What makes different people alike: A solution to the problem of across-subject fMRI decoding

Rajeev Raizada

Neukom Institute, Dartmouth College

http://www.dartmouth.edu/~raj

Joint work with Andrew Connolly

What do different people share in common?

Striving for universality

- We don't just want to understand one particular person's brain
- We want findings that are true of all human brains
- We want theories which succeed at the population level

Neural decoding: what is it, and why bother?

Just seeing that some brain area "lights up" doesn't tell us anything about what that lit-up area is actually doing

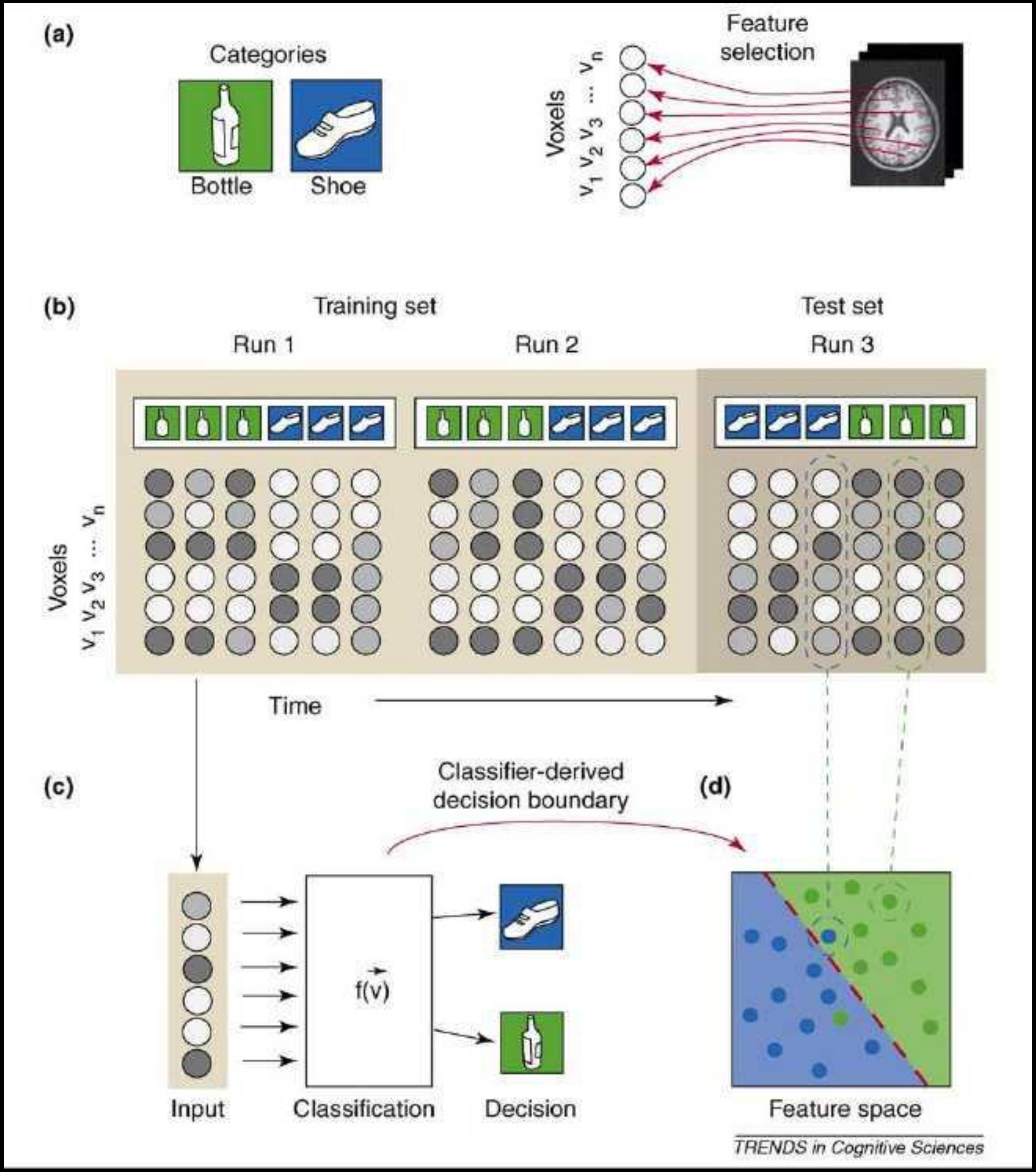
We need to be able to interpret that activation: what mental process is it implementing?

fMRI decoding: given the activation pattern, figure out what task-condition gave rise to it

The problem of across-subject fMRI decoding

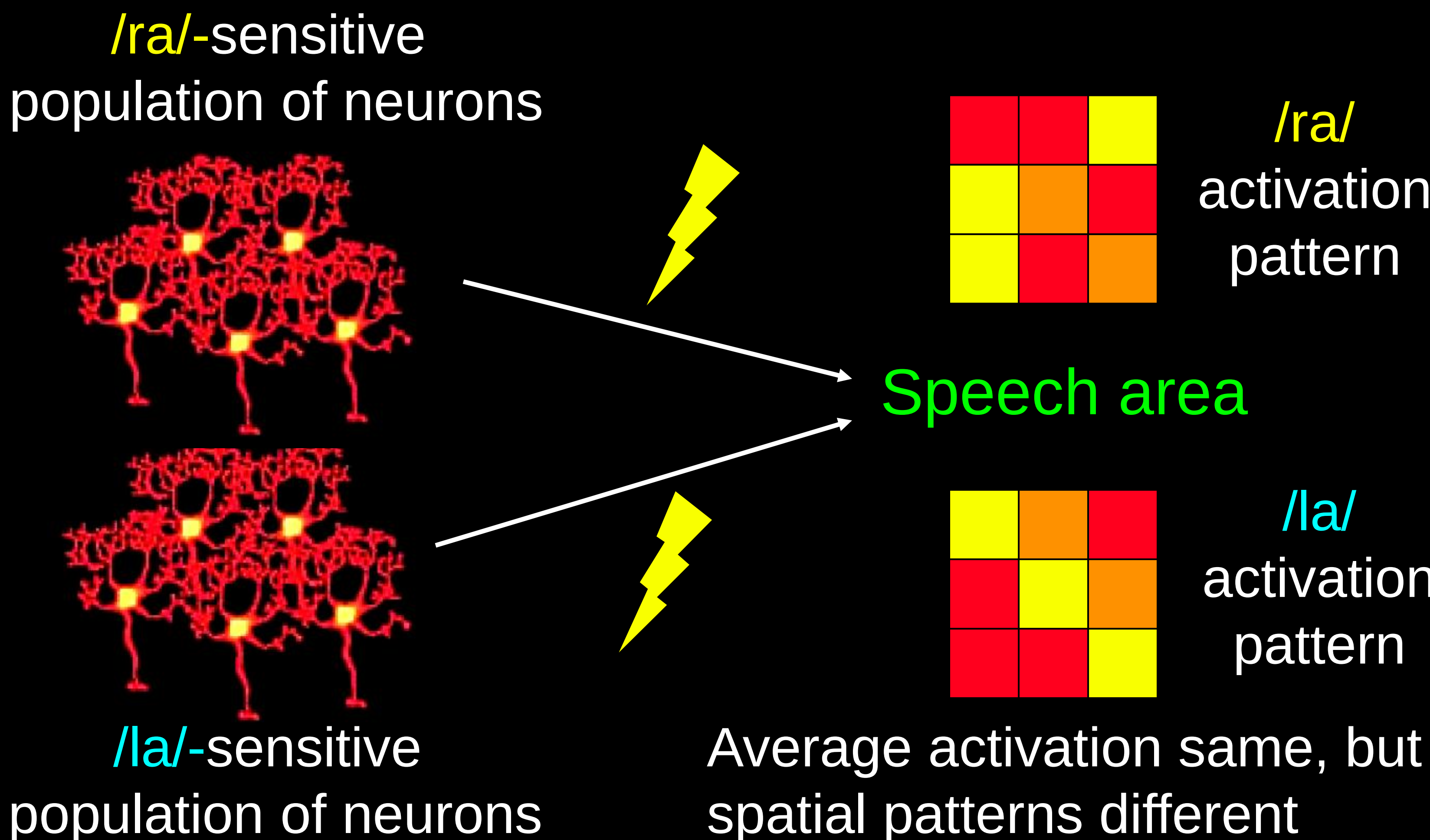
Neural decoding seems to work quite well within-subjects, but not very well across-subjects

What is neural decoding? (within-subject)



From: **Norman, Polyn, Detre & Haxby** (2006), **Trends in CogSci, 10(9), 424–30**

Multivoxel "neural fingerprints" contain stimulus-information



Raizada et al., Cerebral Cortex, 2010

Different people's brains:
alike at coarse-scale, different at fine-scale

I can align my hand to your hand, and the fingers will match up

But the fingerprints won't match up

Just like literal fingerprints, neural fingerprints seem to be subject-unique

Shinkareva, Mitchell and colleagues (PLoS ONE, 2008):

- Attempted both within- and across-subject decoding
- Found that “a critical diagnostic portion of the neural representation of the categories and exemplars is still idiosyncratic to individual participants”

A seemingly obvious idea, which actually turns out to be wrong (in my view)

Obvious idea: To do neural decoding across subjects, you take the subjects' neural activation, and enter it into a decoder

Within-subject neural decoding:

- Pick a set of voxels in the single subject's brain
- Get the activation patterns across those voxels, leaving one run out
- Feed those activation patterns into a classifier
- Predict activation pattern for that same set of voxels for the left-out run

Across-subject neural decoding:

- Pick the same set of voxels, in all subjs' brains
- Get the activation patterns across those voxels, leaving one subject out
- Feed those activation patterns into a classifier
- Predict activation pattern for that same set of voxels for the left-out subject

Why across-subject decoding in neural activation space doesn't work very well

My "neural fingerprints" are not like your neural fingerprints

If I know the fingerprints of nine people, I still can't predict the fingerprints of a 10th person, except in very approximate terms

- "It will be swirly, and it will be on their finger"

We need to abstract away from subject-specific neural patterns

- But what should we abstract-away to ?

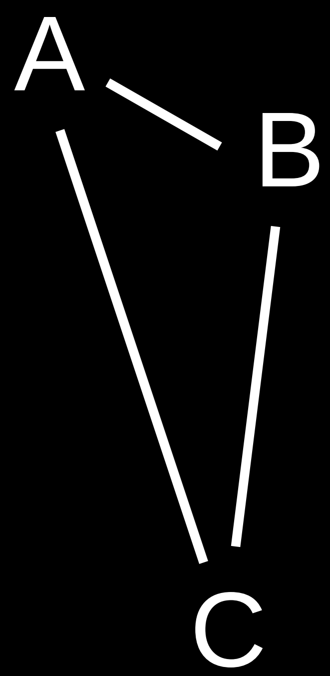
Similarity-space

The set of pairwise similarities between items, as defined by some similarity-measure (or dissimilarity-measure)

Distances between

cities A, B and C

	A	**B**	**C**
A	0	1	5
B	1	0	4
C	5	4	0



Similarity-space: a long history in cognitive psychology and computer science

Roger Shepard (1962), John Kruskal (1964)

- Multidimensional scaling (MDS)
- Takes a set of similarities, and represents them as the best-fitting lower-dimensional projection

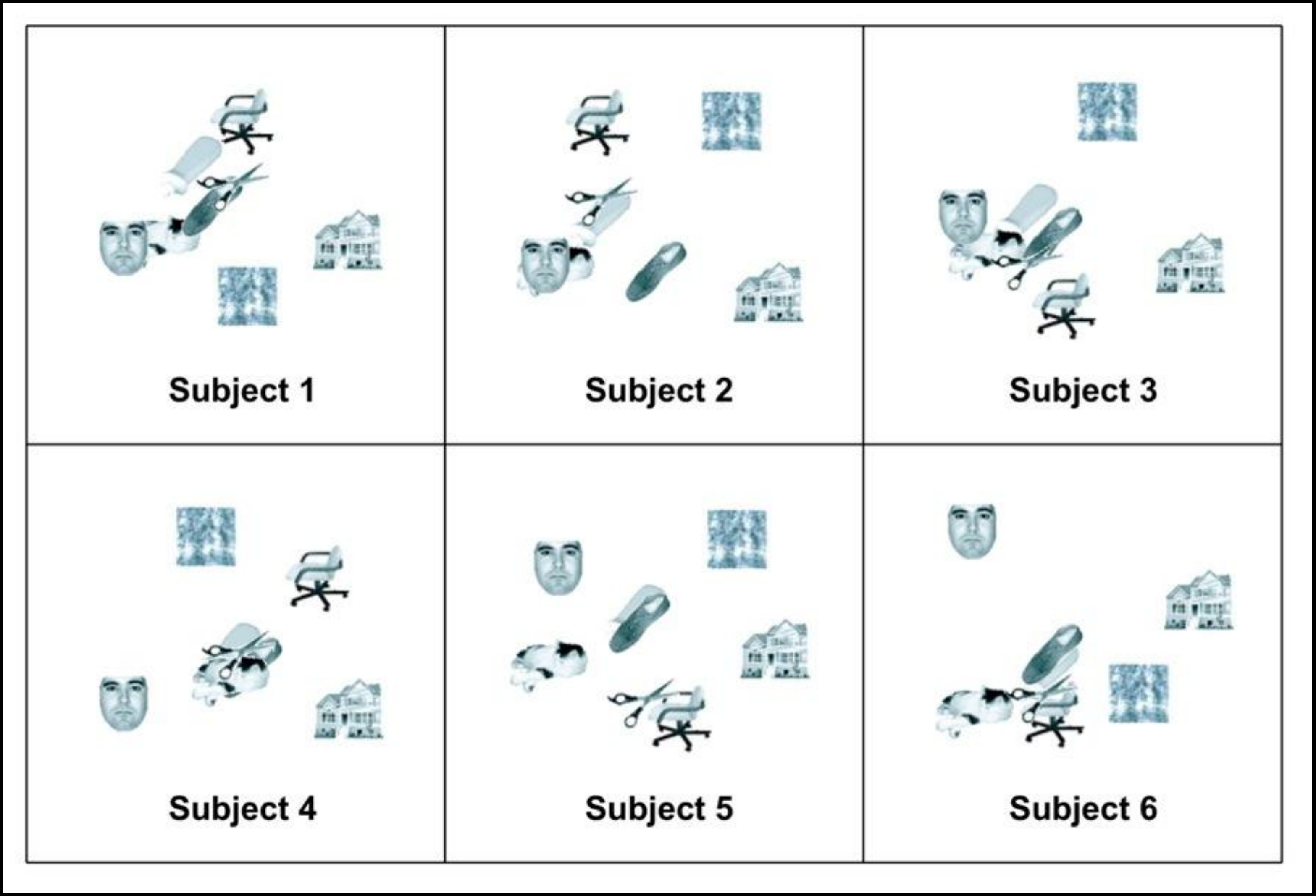
Laakso & Cottrell (1998, 2000)

- Similarity-space of hidden units in neural networks
- Building upon a proposal in philosophy of mind by Paul Churchland

Shimon Edelman (1998)

- “Representation is representation of similarities”
- Computer vision, visual psychophysics

Neural similarity-space: shows representational structure, but does not seem to enable decoding



Neural similarity-space: shows representational structure, but does not seem to enable decoding

N. Kriegeskorte / NeuroImage 56 (2011) 411–421

	stimulus decoding with response-pattern classifier	cross-decoding with response-pattern classifier	voxel receptive-field (RF) modeling	stimulus reconstruction	representational similarity analysis
example studies	Haxby et al., 2001; Kamitani & Tong, 2005	Polyn et al., 2005; Stokes et al., 2009	Kay et al., 2008; Mitchell et al., 2008	Miyawaki et al., 2008; Naselaris et al., 2009	Kriegeskorte et al., 2008a, 2008b
weaknesses	predefined category grouping may be artificial and may miss major variance-explaining factors		difficult to apply to higher regions, where computational models are lacking or have prohibitive parameter complexity for RF-model fitting	engineering, not neuroscience focus: unclear how to test theories or draw specific neuroscientific conclusions	neuroscience, not engineering focus: no prediction of activity patterns or decoding
	no generalization to novel stimuli	limited generalization to novel stimuli			

Why you can’t use similarity-space to perform neural decoding*

Neural decoding:

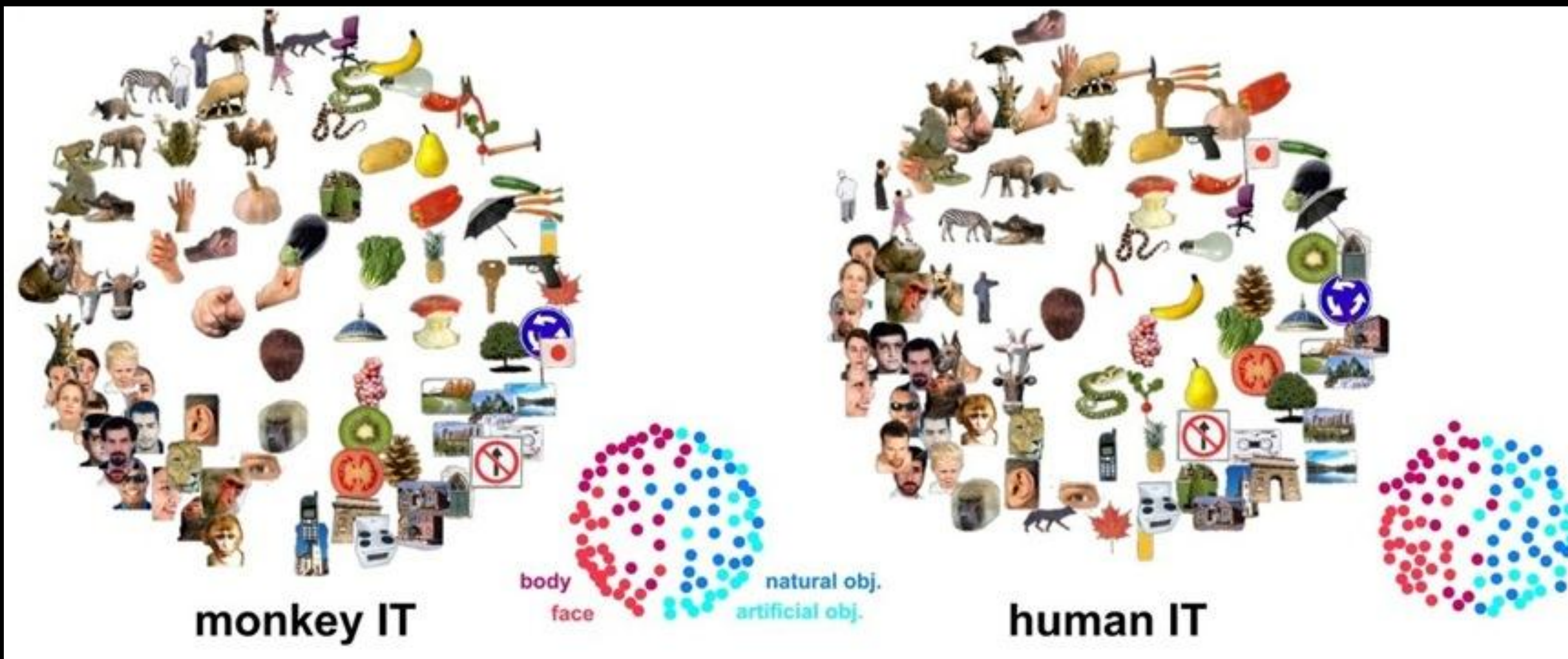
- Take a bunch of neural activation
- Enter it into a decoder (a pattern-classifier)

Similarity-space:

- There is no neural activation, only similarities
- There is no classifier algorithm, only visualisation / dimensionality-reduction algorithms such as MDS

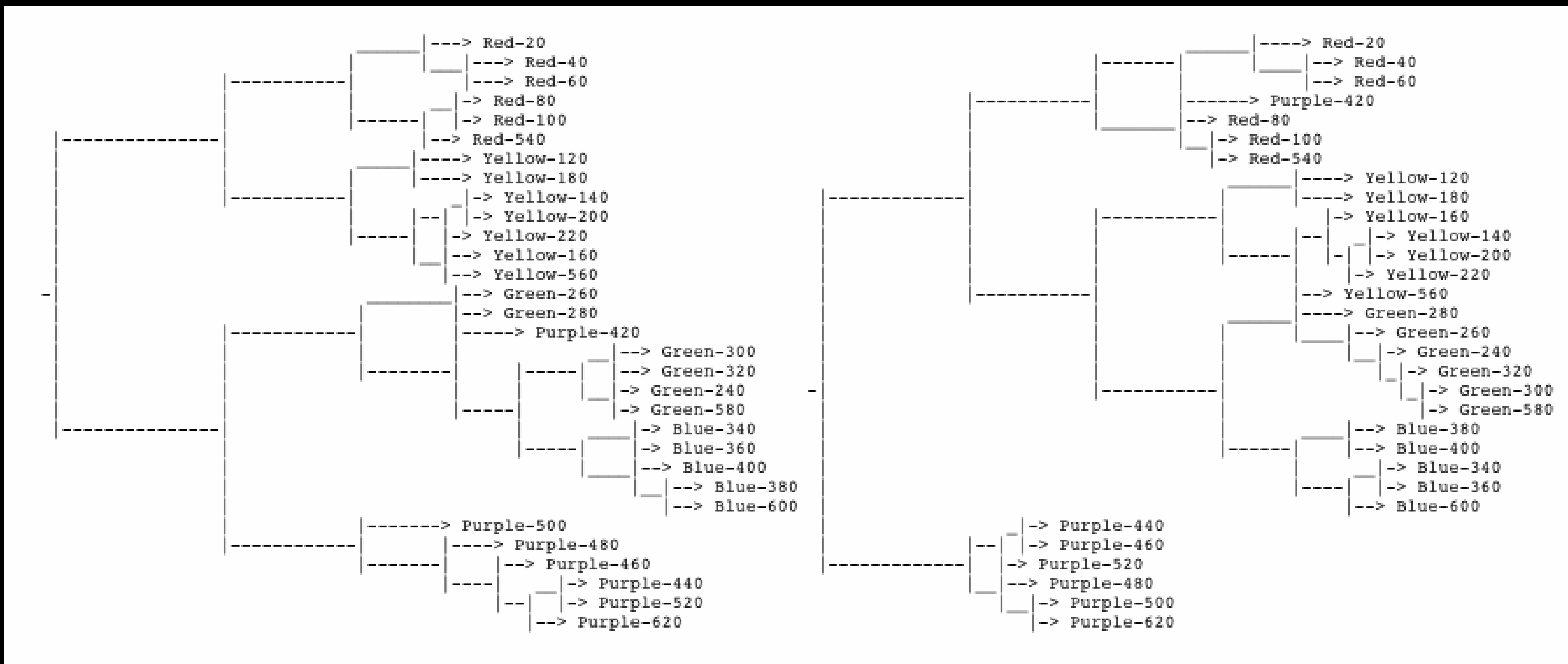
*or at least, why it looks that way

However, comparing similarity-spaces can be very informative, even without decoding



Kriegeskorte, Kiani et al., Neuron, 2008

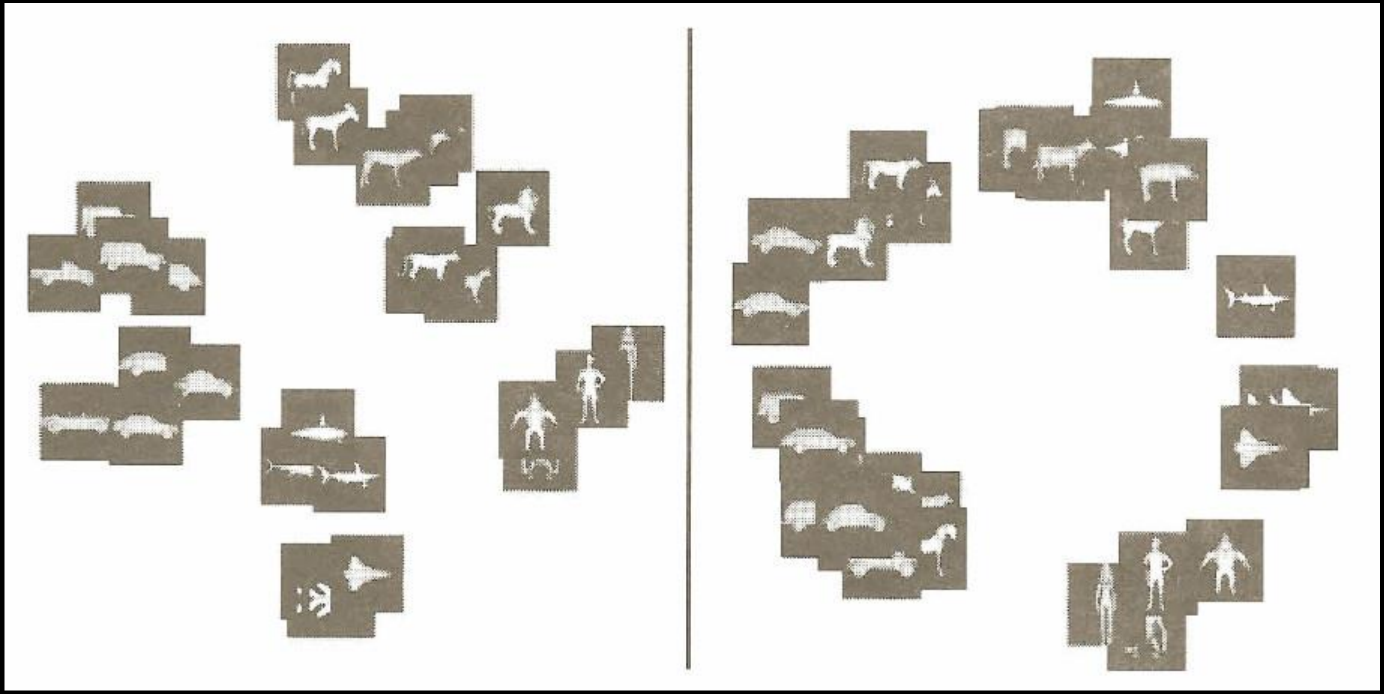
However, comparing similarity-spaces can be very informative, even without decoding



Laakso & Cottrell (1998, 2000):

Comparing similarity-spaces of hidden-unit activations, in neural networks with different architectures trained on the same data

However, comparing similarity-spaces can be very informative, even without decoding



Behavioural similarity Neural similarity

Edelman, S., Grill-Spector, K, Kushnir, T & Malach, R. (1998) Toward direct visualization of the internal shape representation space by fMRI. Psychobiology, 26, 309-321.

So, why bother about decoding?

Goal:

- Demonstrate conceptual similarity across neural diversity
- In other words, show that two different people's neural representational schemes are the same

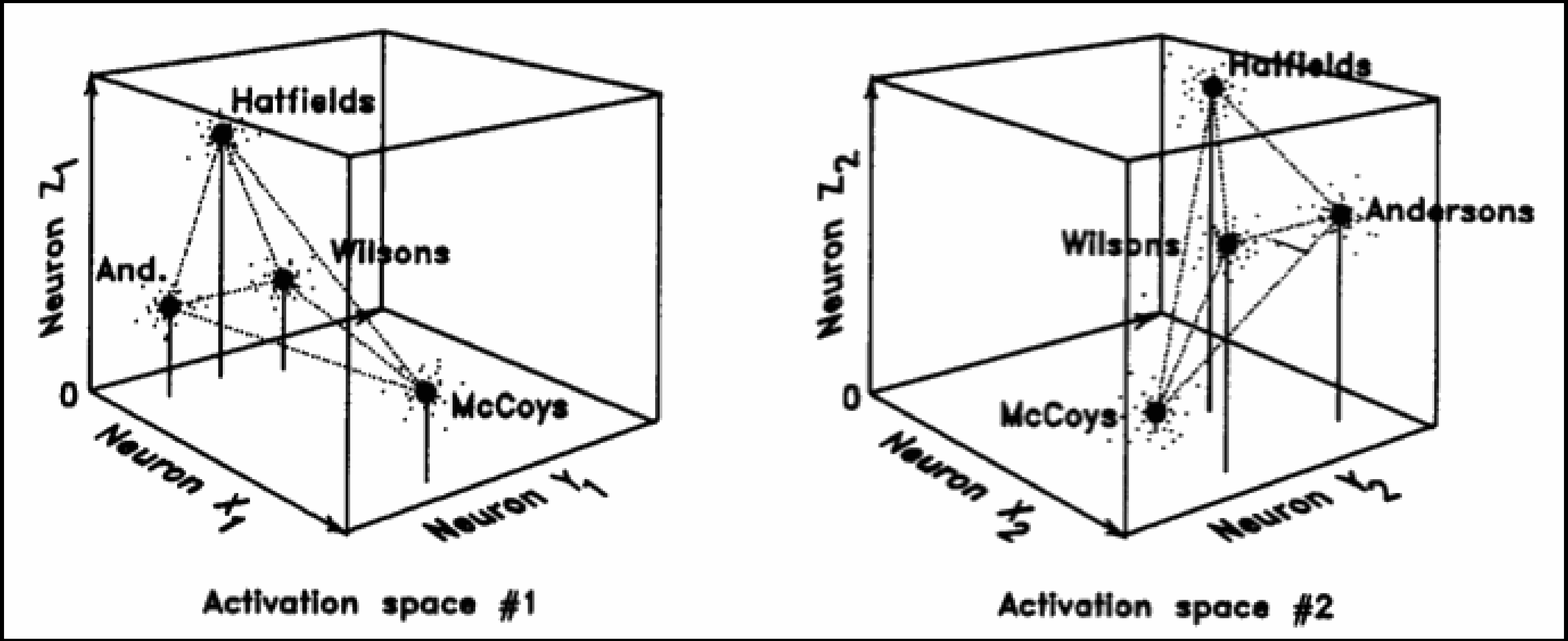
Donald Davidson (1974)

- "On the very idea of a conceptual scheme"
- In order to show that two conceptual schemes are the same, you need to be able to translate between them
- Translating between different people's neural representations = Across-subject neural decoding

The problem of conceptual similarity across neural diversity

- Suppose you and I are looking at the same object, e.g. an apple.
- It elicits some neural patterns inside my head
- It elicits different neural patterns inside your head
- But at some level, we both have the same thought: “apple”
- What, then, is the neural processing that my brain and your brain share in common?
- Problem described by Paul Churchland (1986, 1998)

Churchland's proposed solution: similarity-space is what is shared



P.M. Churchland (1998) "Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered", J. Philosophy, 95, 5-32.

Drawing upon neural network simulations by Laakso & Cottrell (1998)

Unanswered questions

Does the brain actually do anything like this?

Can we show that the representations in one person's similarity-space are the same as the representations in another person's?

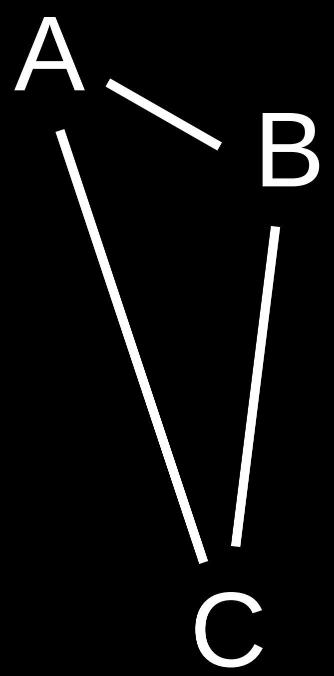
In other words, can we use similarity-space to perform across-subject neural decoding?

Similarity-space (again)

The set of pairwise similarities between items, as defined by some similarity-measure (or dissimilarity-measure)

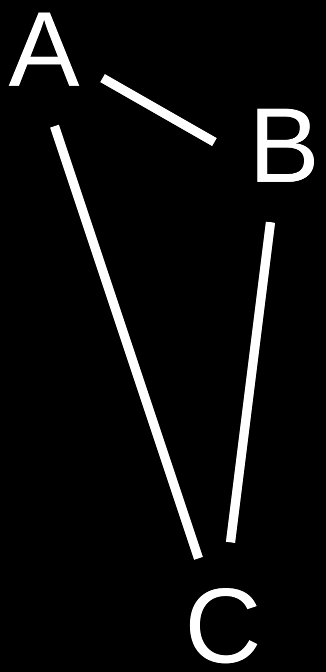
Distances between cities A, B and C

	A	**B**	**C**
A	0	1	5
B	1	0	4
C	5	4	0



How to decode in similarity-space: a simple solution

Distances between cities A, B and C



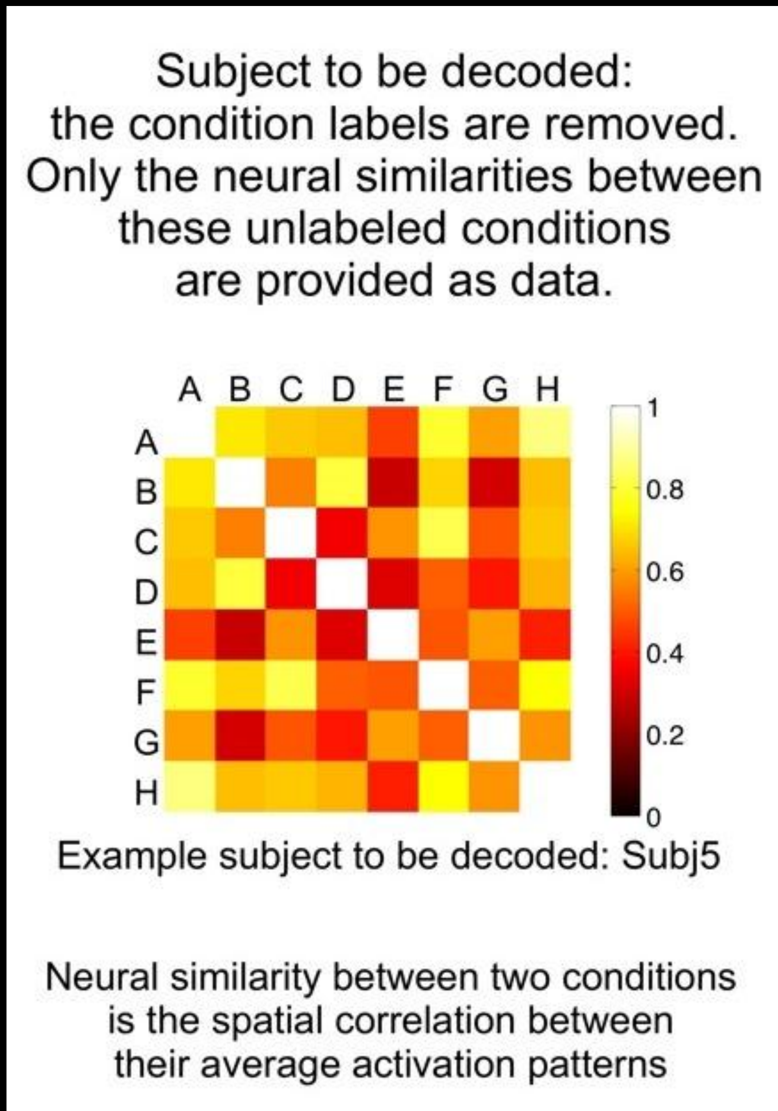
Three labels for the cities. Which ones correspond to A, B and C?

San Diego

Boston

NYC

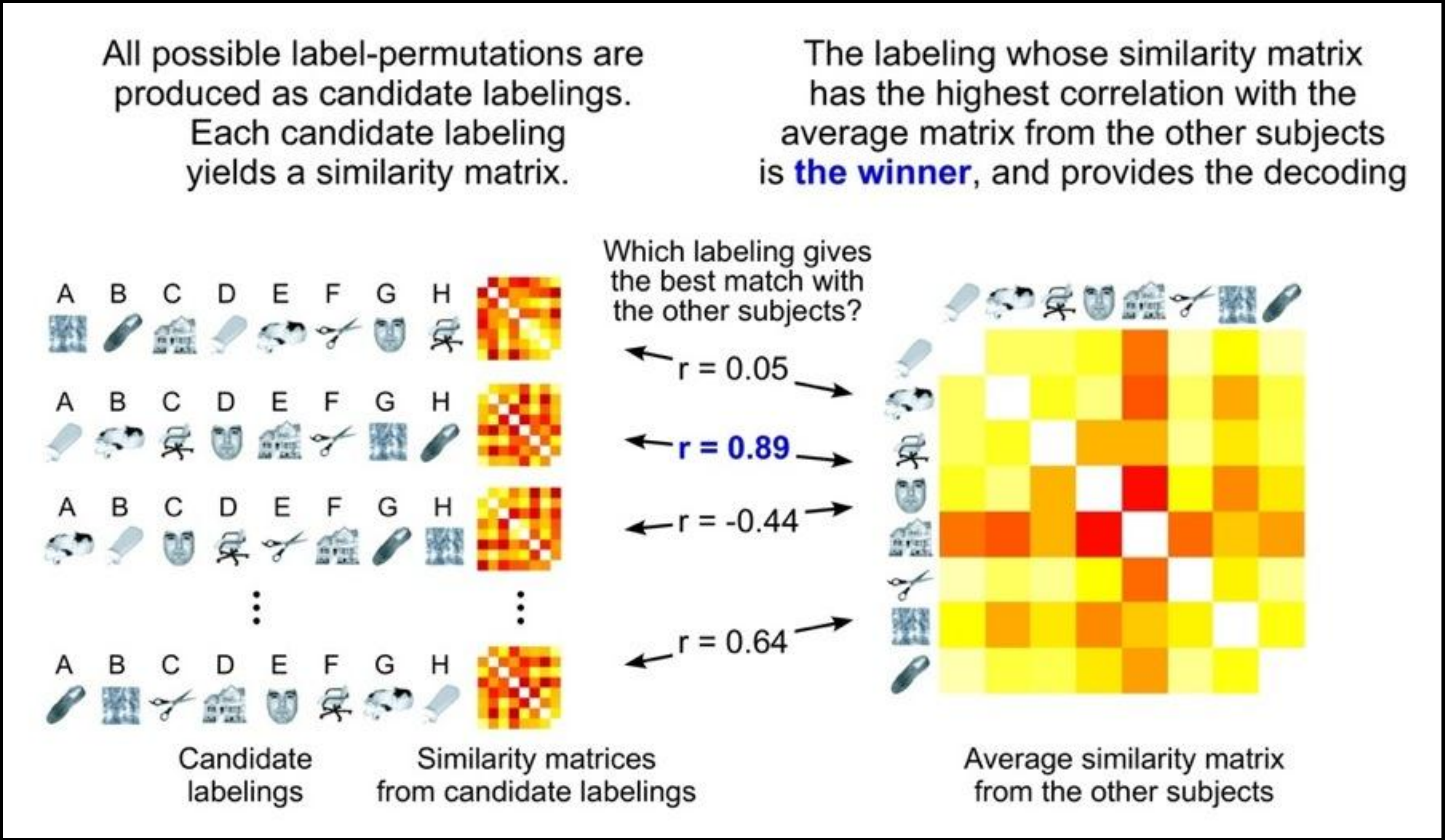
Across-subject decoding via neural similarity-space



Goal:

Figure out which labels correspond to which stimulus conditions, using only condition-labels from the other subjects

A simple solution: permute the labels, and see which permutation matches best



How well does this work with actual neural data?

Dataset:

- Haxby et al, Science, 2001. “Distributed and overlapping representations of faces and objects in ventral temporal cortex.”
- http://dev.pymvpa.org/datadb/haxby2001.html

Six subjects, eight stimulus categories:

- bottles, cats, chairs, faces, houses, scissors, scrambled-pictures, shoes

Voxels from ventral-temporal (VT) cortex masks

- Lingual, parahippocampal, fusiform, and inferior temporal gyri

How well does this work with actual neural data?

Result: 91.7% correct

- 44 out of 48 decodings correct
- 48 = 6x8: 6 subjects, 8 categories per subj

Five subjects: all eight categories correct

One subject: 4 out of 8 correct

- Confusions: bottle-scissors, shoe-chair

Shared hierarchy of representations across subjects

- Not just animate vs. inanimate
- Multiple animate and inanimate subcategories

How likely is it to get N labelings correct, just by chance?

For 8 categories, there are 8-factorial possible labelings

- 8! = 40320

Only one of those 40320 labelings gets all 8-out-of-8 labels correct.

- 5 of the 6 subjects achieved that perfect labeling

Permutation distribution:

p < 0.05 critical number of correct labeliings: n = 3

What about neural diversity?

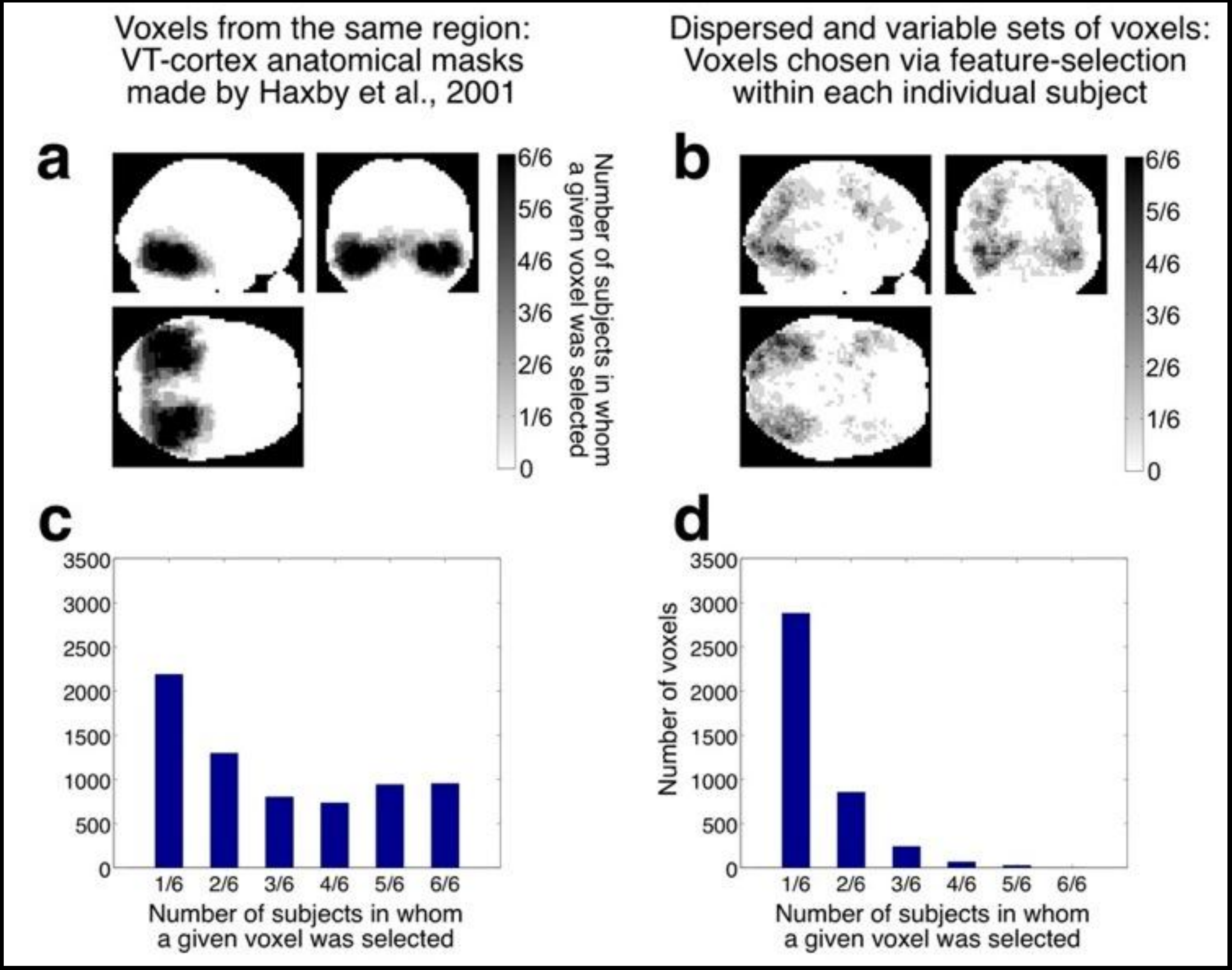
In that analysis, everyone's voxels were drawn from the same region: VT-cortex

What about if we use different sets of voxels for each individual subject?

Feature-selection. Pick voxels which are:

- In the top 5% of active voxels (univariate t-test, objects > rest), and
- In the top 5% of discriminative voxels (univariate F-test: between/with class-variance)

What about neural diversity?



Results in the presence of neural diversity

Result: 87.5% correct

- 42 out of 48 decodings correct

Four subjects: all eight categories correct

One subject: 4 out of 8 correct

- Confusions: cat, chair, face and scissors

Other subject: 6 out of 8 correct

- Confusions: bottles and scissors

Number of voxels selected within each subject ranged from 473 to 1346

Comparison to previous studies of across-subject fMRI decoding

People's large-scale brain-states are alike

It's their fine-scale representations which differ

Across-subject decoding of large-scale brain-states:

- Reading a sentence vs. looking at a picture (Wang et al, 2003)
- Face-matching vs. location-matching (Mourao-Miranda et al, 2005)
- Decoding between different cognitive tasks (Poldrack et al, 2009)
- Monetary-reward vs. viewing attractive face (Clithero et al, 2010)

Two very interesting approaches: Tom Mitchell, Jim Haxby

These approaches are still applying decoding to the neural activation itself, rather than to similarity-space

Shinkareva, Mitchell and colleagues (PLoS ONE, 2008)

- Could decode category: tool vs. dwelling
- Also: which of the five specific exemplars within each category the subjects were looking at

However:

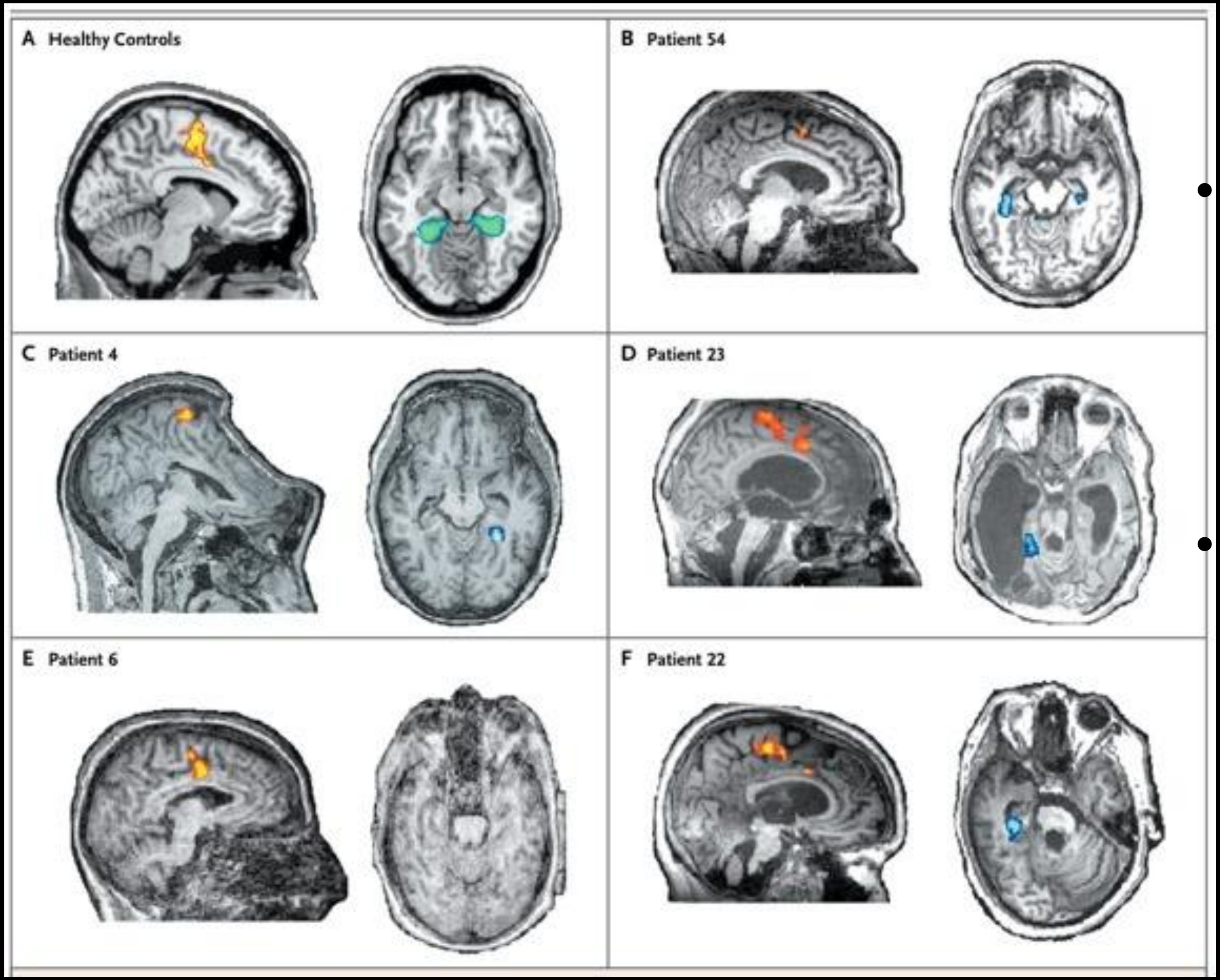
- Low accuracy (reported rank-accuracy, not percentage correct)
- Above chance for eight of the twelve subjects

Haxby, Guntupalli and colleagues

- Have proposed a high-dimensional mapping, called “hyper-alignment”, of one person's voxel space onto another’s

Who cares?

If you can assign meaning to neural activation across different individuals, you can do interesting things



- Monti et al. “Willful modulation of brain activity in disorders of consciousness”, NEJM, 2010.
- Answer “Yes” or “No”, by imagining playing tennis, or by imagining navigating the streets of a familiar city

Who cares?

Why just the two responses of “Yes” and “No” ?

Why use motor imagery and navigation?

- Those are amongst the very few types of activation that will reliably occur in the same part of brain across different people
- You can “decode” as “yes” or “no” simply by seeing which part of the brain lights up
- This decoding will work for any individual, because everybody has motor-cortex and parahippocampal gyrus in more or less the same place

What if we could decode a broader range of meanings, at a finer-grain?

And if we were not restricted to tasks which happen to be obliging enough to specifically light up a single region?

Thanks!