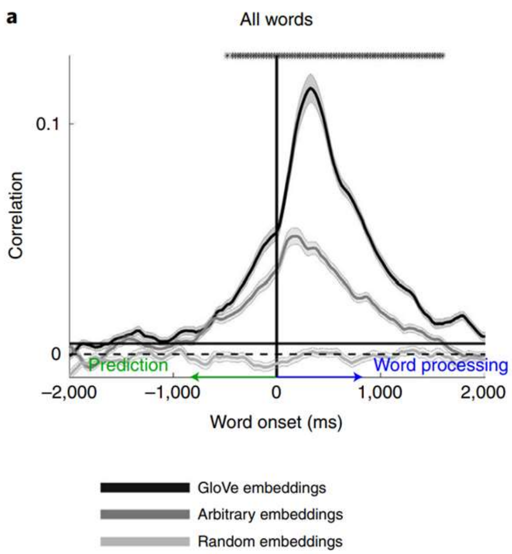


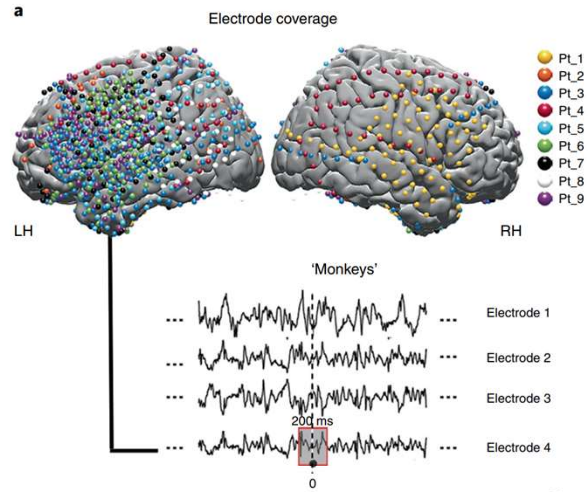


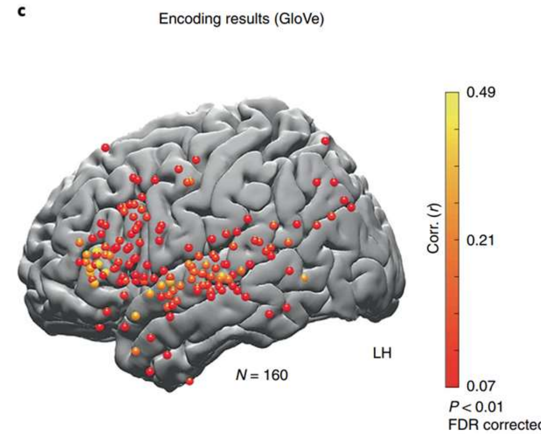
Intracranial information about the neural codes for language

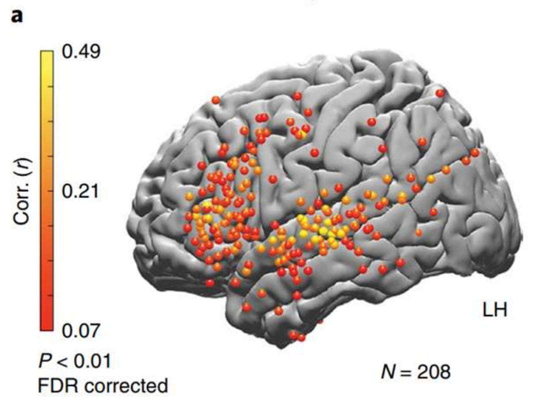
Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., … Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. Nature Neuroscience, 25(3), Art. 3. https://doi.org/10.1038/s41593-022-01026-4

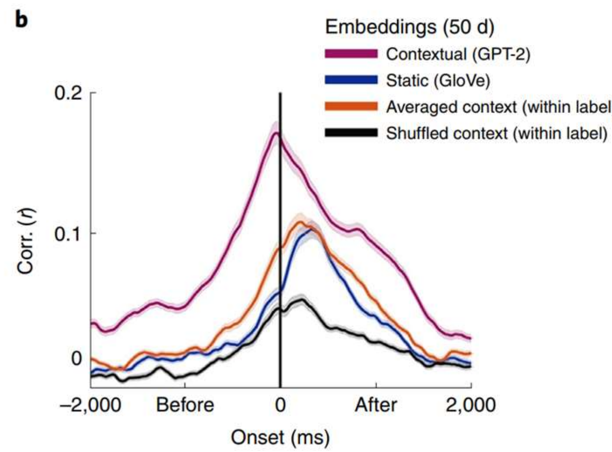
Glove encodings can also be used to predict intracranial ECOG signals from 9 patients with ECOG.
Interestingly, the signals arise even before word onset (and many additional analyses show that the brain anticipates the upcoming word)
Indeed, "autoregressive" models such as GPT2, which are trained to predict the upcoming word, provide a better match to brain signals.











Conclusions

Words can be represented as vectors in a high-dimensional semantic space, by several techniques (behavioral, introspective judgments, or distributional statistics of large text corpuses).

It is often surprising how **deep semantic relations** between words can be discovered automatically in those vectors through distributional statistics (gender, capitals, etc).

Experimental validations of the proposed vectors include
- **Behavioral** measures of conceptual similarity
- Capacity of those vectors to model **brain activity**

It should be noted that, at the brain level, there are few if any demonstrations of fine-grained vectors *within* an area – most of the work indicates that **different semantic dimensions are attributed to different regions**, and therefore to different neural populations.

Future work should try to clarify under which conditions the brain learns to assign dimensions to distinct neurons, and under which conditions the neural code is fully distributed.

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)