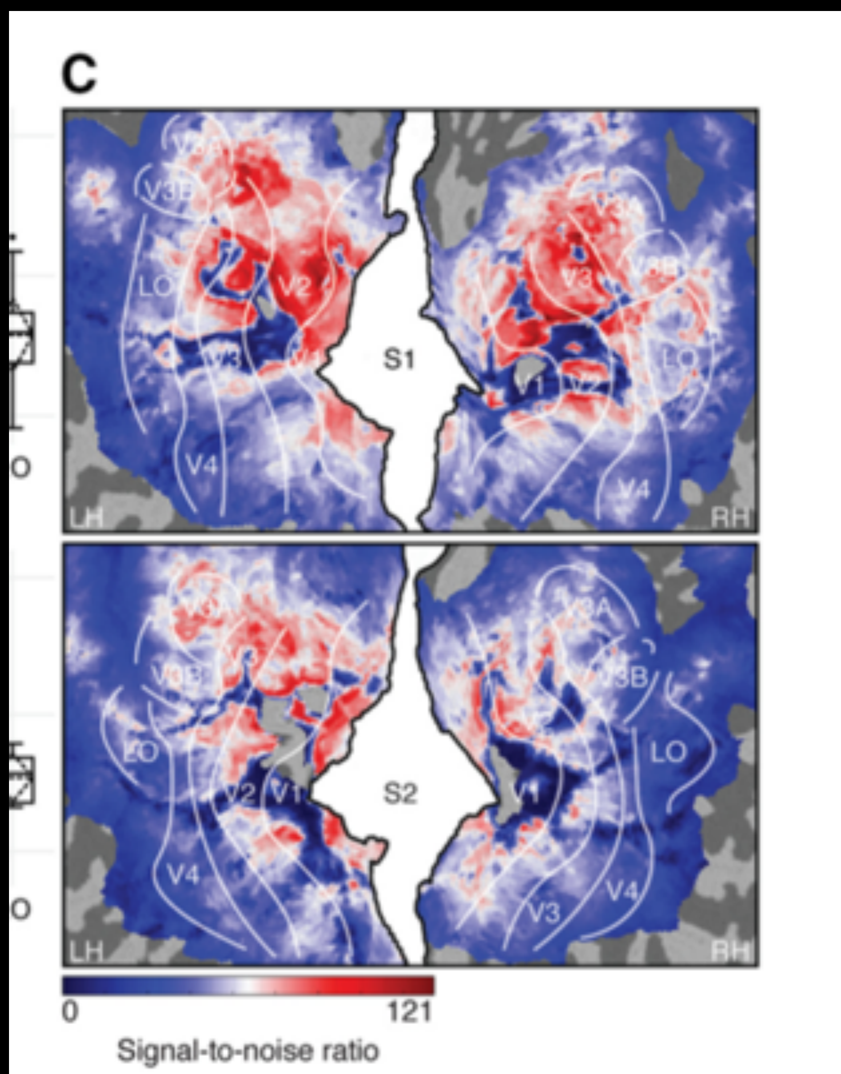
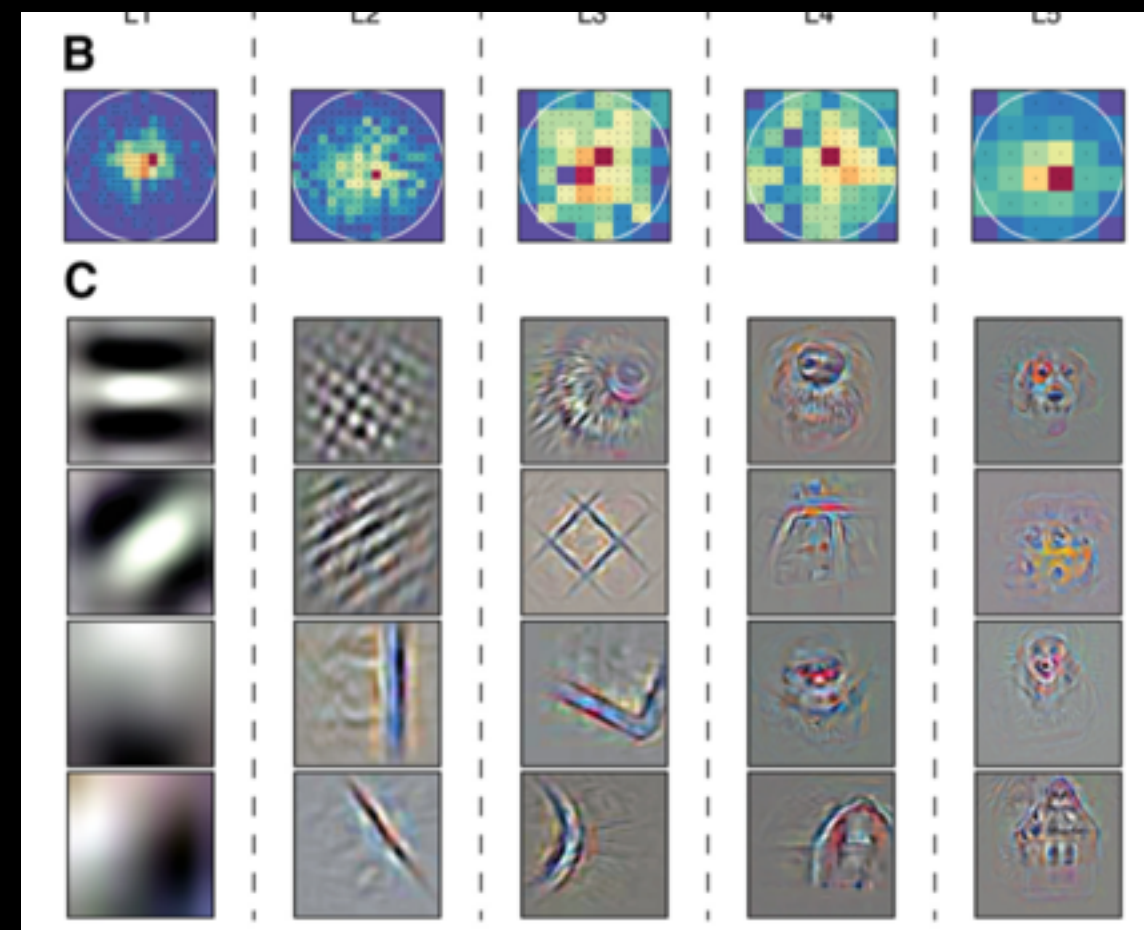
Machine Learning applications in Cognitive Neuroscience and psychology





Gilles de Hollander
10 May 2016

Overview

1. Machine Learning as a statistical tool
    - “Brain reading”
2. Machine learning as a model of the brain
    - Mechanistic model of visual cortex
    - Cognitive model of categorisation
3. Machine Learning and the Simon Task

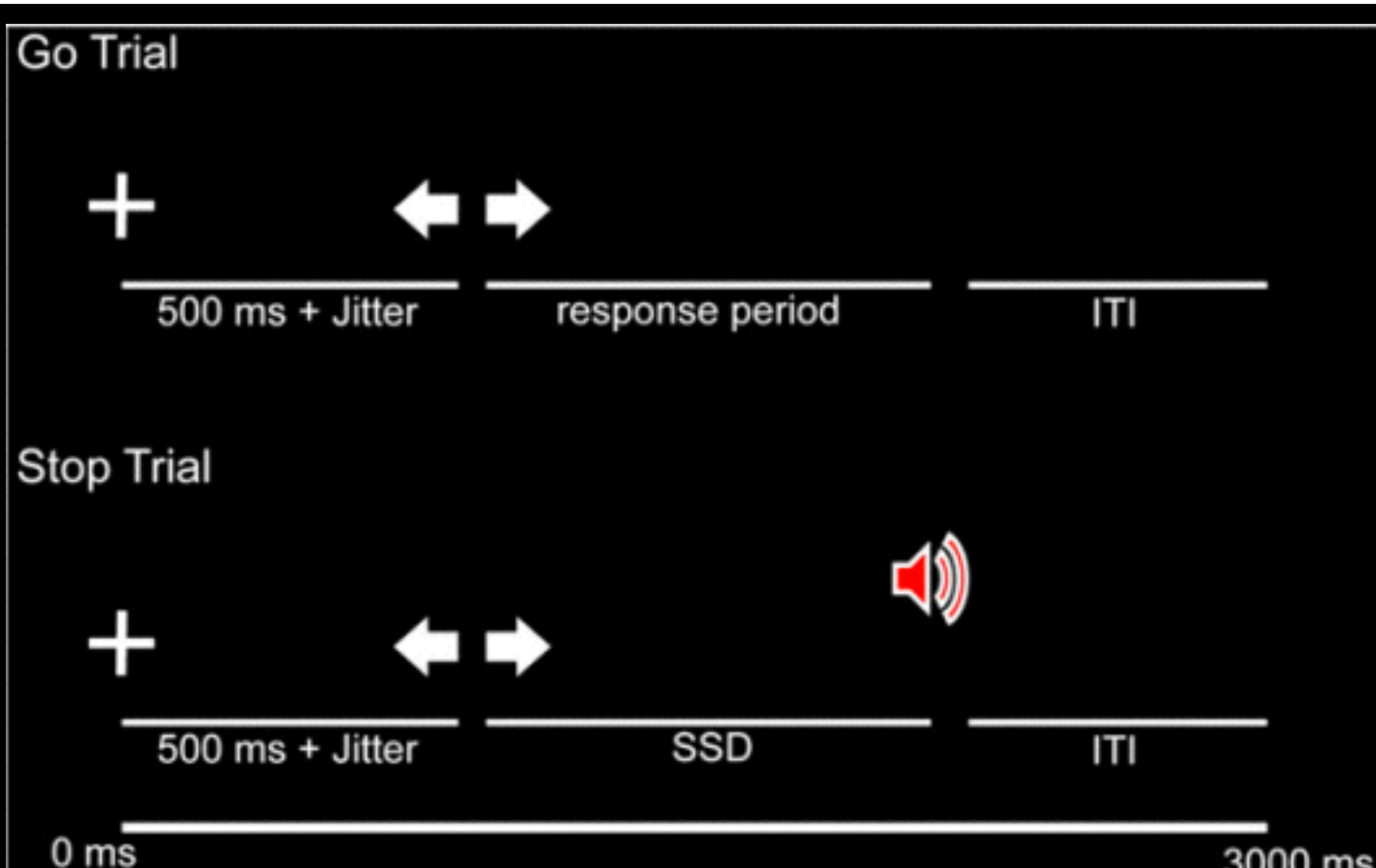
16286 • The Journal of Neuroscience, December 3, 2014 • 34(49):16286–16295

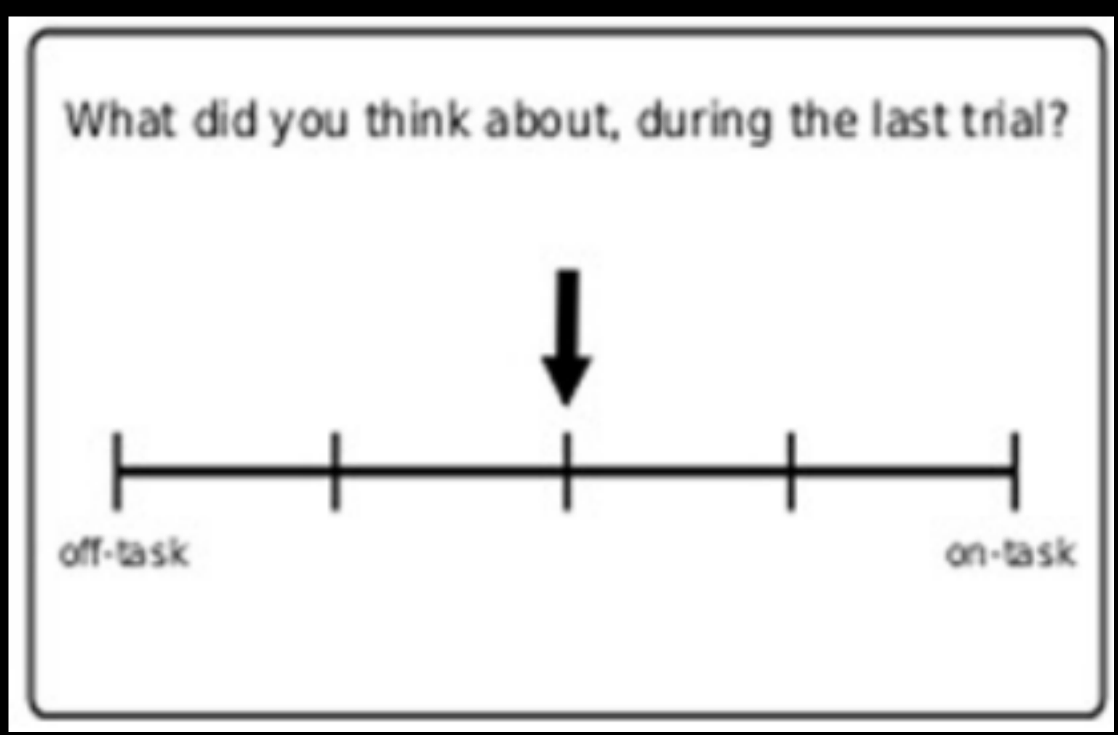
Behavioral/Cognitive

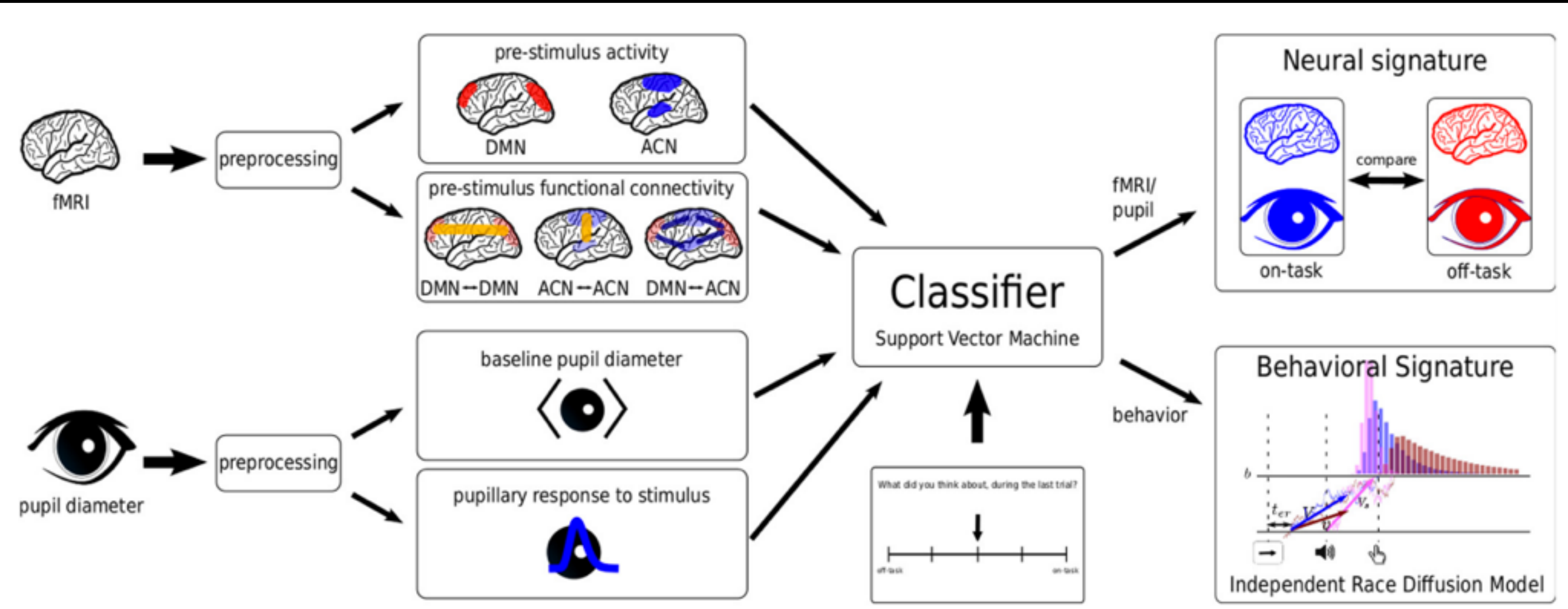
When the Brain Takes a Break: A Model-Based Analysis of Mind Wandering

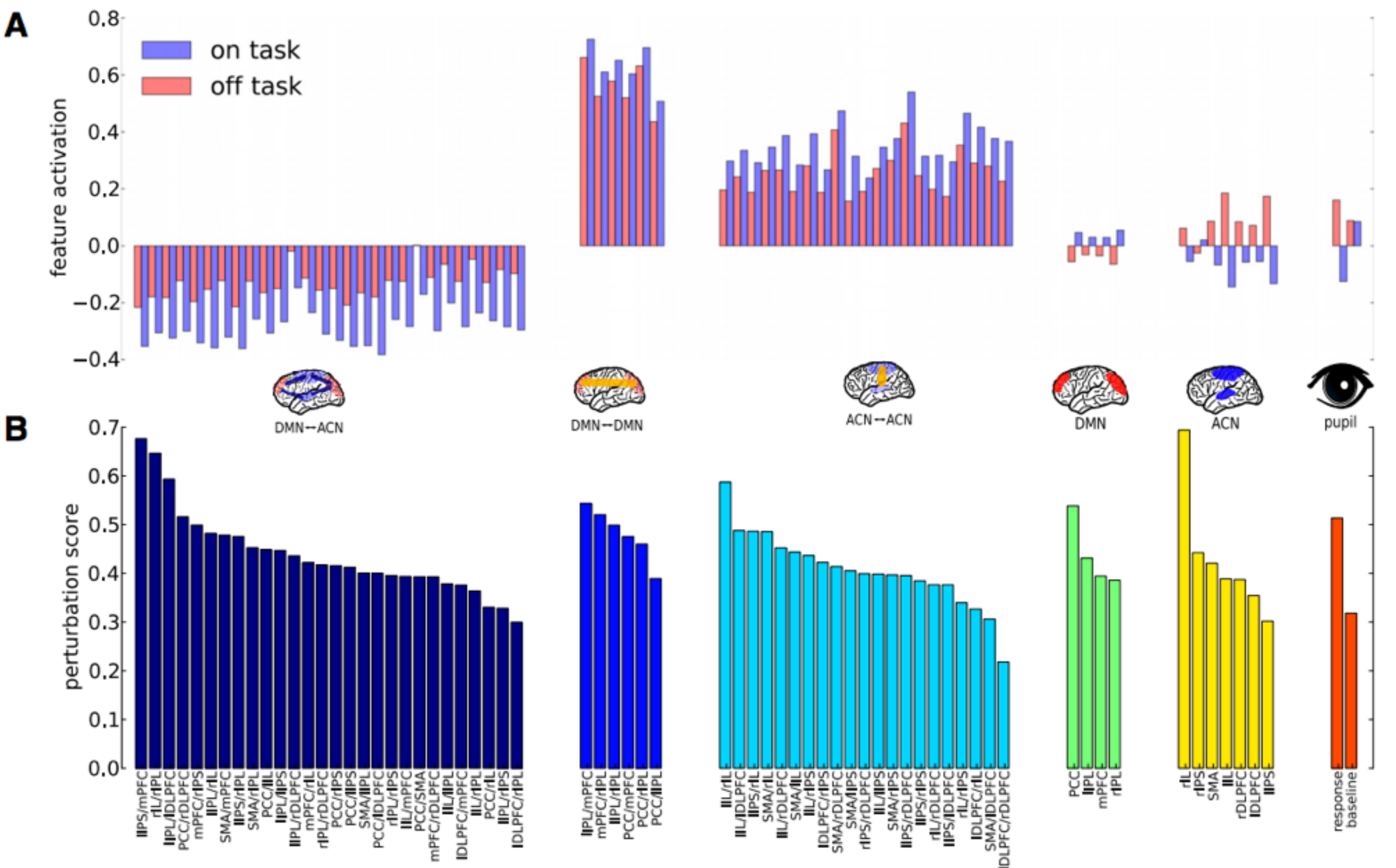
Matthias Mittner,[1] Wouter Boekel,[2] Adrienne M. Tucker,[2] Brandon M. Turner,[4] Andrew Heathcote,[3] and Birte U. Forstmann[2]

[1]Department of Psychology, University of Tromsø, 9037 Tromsø, Norway, [2]Cognitive Science Center Amsterdam, 1018 VZ Amsterdam, The Netherlands, [3]Newcastle Cognition Laboratory, School of Psychology, University of Newcastle, 2308 Newcastle, Australia, and [4]Stanford University Center for Mind, Brain and Computation, Department of Psychology, Stanford University, Stanford, California 94305



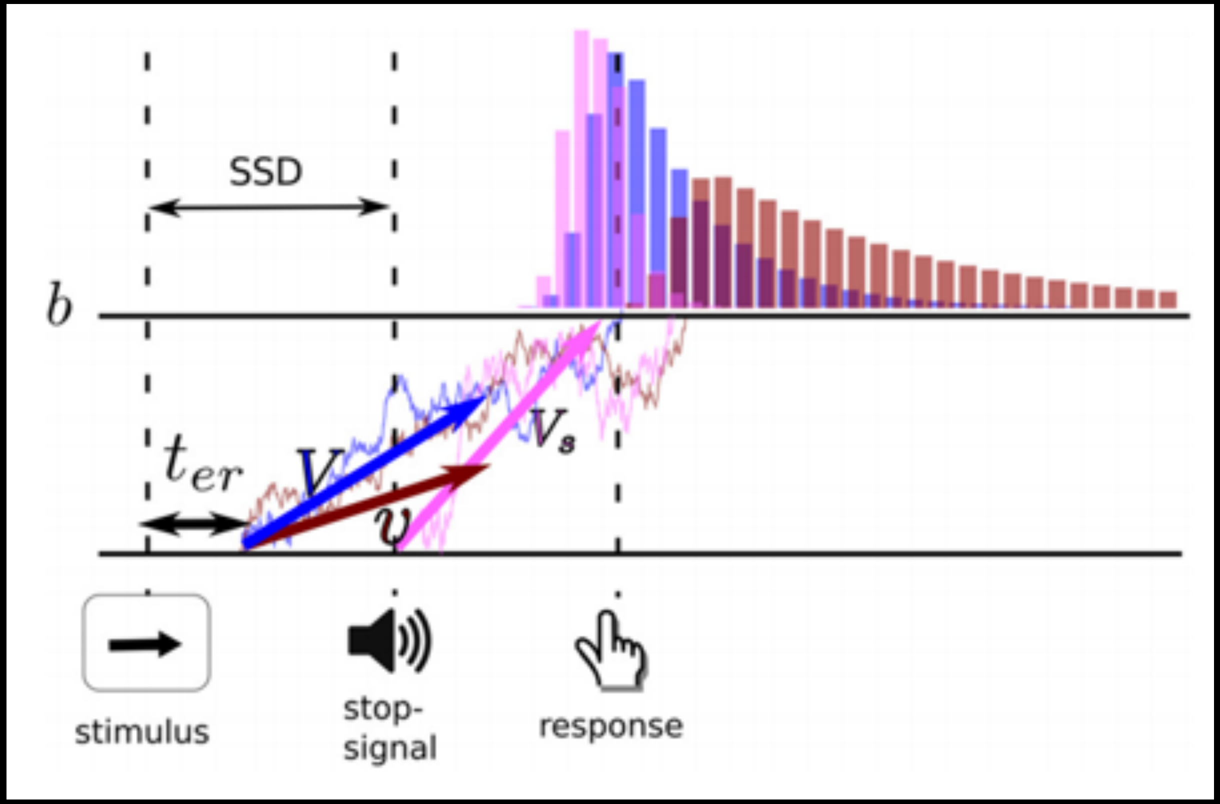



fMRI
preprocessing
pupil diameter
preprocessing
pre-stimulus activity
DMN
ACN
pre-stimulus functional connectivity
DMN↔DMN
ACN↔ACN
DMN↔ACN
baseline pupil diameter
pupillary response to stimulus
Classifier
Support Vector Machine
What did you think about, during the last trial?
fMRI/
pupil
behavior
Neural signature
compare
on-task
off-task
Behavioral Signature
Independent Race Diffusion Model

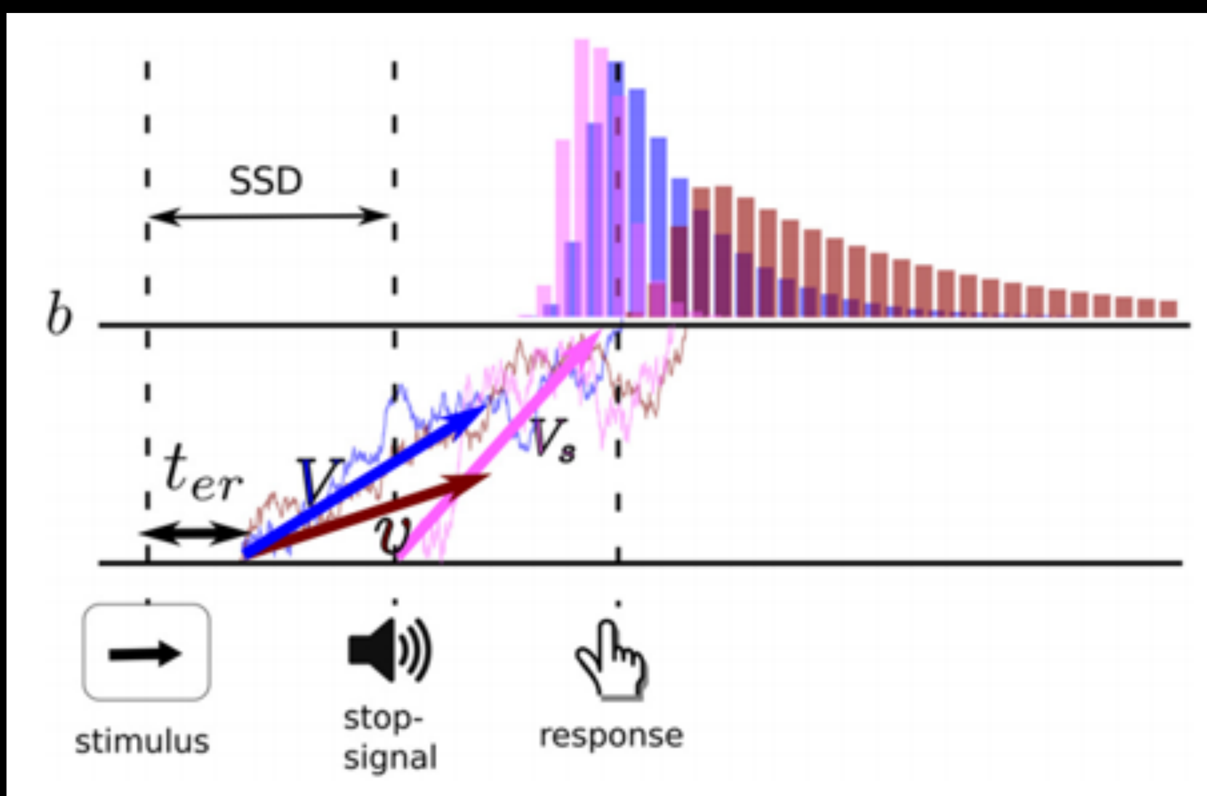


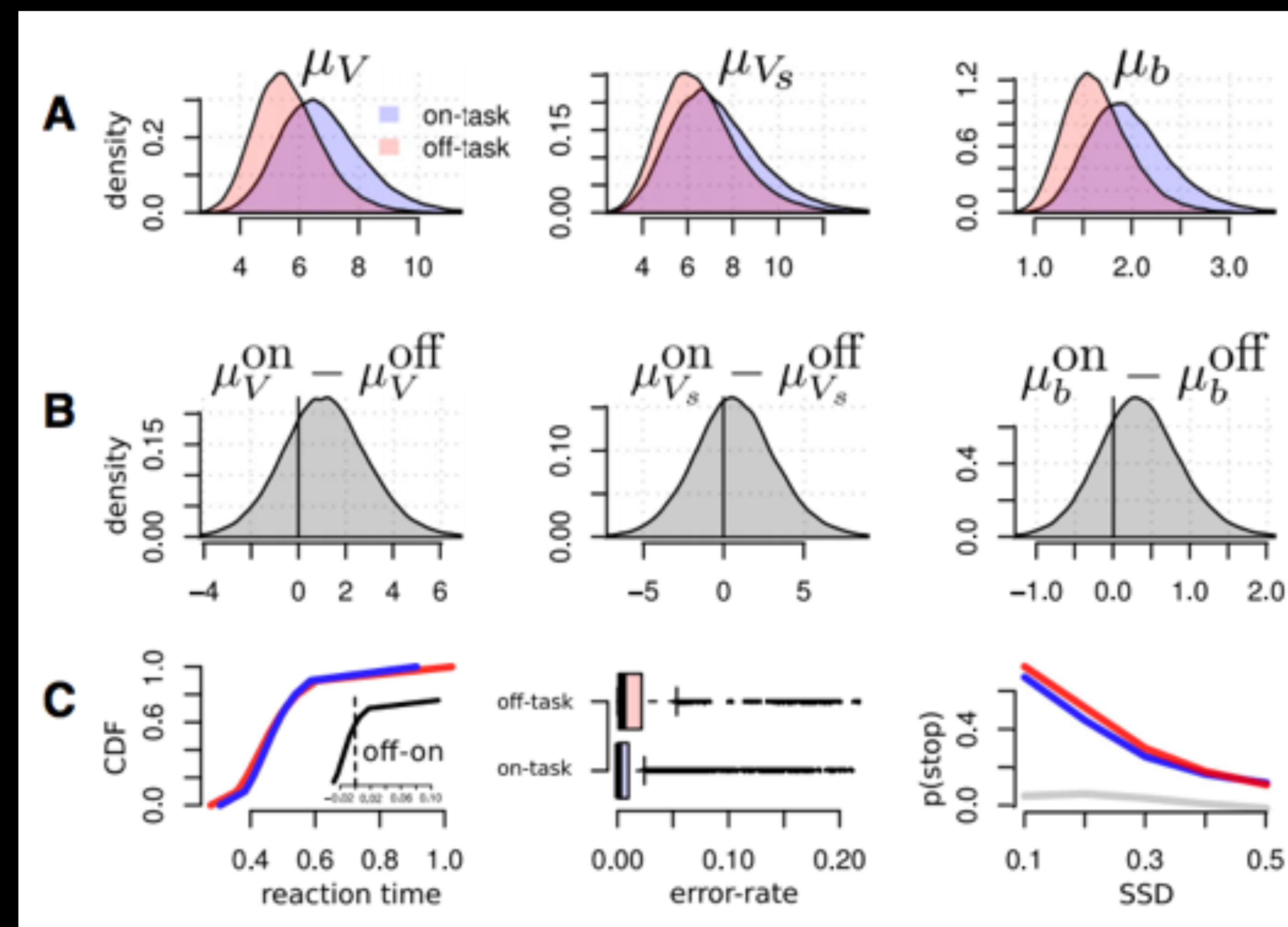
A
on task
off task
feature activation
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
DMN↔ACN
DMN↔DMN
ACN↔ACN
DMN
ACN
pupil
B
perturbation score
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
lIPS/mPFC
rIL/rIPL
lIPL/lDLPFC
PCC/rDLPFC
mPFC/rIPS
lIPL/rIL
SMA/mPFC
lIPS/rIPL
SMA/rIPL
PCC/lIL
lIPL/lIPS
lIPL/rDLPFC
mPFC/rIL
rIPL/rDLPFC
PCC/rIPS
PCC/lIPS
SMA/lIPL
PCC/lDLPFC
rIPL/rIPS
lIL/mPFC
PCC/SMA
mPFC/rDLPFC
lIL/lIPL
lDLPFC/mPFC
lIL/rIPL
PCC/rIL
lIPL/rIPS
lDLPFC/rIPL
lIPL/mPFC
mPFC/rIPL
lIPL/rIPL
PCC/mPFC
PCC/rIPL
PCC/lIPL
lIL/rIL
lIL/lDLPFC
lIPS/rIL
SMA/rIL
lIL/rDLPFC
SMA/lIL
lIL/rIPS
lDLPFC/rIPS
SMA/rDLPFC
SMA/lIPS
rIPS/rDLPFC
lIL/lIPS
SMA/rIPS
lIPS/rDLPFC
lIPS/rIPS
rIL/rDLPFC
lIPS/lDLPFC
rIL/rIPS
lDLPFC/rIL
SMA/lDLPFC
lDLPFC/rDLPFC
PCC
lIPL
mPFC
rIPL
rIL
rIPS
SMA
lIL
rDLPFC
lDLPFC
lIPS
response
baseline

Race stopping model



Race stopping model

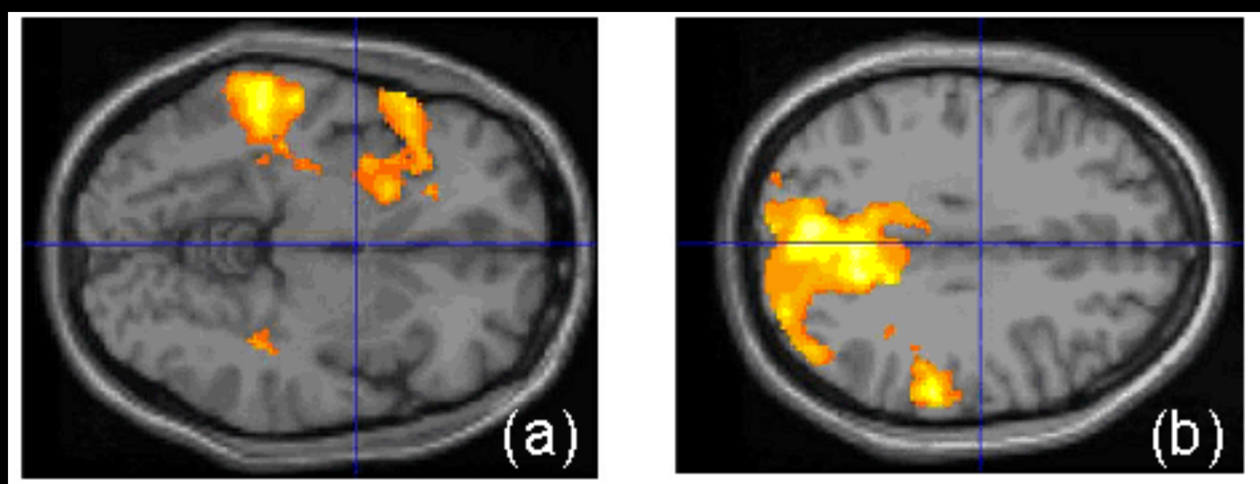


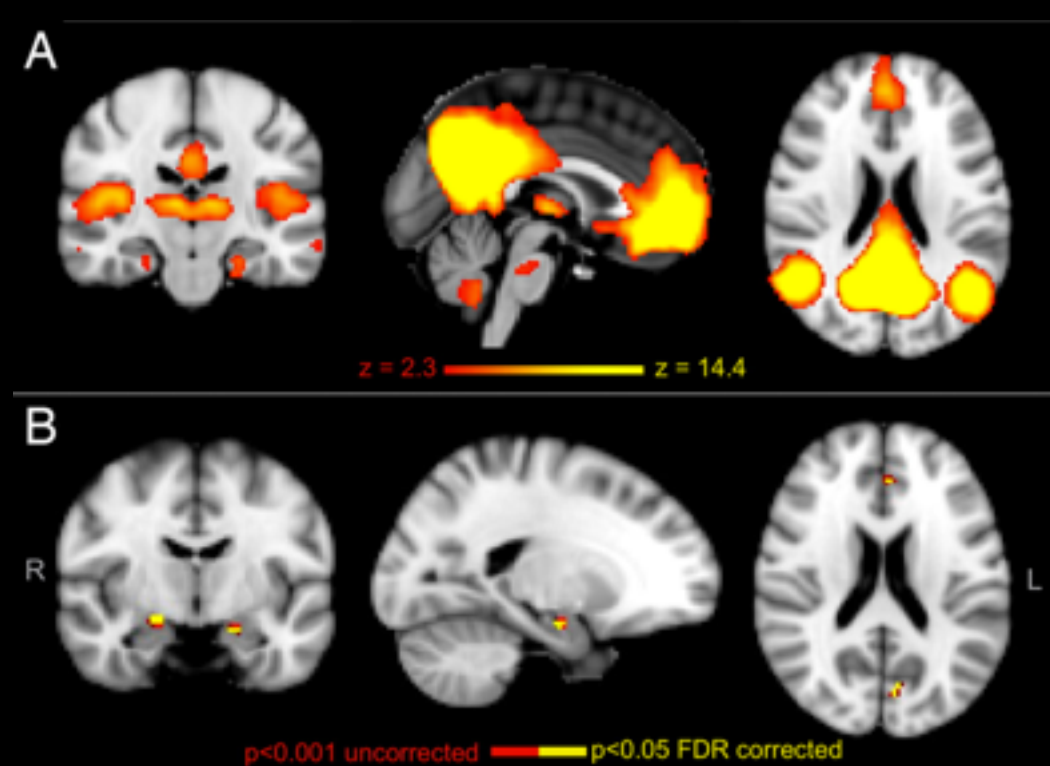


What do we learn

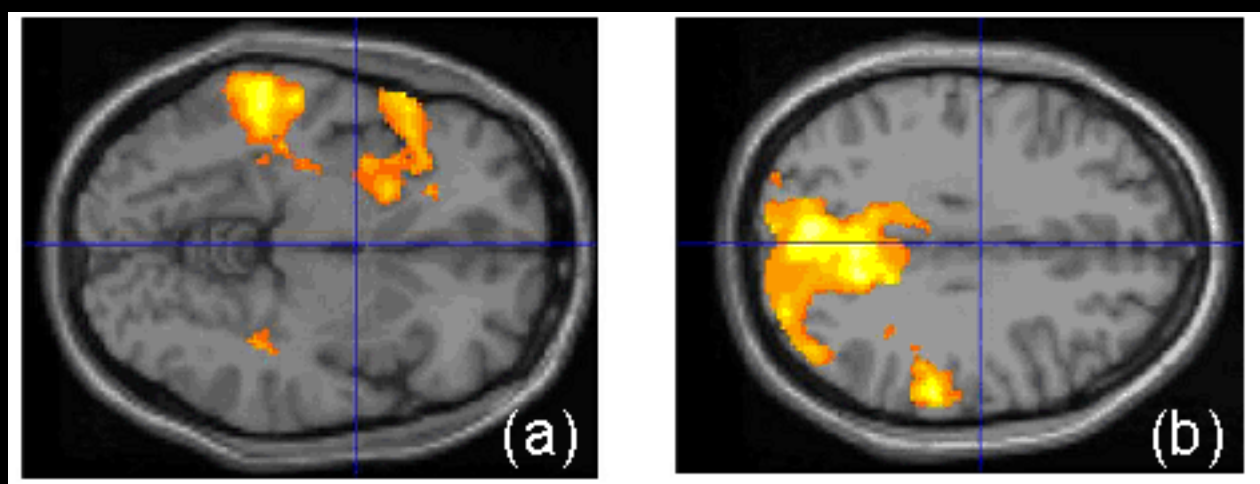
- We can “decode” mind-wandering
    - All modalities are informative
- No real mechanistic understanding
    - Cognitive model can help to some extent

Functional MRI



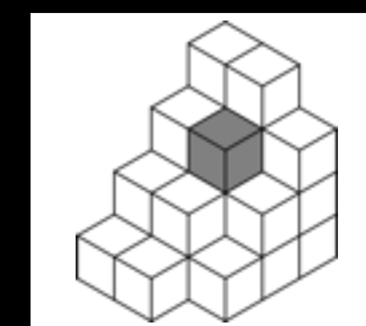


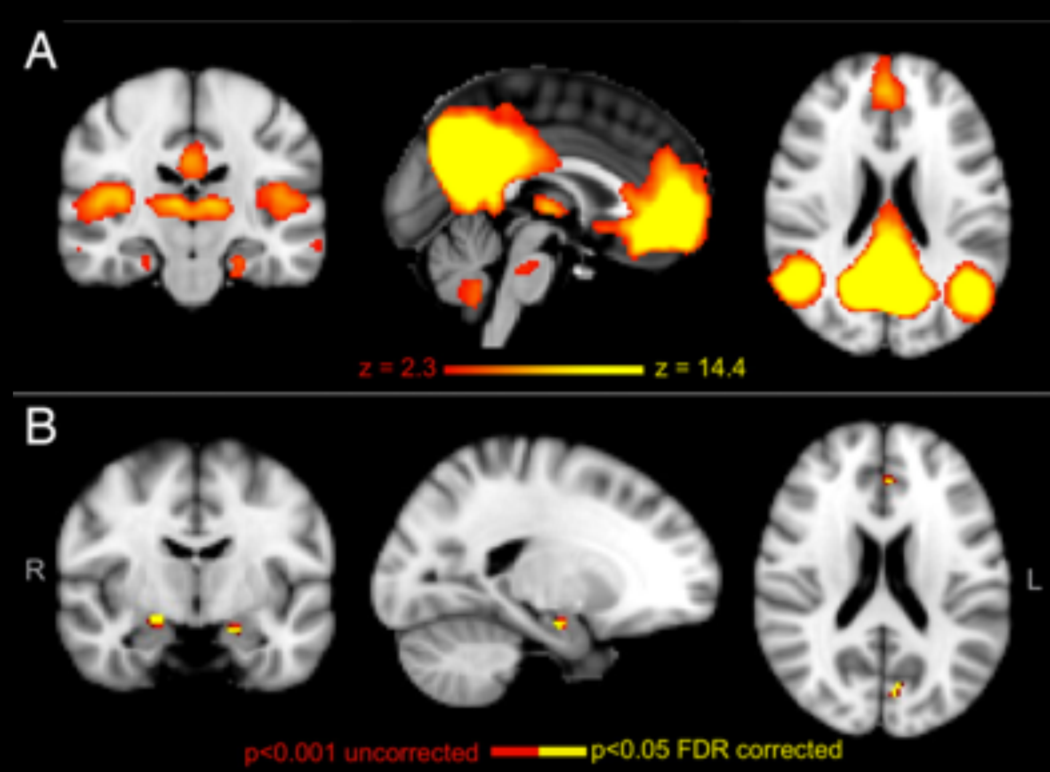
Functional MRI



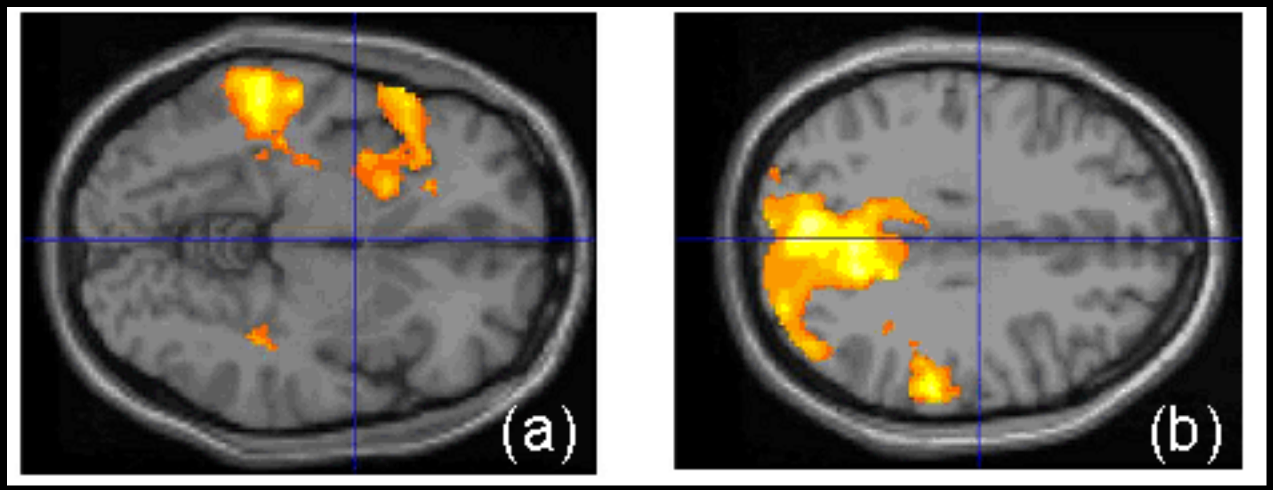
Massively univariate

100.000 t-tests



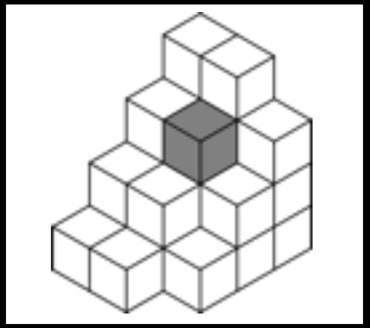


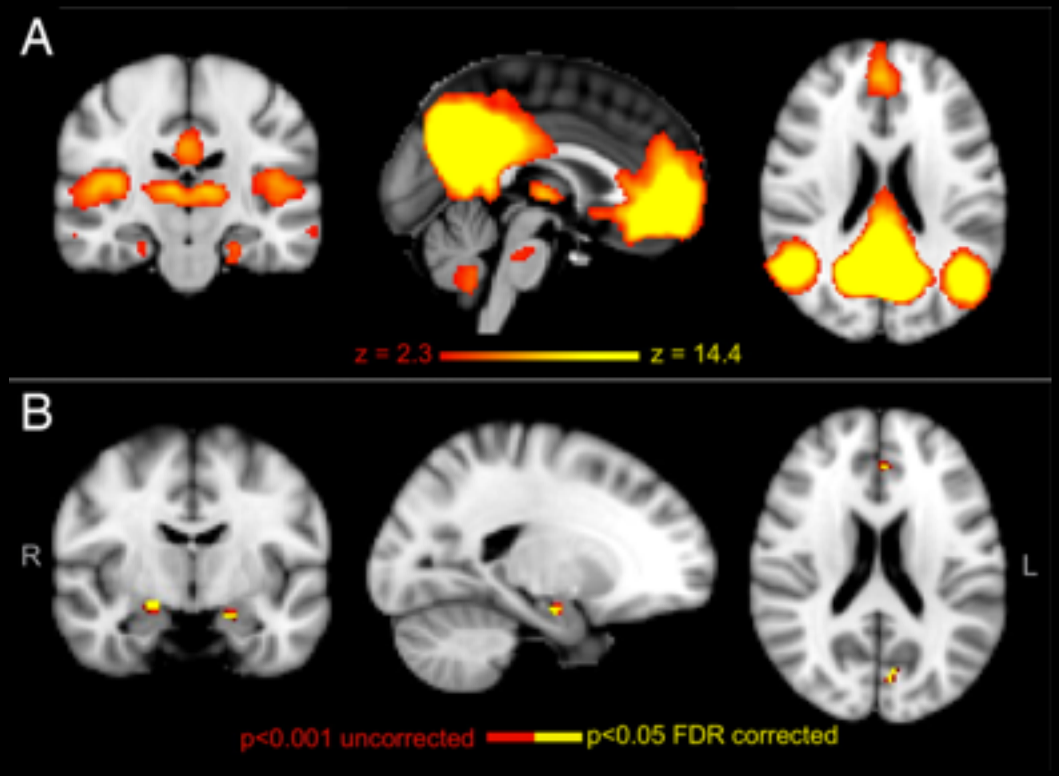
Functional MRI

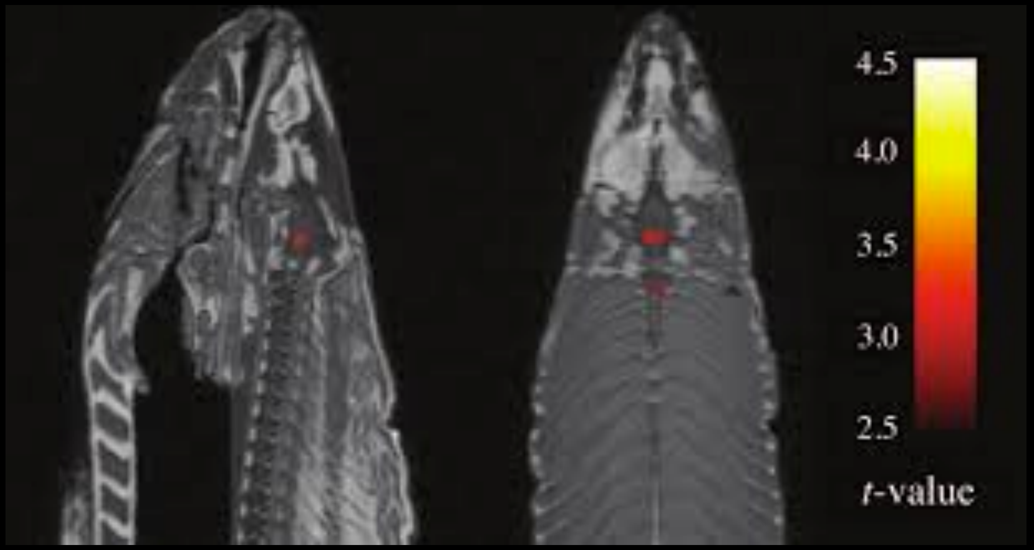


Massively univariate

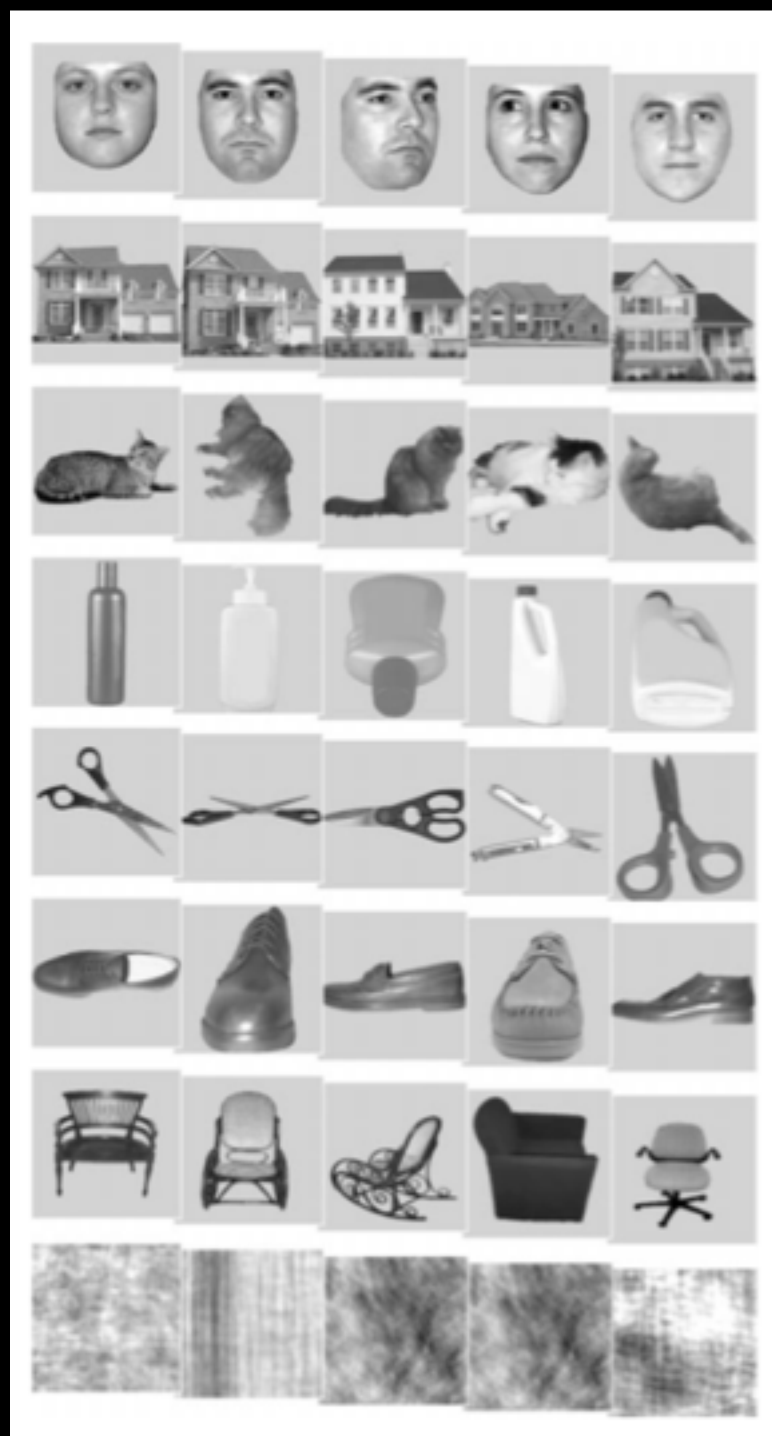
100.000 t-tests

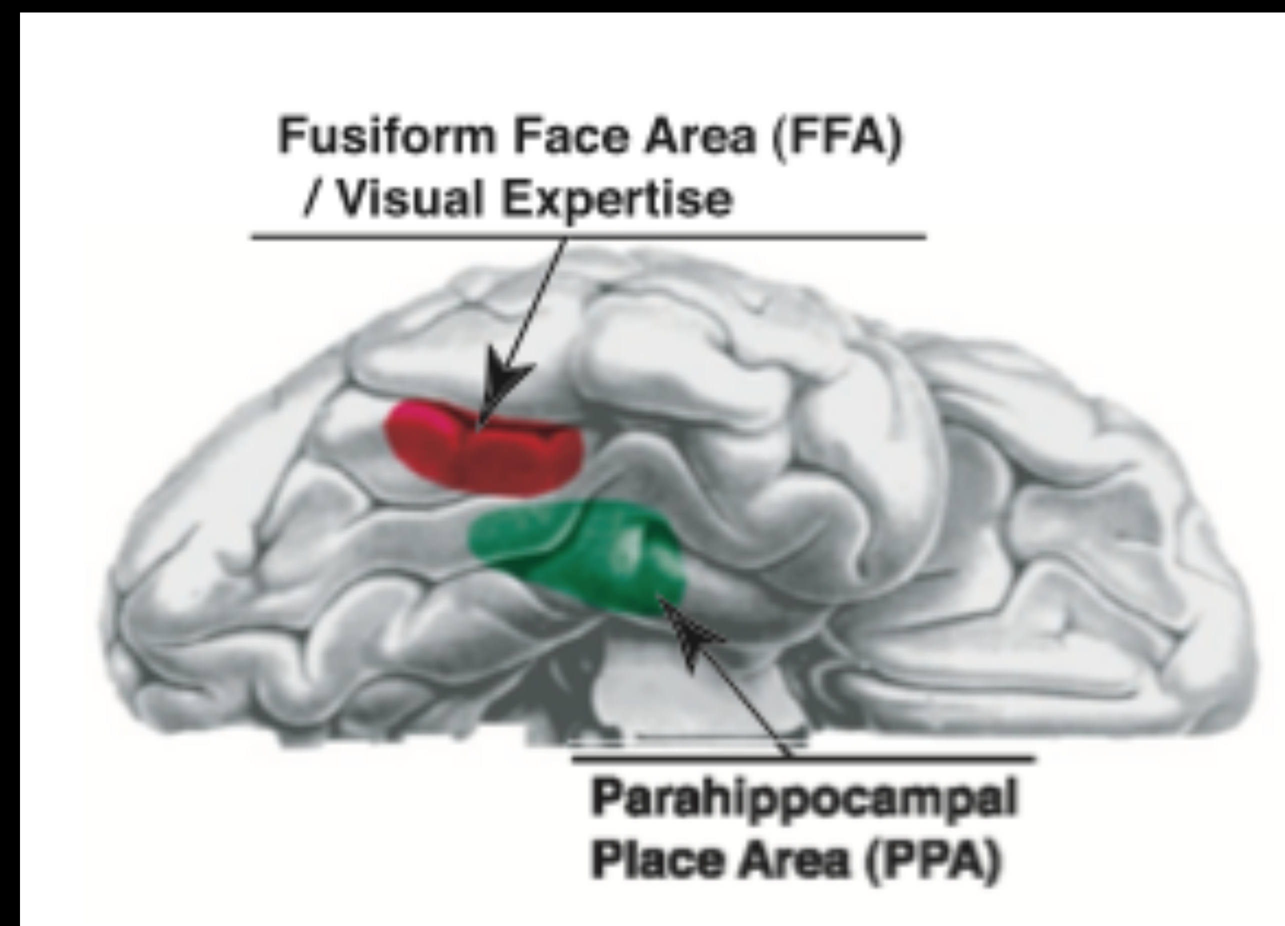


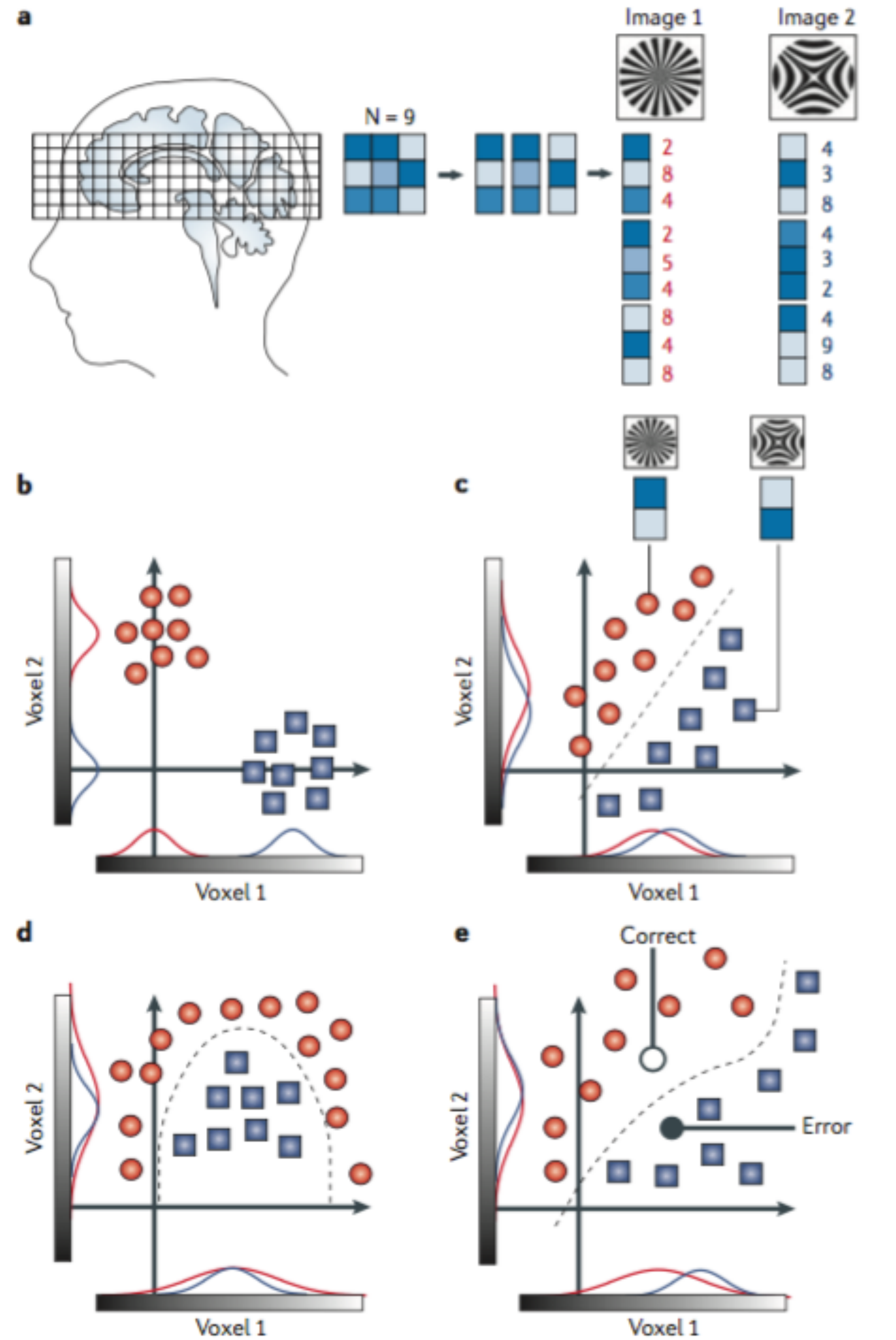


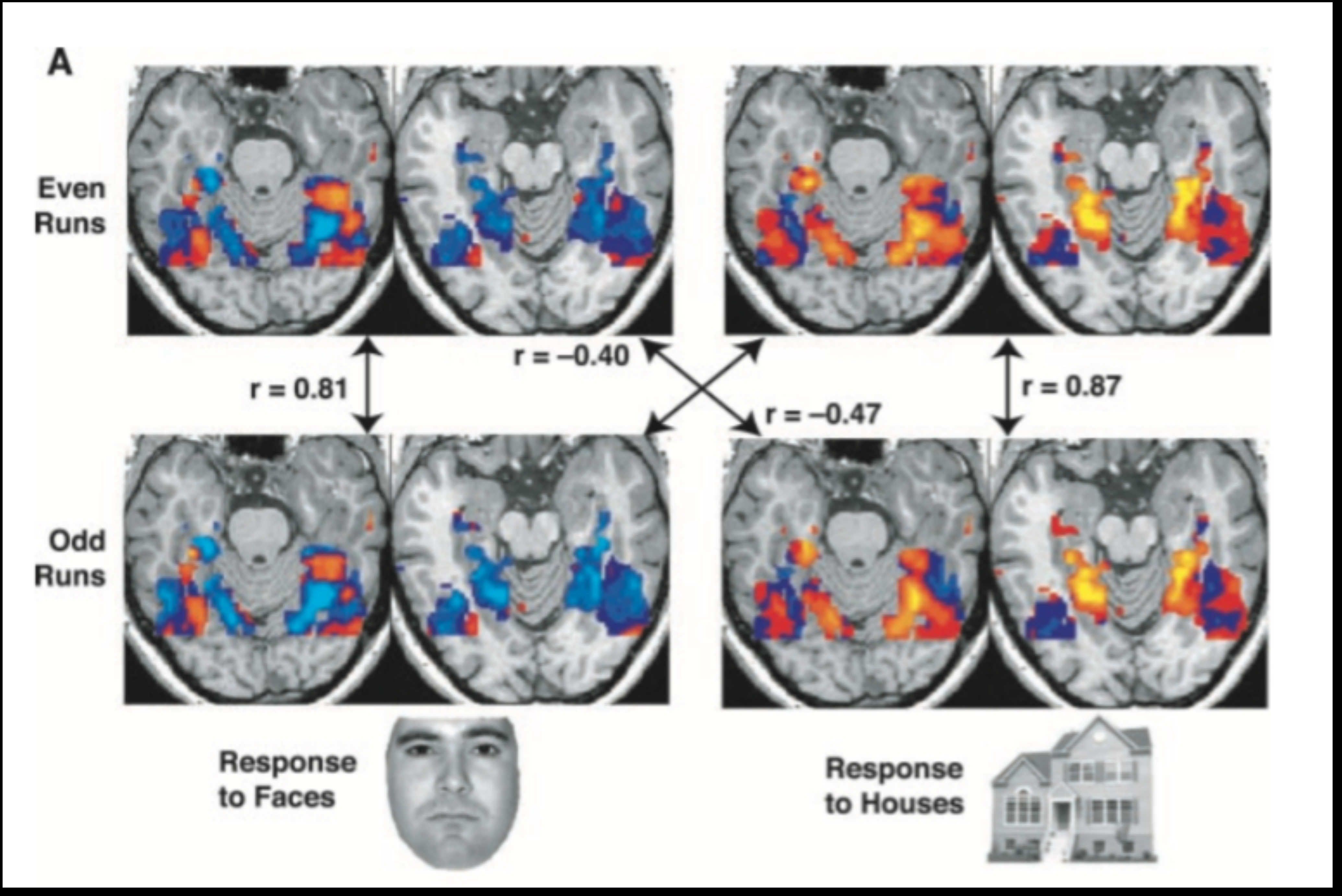


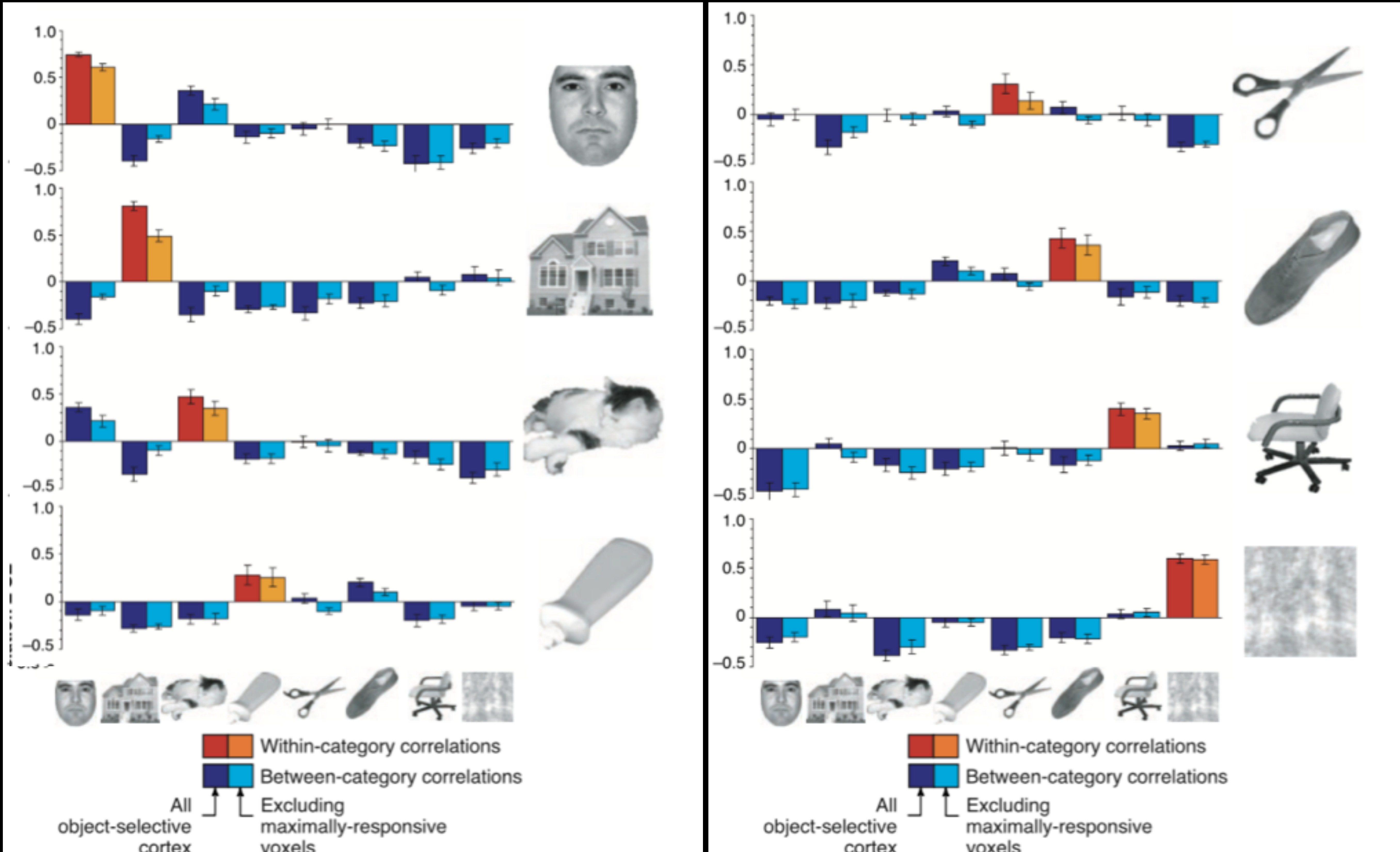
Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex
James V. Haxby *et al.*
Science **293**, 2425 (2001);
DOI: 10.1126/science.1063736






a
Image 1
Image 2
N = 9
2
8
4
2
5
4
8
4
8
4
3
8
4
3
2
4
9
8
b
Voxel 2
Voxel 1
c
Voxel 2
Voxel 1
d
Voxel 2
Voxel 1
e
Correct
Error
Voxel 2
Voxel 1


A
Even Runs
Odd Runs
r = 0.81
r = –0.40
r = –0.47
r = 0.87
Response to Faces
Response to Houses



1.0
0.5
0
−0.5
Within-category correlations
Between-category correlations
All object-selective cortex
Excluding maximally-responsive voxels

NEUROIMAGING

Decoding mental states from brain activity in humans

John-Dylan Haynes*‡§ and Geraint Rees‡§

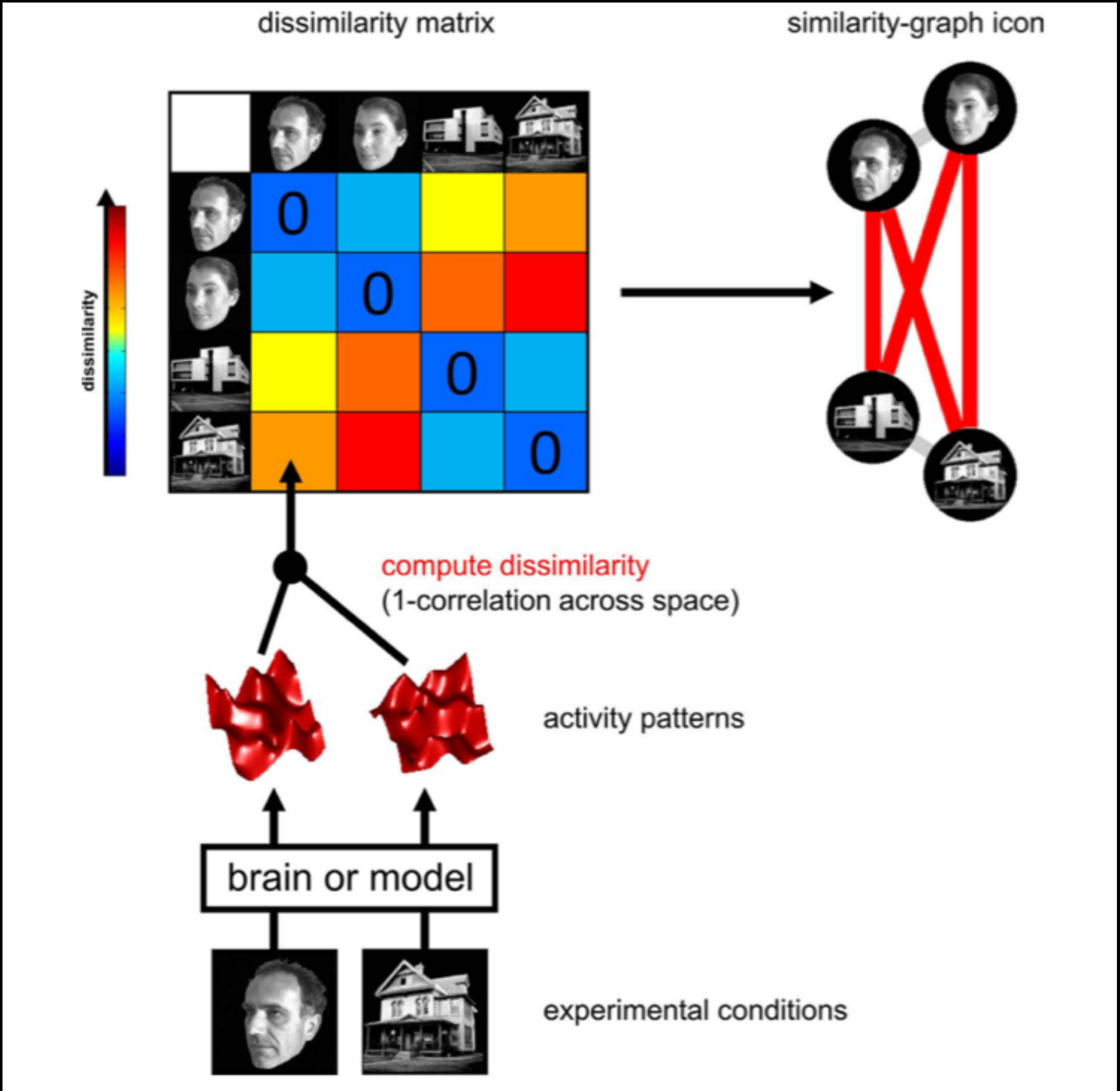
Representational similarity analysis – connecting the branches of systems neuroscience

Nikolaus Kriegeskorte[1,*], Marieke Mur[1,2] and Peter Bandettini[1]

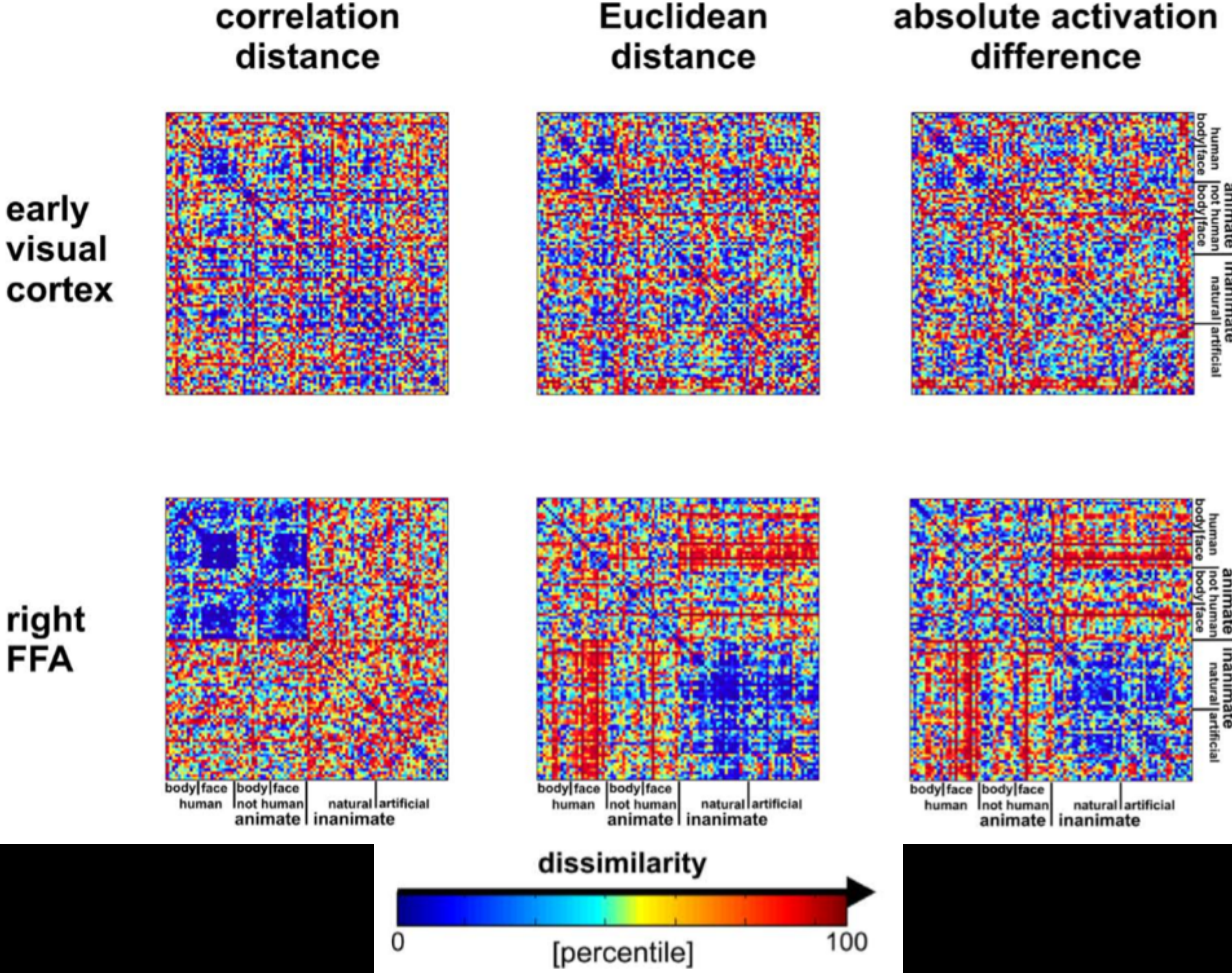
Beyond mind-reading: multi-voxel pattern analysis of fMRI data

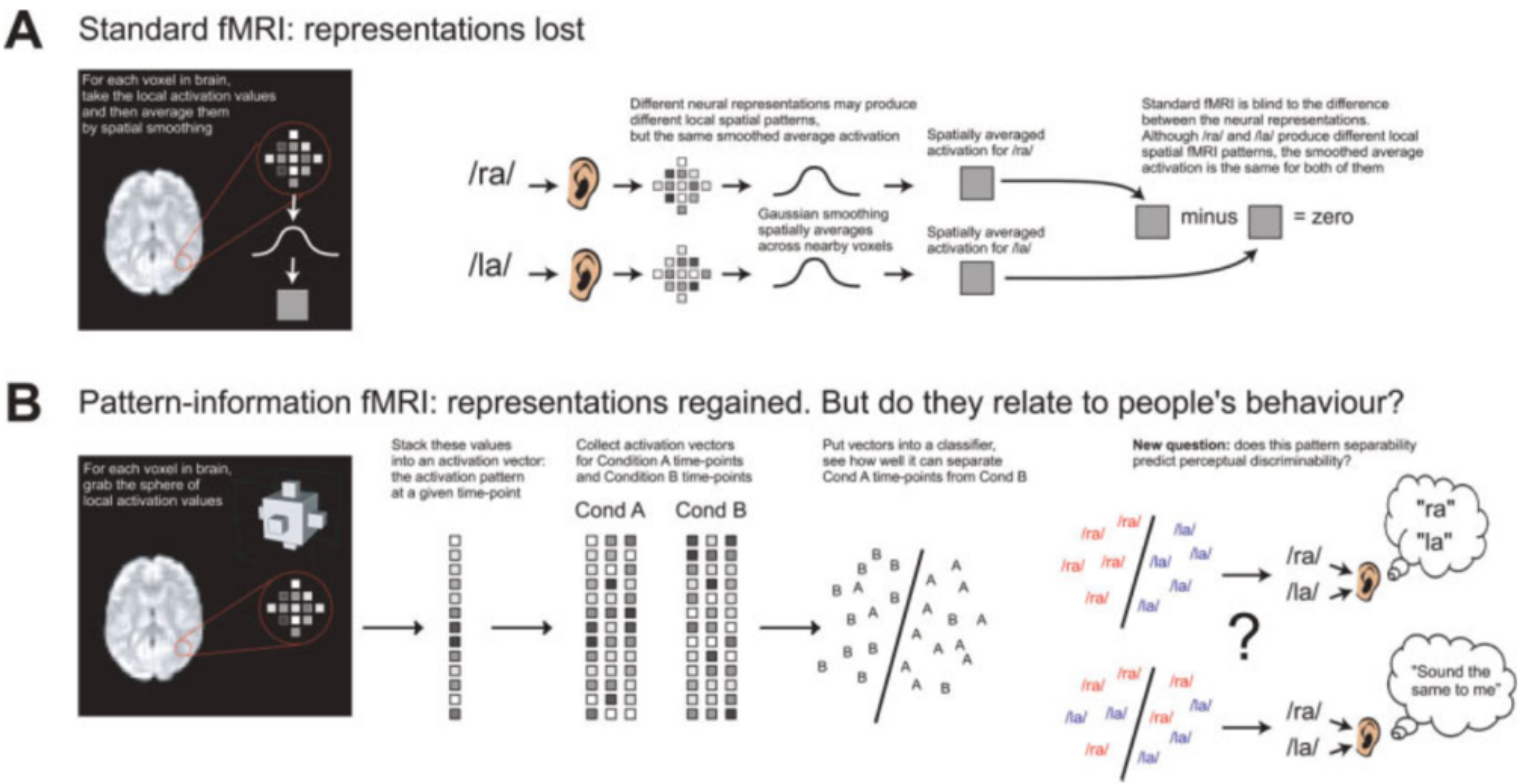
Kenneth A. Norman[1], Sean M. Polyn[2], Greg J. Detre[1] and James V. Haxby[1]

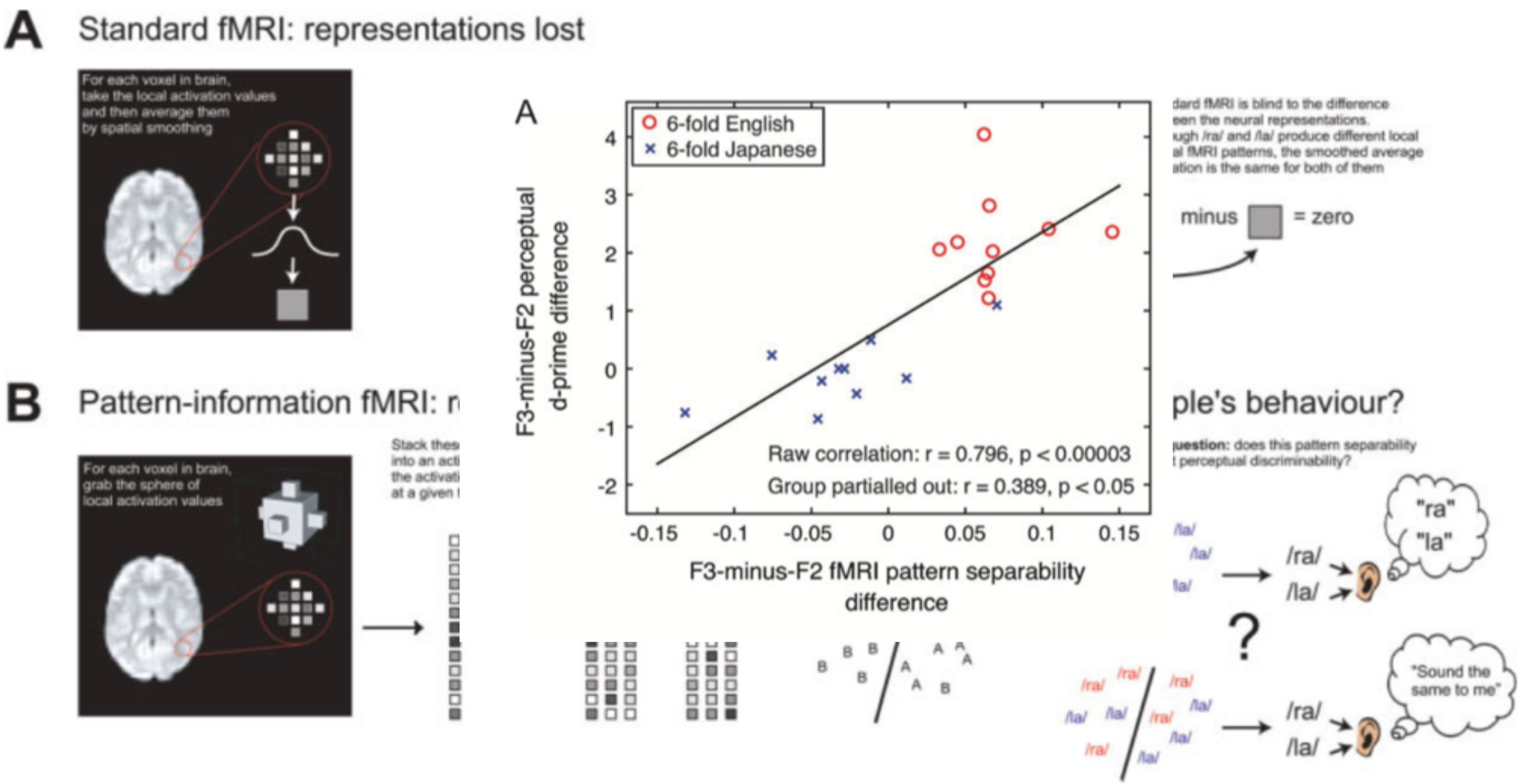
[1] Department of Psychology, Princeton University, Green Hall, Washington Road, Princeton, NJ 08540, USA
[2] Department of Psychology, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104, USA



dissimilarity matrix
similarity-graph icon
dissimilarity
0
0
0
0
compute dissimilarity
(1-correlation across space)
activity patterns
brain or model
experimental conditions


correlation distance
Euclidean distance
absolute activation difference
early visual cortex
right FFA
body | face
human
body | face
not human
animate
natural | artificial
inanimate
dissimilarity
0
[percentile]
100


A Standard fMRI: representations lost
For each voxel in brain, take the local activation values and then average them by spatial smoothing
Different neural representations may produce different local spatial patterns, but the same smoothed average activation
/ra/
/la/
Gaussian smoothing spatially averages across nearby voxels
Spatially averaged activation for /ra/
Spatially averaged activation for /la/
Standard fMRI is blind to the difference between the neural representations. Although /ra/ and /la/ produce different local spatial fMRI patterns, the smoothed average activation is the same for both of them
minus
= zero
B Pattern-information fMRI: representations regained. But do they relate to people's behaviour?
For each voxel in brain, grab the sphere of local activation values
Stack these values into an activation vector: the activation pattern at a given time-point
Collect activation vectors for Condition A time-points and Condition B time-points
Cond A
Cond B
Put vectors into a classifier, see how well it can separate Cond A time-points from Cond B
New question: does this pattern separability predict perceptual discriminability?
"ra"
"la"
"Sound the same to me"
?



A
Standard fMRI: representations lost
For each voxel in brain, take the local activation values and then average them by spatial smoothing
dard fMRI is blind to the difference
een the neural representations.
ugh /ra/ and /la/ produce different local
al fMRI patterns, the smoothed average
ation is the same for both of them
minus
= zero
B
Pattern-information fMRI: r
ple's behaviour?
For each voxel in brain, grab the sphere of local activation values
Stack thes
into an acti
the activati
at a given t
question: does this pattern separability
t perceptual discriminability?
"ra"
"la"
/ra/
/la/
?
"Sound the same to me"
A
6-fold English
6-fold Japanese
F3-minus-F2 perceptual d-prime difference
F3-minus-F2 fMRI pattern separability difference
Raw correlation: r = 0.796, p < 0.00003
Group partialled out: r = 0.389, p < 0.05

Supervised MVPA

- Most fMRI research use some sort of paradigm with clearly separable conditions

Correspondence of the brain's functional architecture during activation and rest

Stephen M. Smith[a,1], Peter T. Fox[b], Karla L. Miller[a], David C. Glahn[b,c], P. Mickle Fox[b], Clare E. Mackay[a], Nicola Filippini[a], Kate E. Watkins[a], Roberto Toro[d], Angela R. Laird[b], and Christian F. Beckmann[a,e]

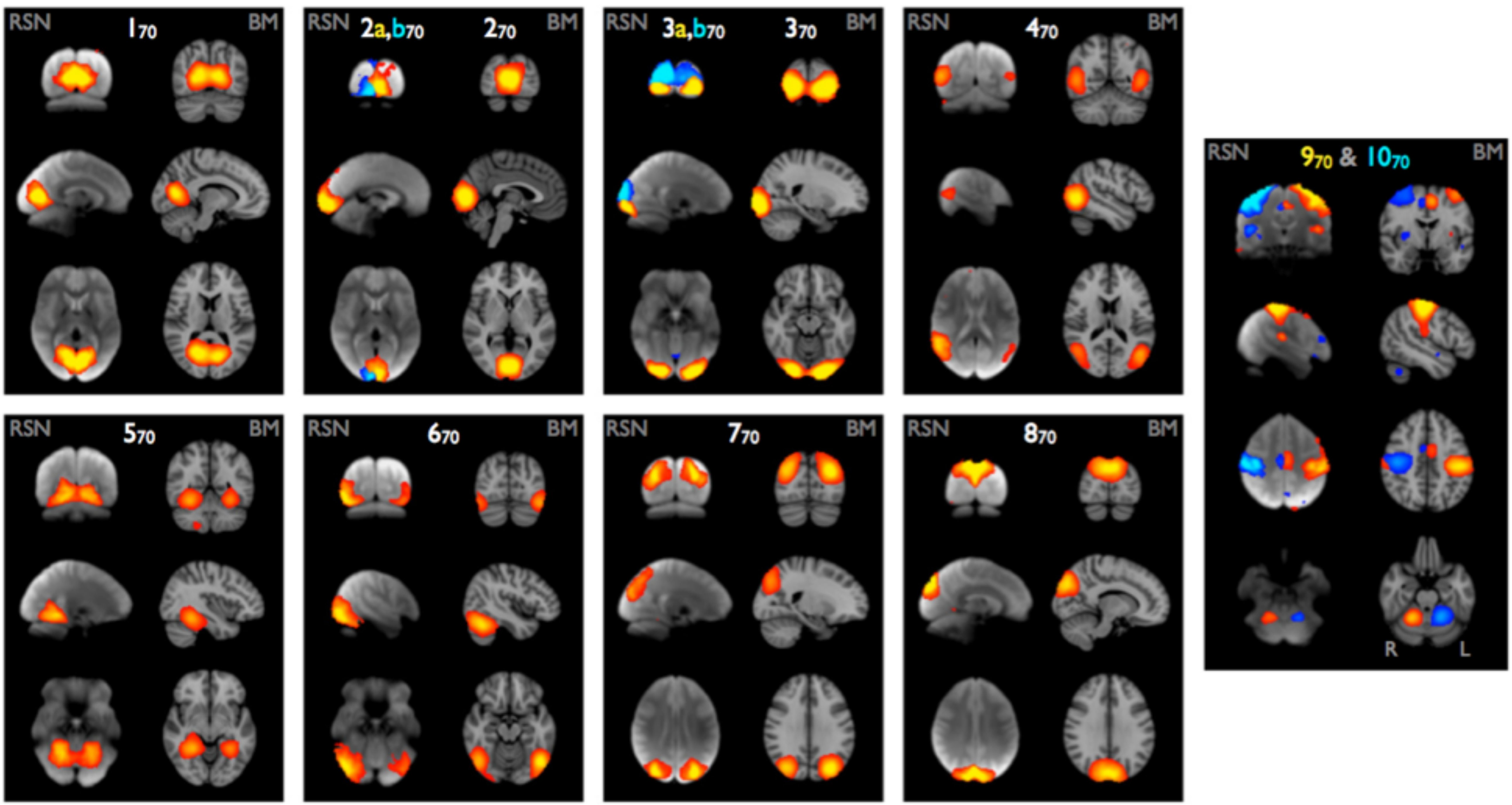


Fig. 3. Eight well-matched pairs of networks in visual areas (1–8_{70}), and 2 pairs from the sensorimotor areas (9, 10_{70}), from the 70-component analyses of the BrainMap activation database and the resting FMRI dataset. All Gaussianized ICA maps are thresholded at $Z = 4$ (higher than for the 20-dimensional results for comparability, because the higher-dimensional analysis, by definition, has reduced ICA residuals).

MVPA

- Exploiting multivariate nature of functional MRI data
  - Supervised learning in task paradigms
    - Decoding
    - Similarity measures
  - Unsupervised learning in resting state

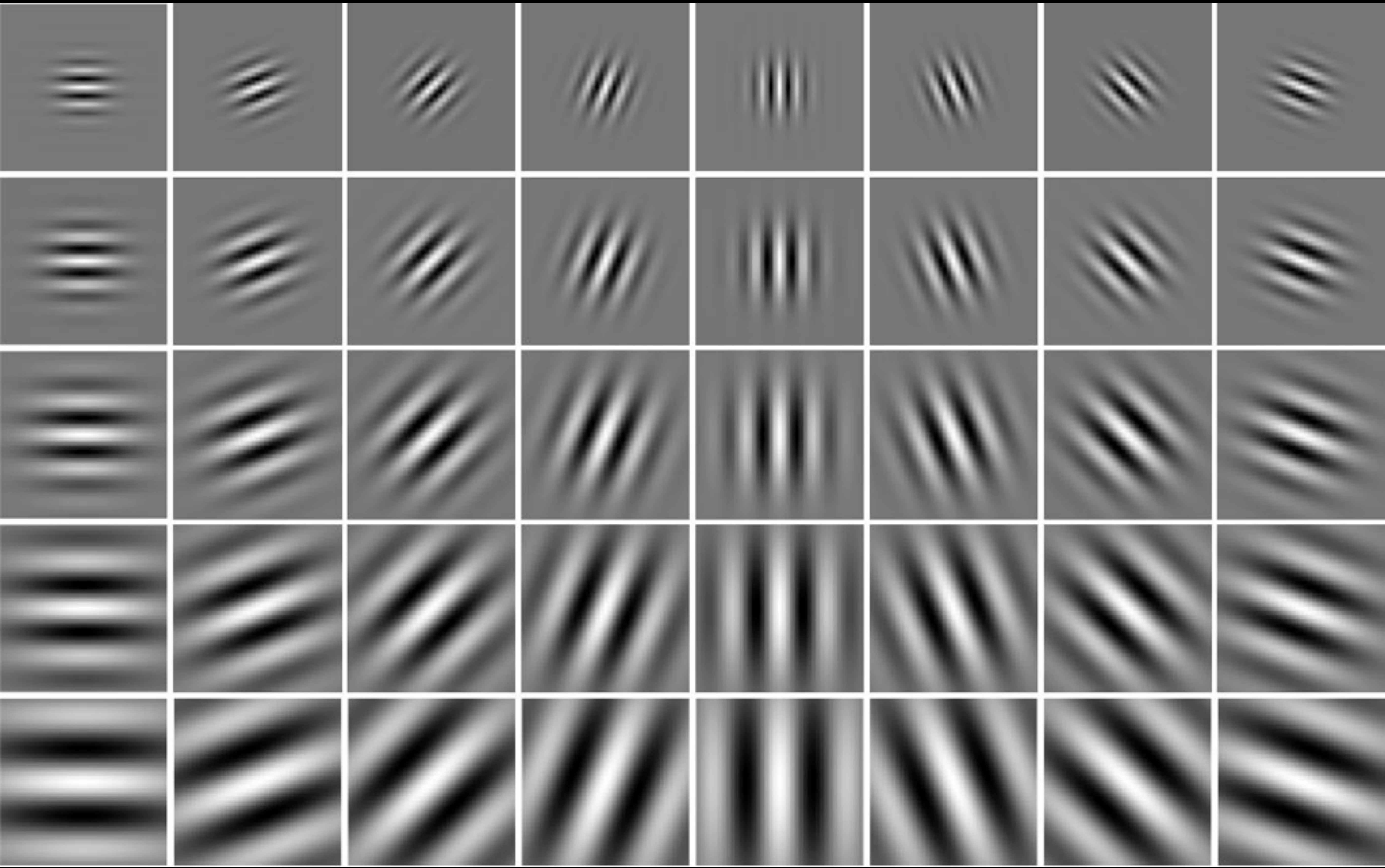
MVPA

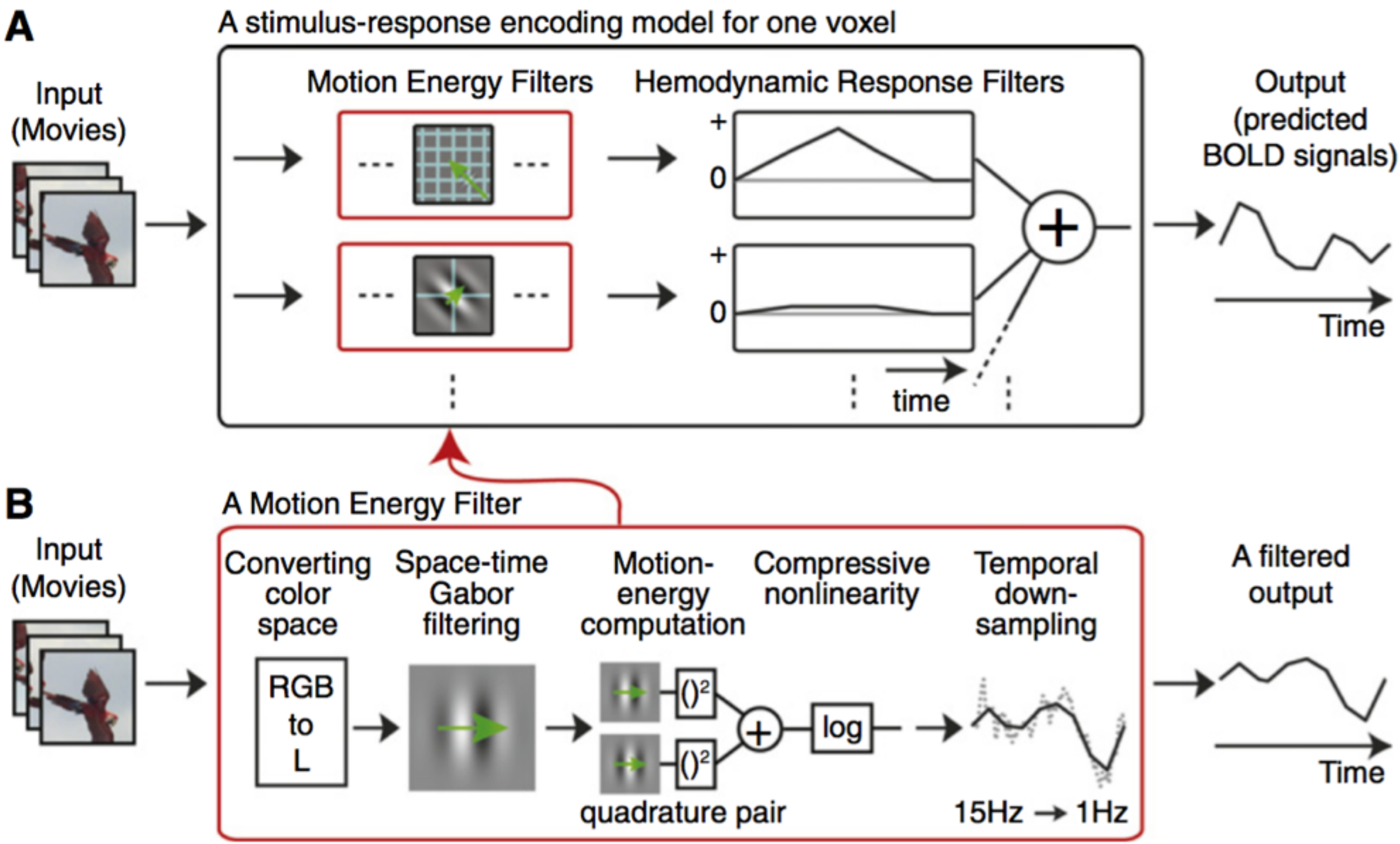
- MVPA can find effects that univariate analysis cannot
    - Neuroscientists love this (note that fMRI dataset ~= 15,000 - 50,000 euros)
- Kriegeskorte: MVPA can learn us something about how the brain represent information

Generative models

- Lately: Bayesian approach to modelling of brain data
  - Generative model stimulus -> brain activity
    - P(brain_activity | stimulus)
  - With Bayes' rule we can now infer the stimulus using the brain data!
    - P(stimulus | brain_activity)

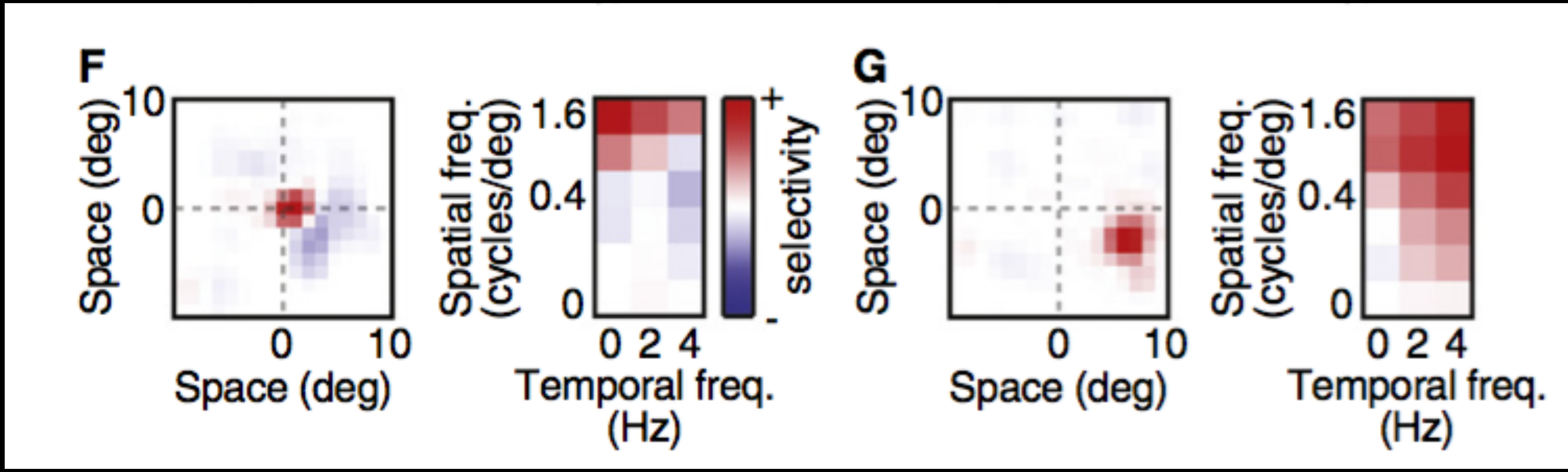
Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies

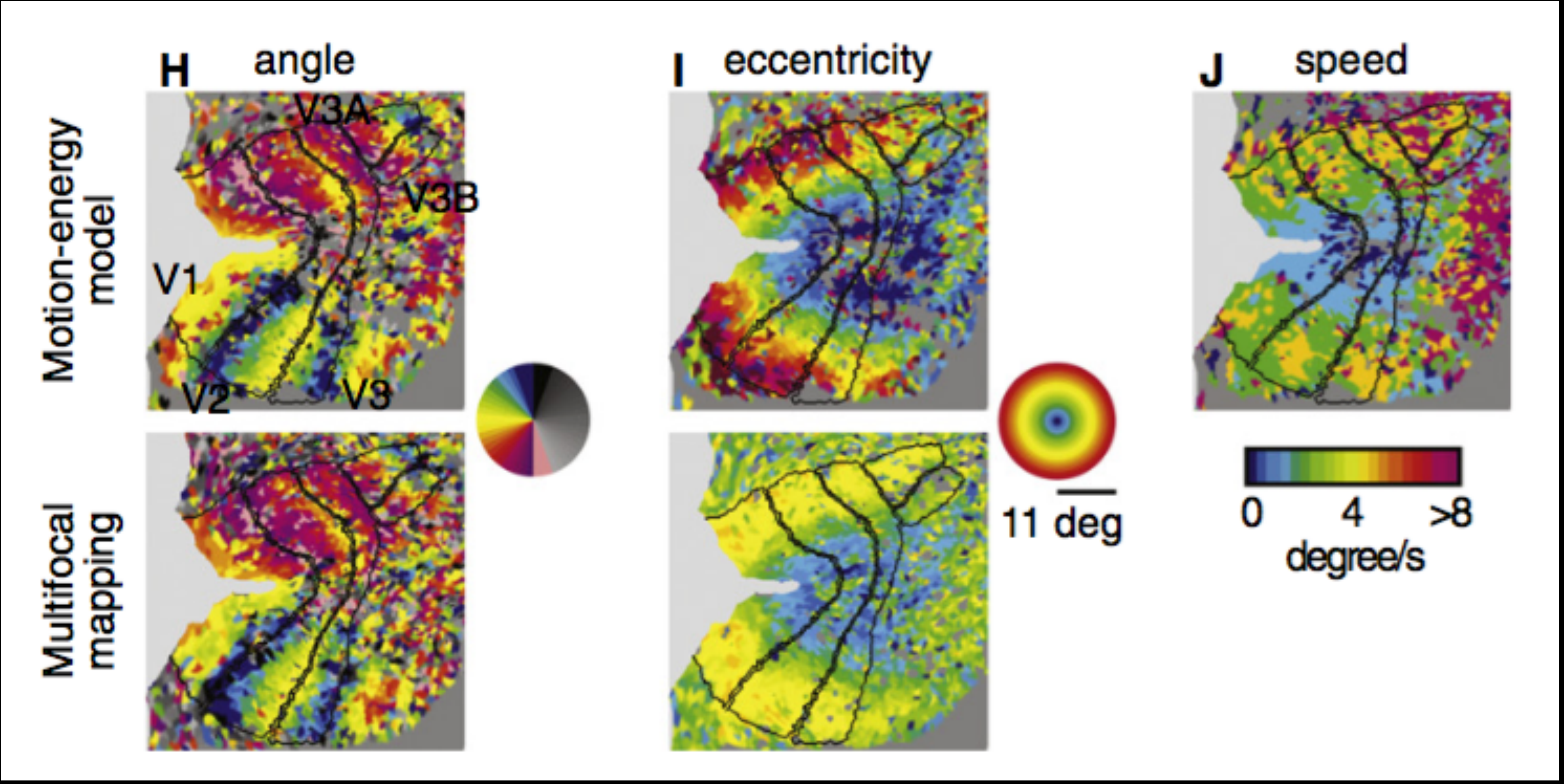


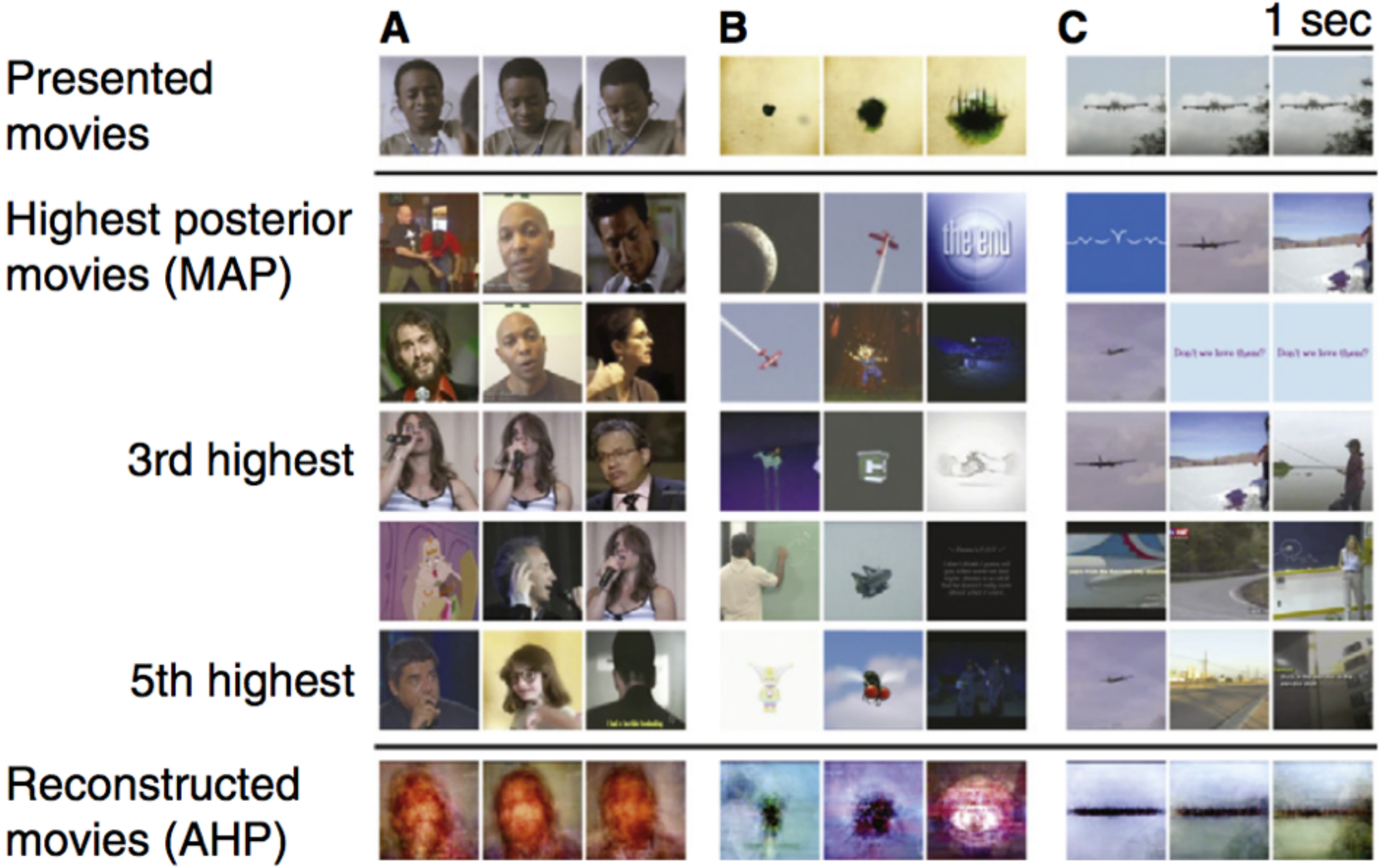

A
A stimulus-response encoding model for one voxel
Input (Movies)
Motion Energy Filters
Hemodynamic Response Filters
+
0
time
Output (predicted BOLD signals)
Time
B
A Motion Energy Filter
Input (Movies)
Converting color space
Space-time Gabor filtering
Motion-energy computation
Compressive nonlinearity
Temporal down-sampling
RGB to L
()²
log
quadrature pair
15Hz → 1Hz
A filtered output
Time

voxel 1

voxel 2




H
angle
I
eccentricity
J
speed
Motion-energy model
Multifocal mapping
V3A
V3B
V1
V2
V3
11 deg
0
4
>8
degree/s


A
B
C
1 sec
Presented movies
Highest posterior movies (MAP)
3rd highest
5th highest
Reconstructed movies (AHP)
the end

Reconstructing visual experiences from brain activity evoked by natural movies

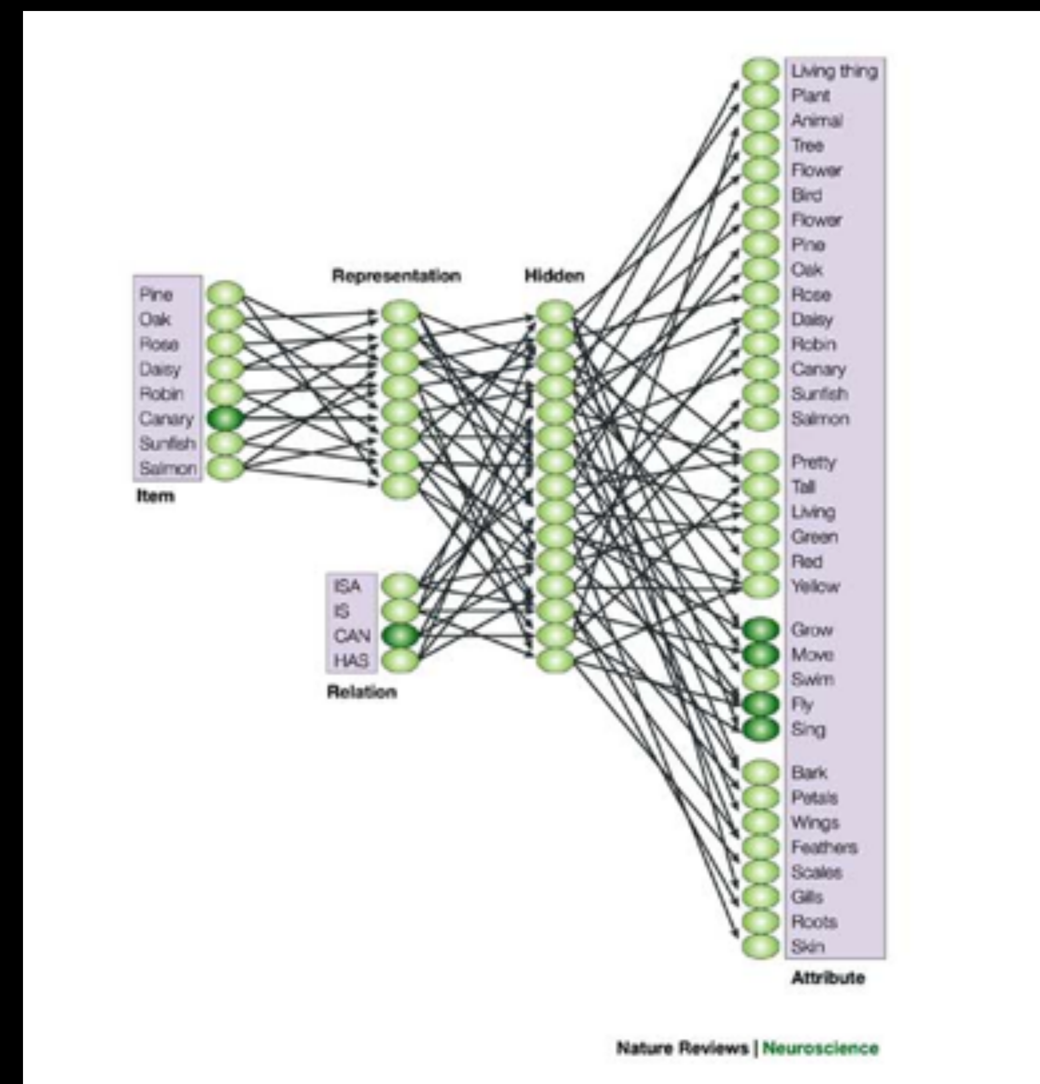
Shinji Nishimoto, An T. Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, Jack L. Gallant

Supplemental movie S1

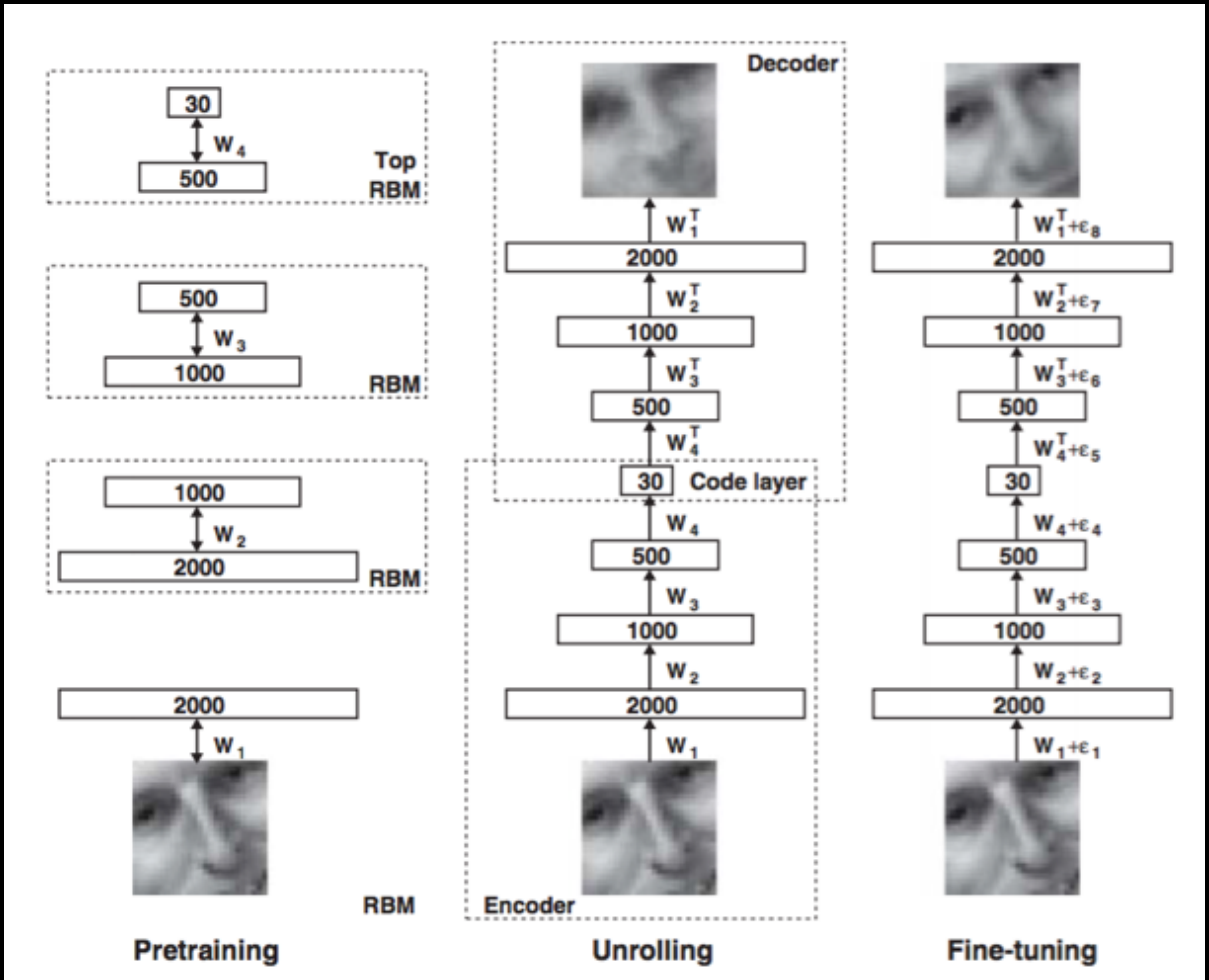
Deep Learning

Deep Learning

- Artificial Neural Networks
    - Hype in the 80s
    - Somewhere in-between supervised and unsupervised learning
- Interest declined in 90s and 00s
- 2006: some crucial papers by Geoffrey Hinton’s group in Toronto
- 2016: *Big idea* of artificial intelligence. Lots of media and business attention



Deep Learning



Deep Learning

Imagenet: 1.2 miljoen plaatjes met 1000 klasse

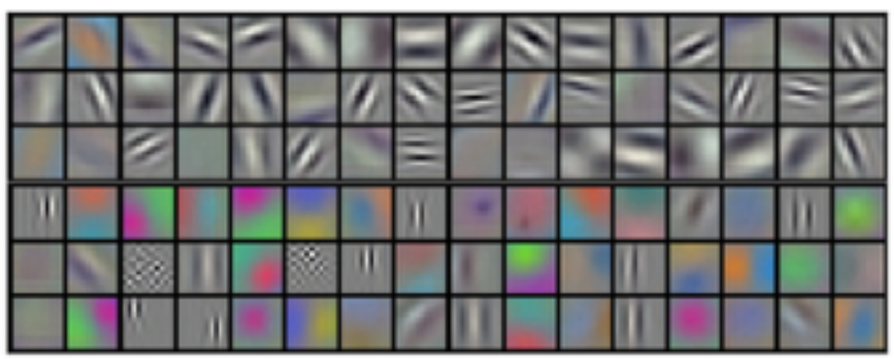
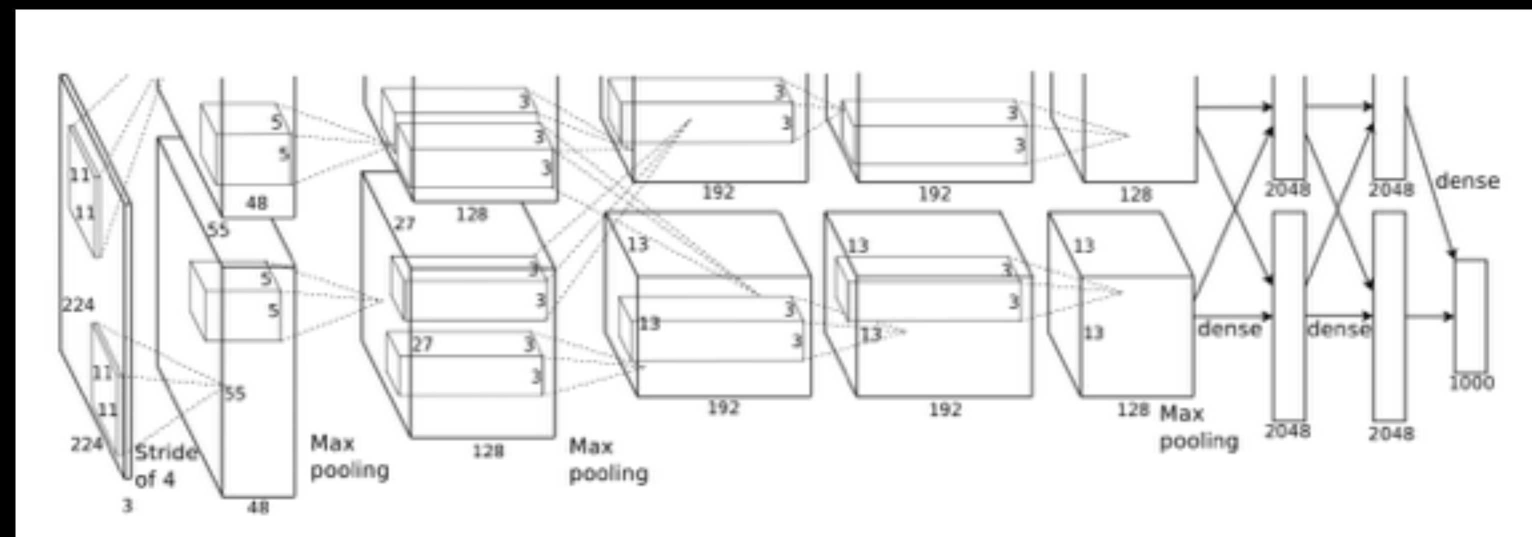
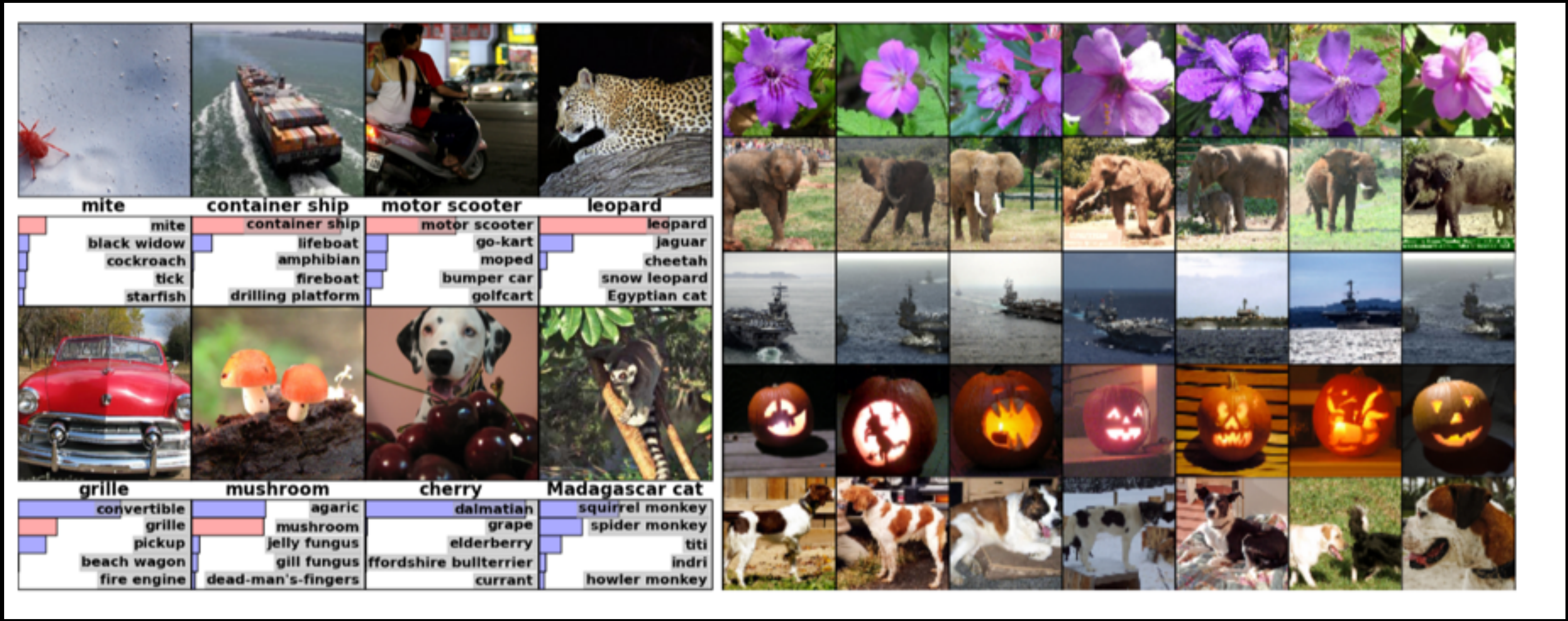


Figure 3: 96 convolutional kernels of size 11×11×3 learned by the first convolutional layer on the 224×224×3 input images. The top 48 kernels were learned on GPU 1 while the bottom 48 kernels were learned on GPU 2. See Section 6.1 for details.



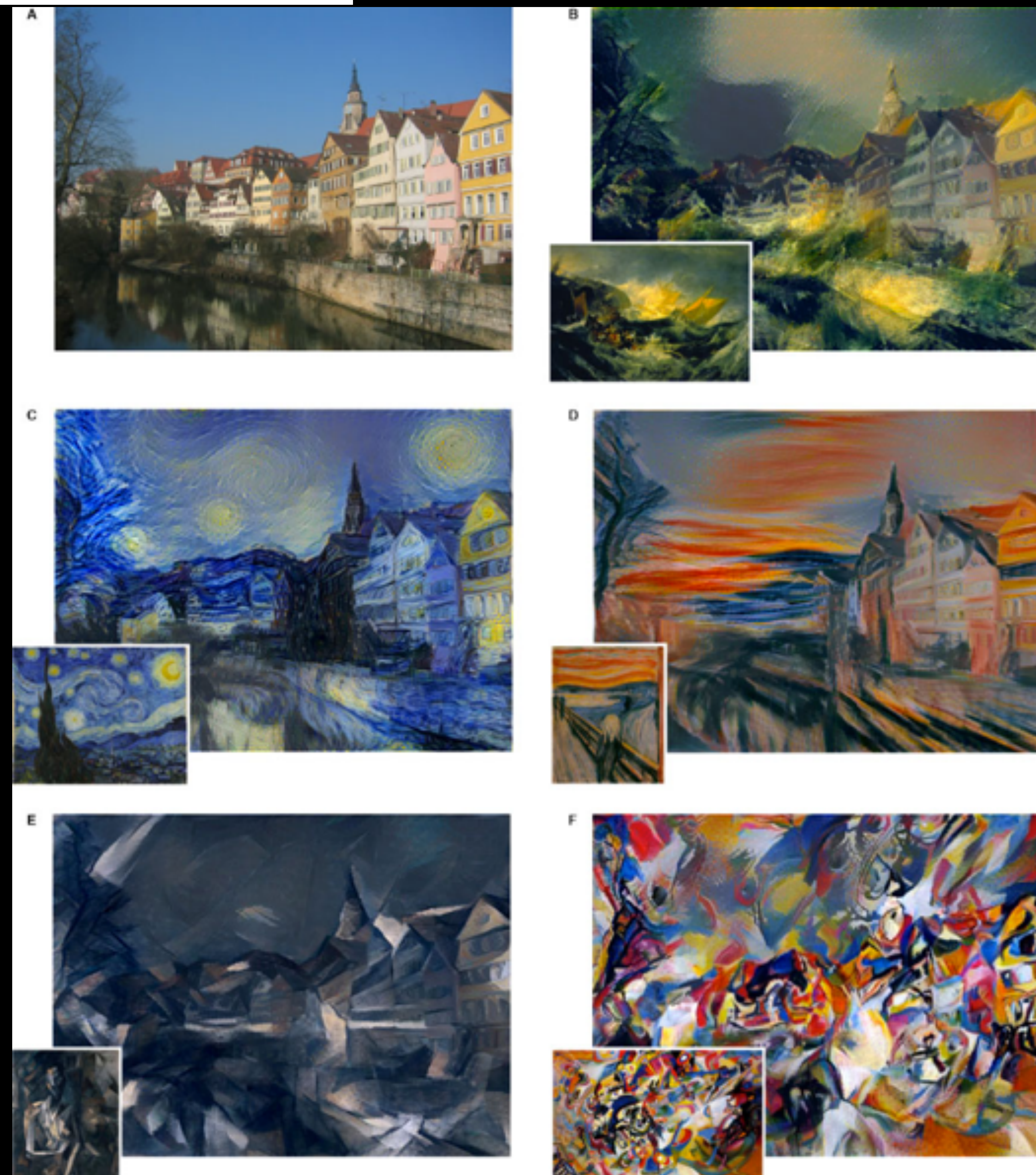
Deep Learning

Imagenet: 1.2 miljoen plaatjes met 1000 klasse
17% correct



A Neural Algorithm of Artistic Style

Leon A. Gatys,[1,2,3*] Alexander S. Ecker,[1,2,4,5] Matthias Bethge[1,2,4]

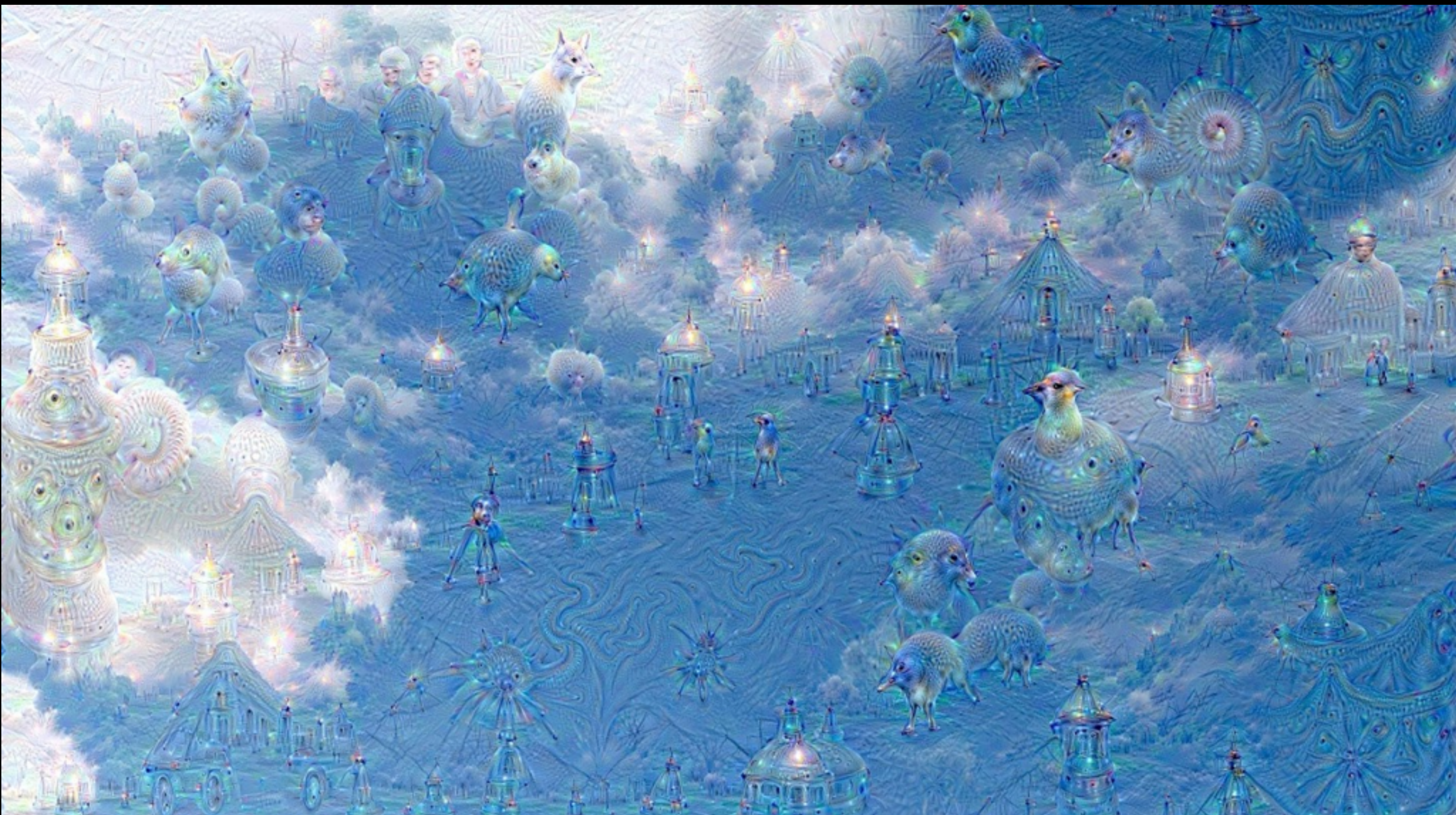


Inceptionism: Going deeper into Neural Networks

12 dec. 2008–18 jun. 2015

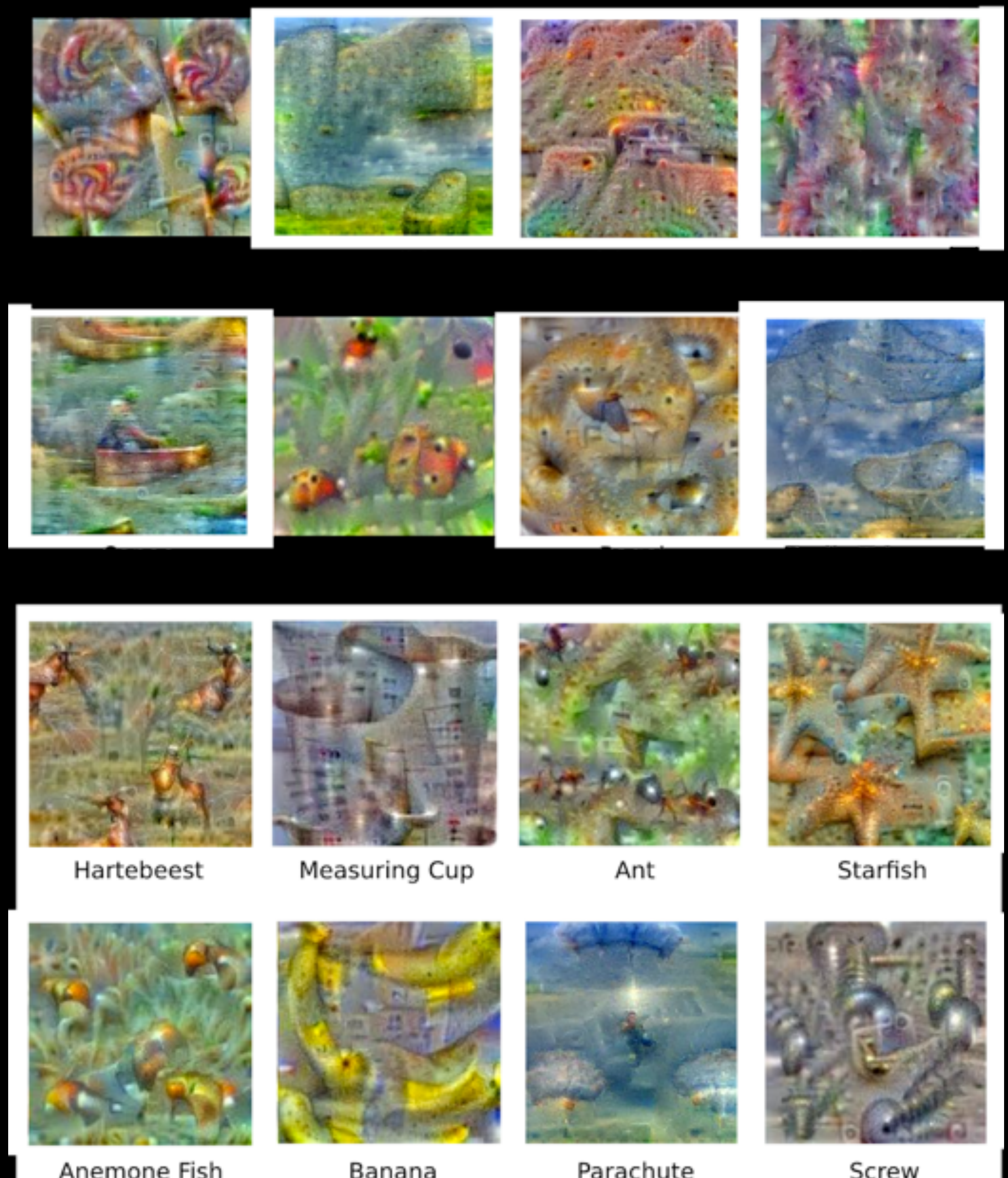
Inceptionism: Going deeper into Neural Networks

12 dec. 2008–18 jun. 2015



Inceptionism: Going deeper into Neural Networks

12 dec. 2008–18 jun. 2015



Inceptionism: Going deeper into Neural Networks

12 dec. 2008–18 jun. 2015

Show and Tell: A Neural Image Caption Generator

Oriol Vinyals
Google

Alexander Toshev
Google

Samy Bengio
Google

Dumitru Erhan
Google

LETTER

doi:10.1038/nature14236

Human-level control through deep reinforcement learning

Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]

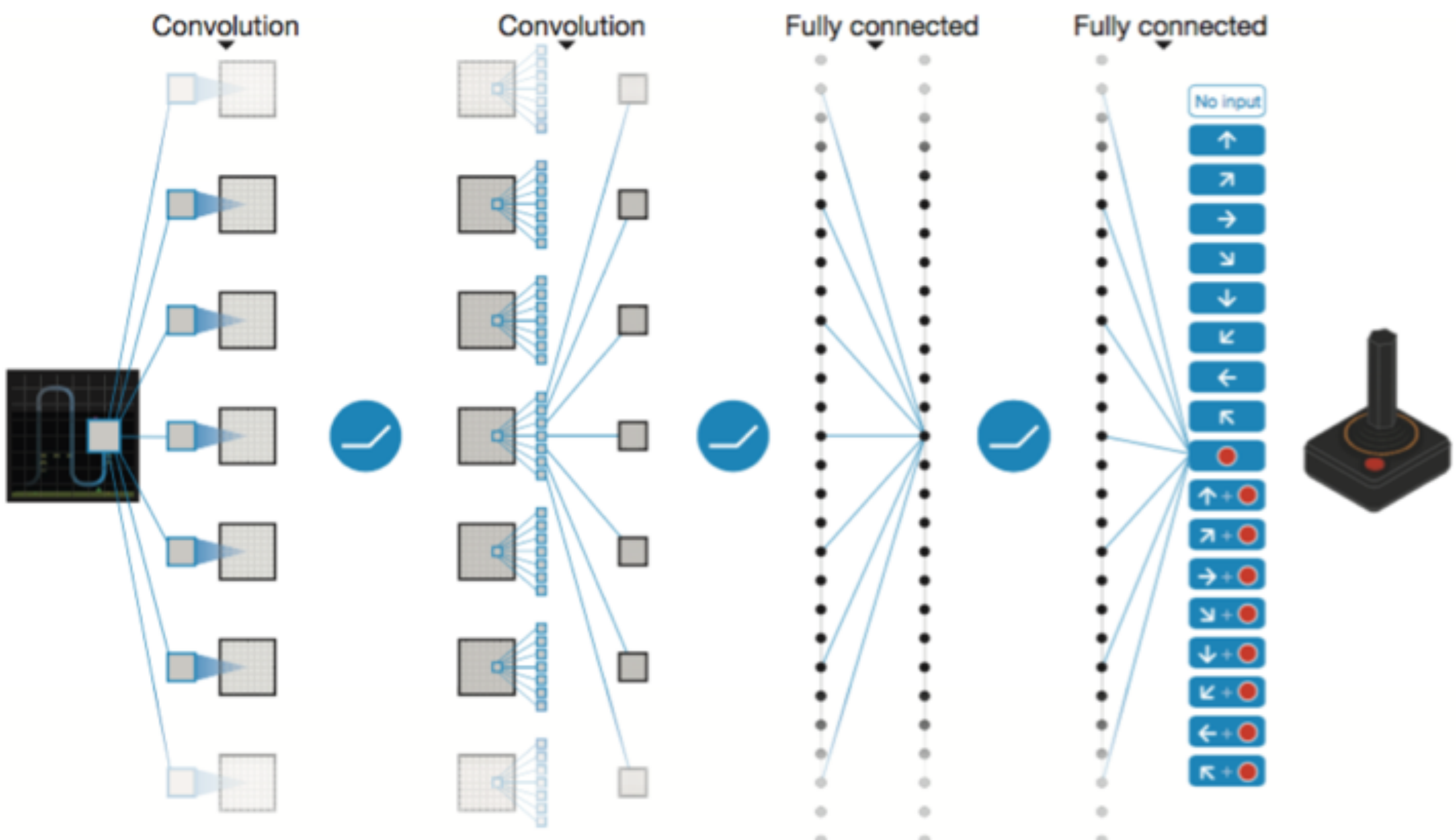


Figure 1 | Schematic illustration of the convolutional neural network. The details of the architecture are explained in the Methods. The input to the neural network consists of an $84 \times 84 \times 4$ image produced by the preprocessing map ϕ, followed by three convolutional layers (note: snaking blue line symbolizes sliding of each filter across input image) and two fully connected layers with a single output for each valid action. Each hidden layer is followed by a rectifier nonlinearity (that is, $\max(0,x)$).

Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream

Umut Güçlü and Marcel A. J. van Gerven
Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands

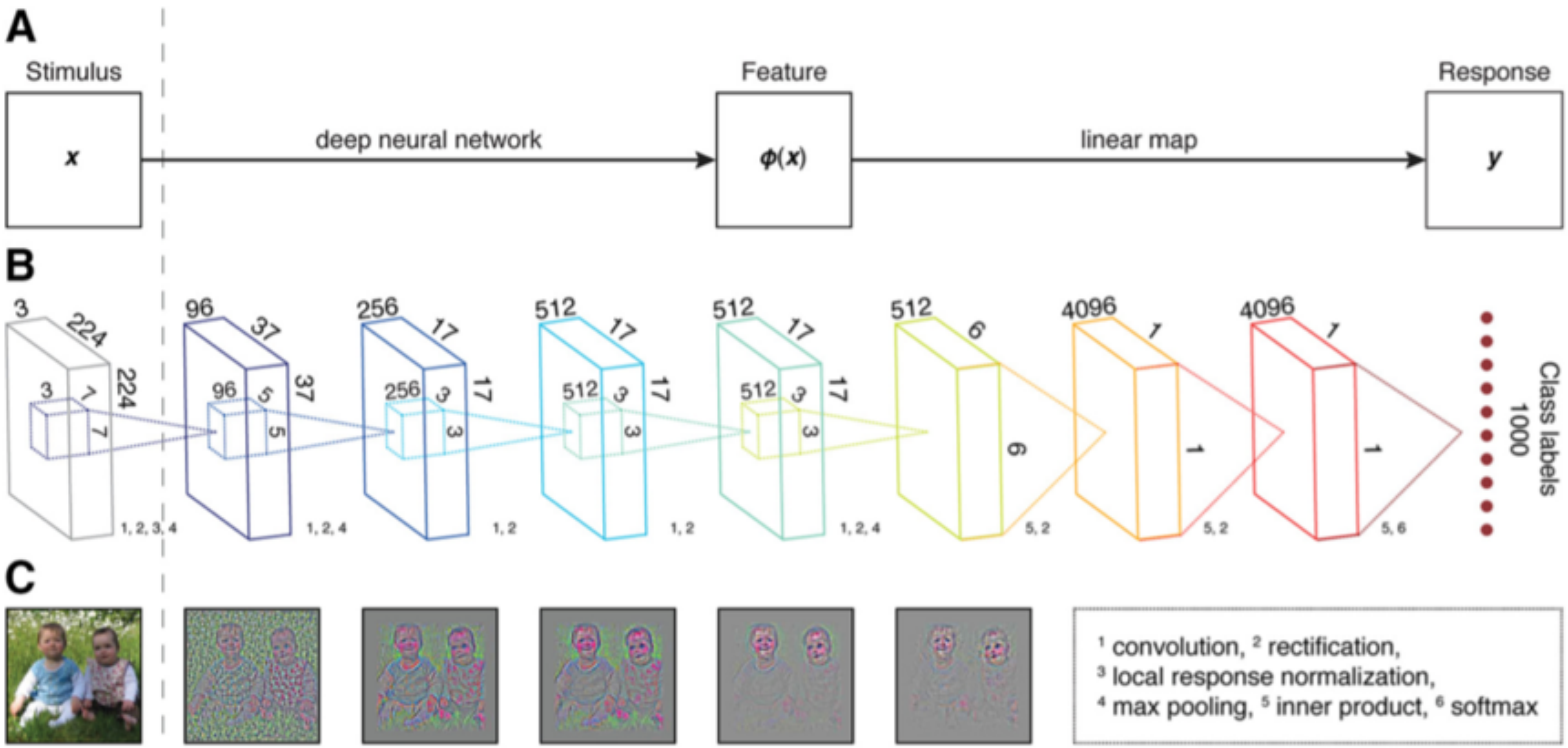
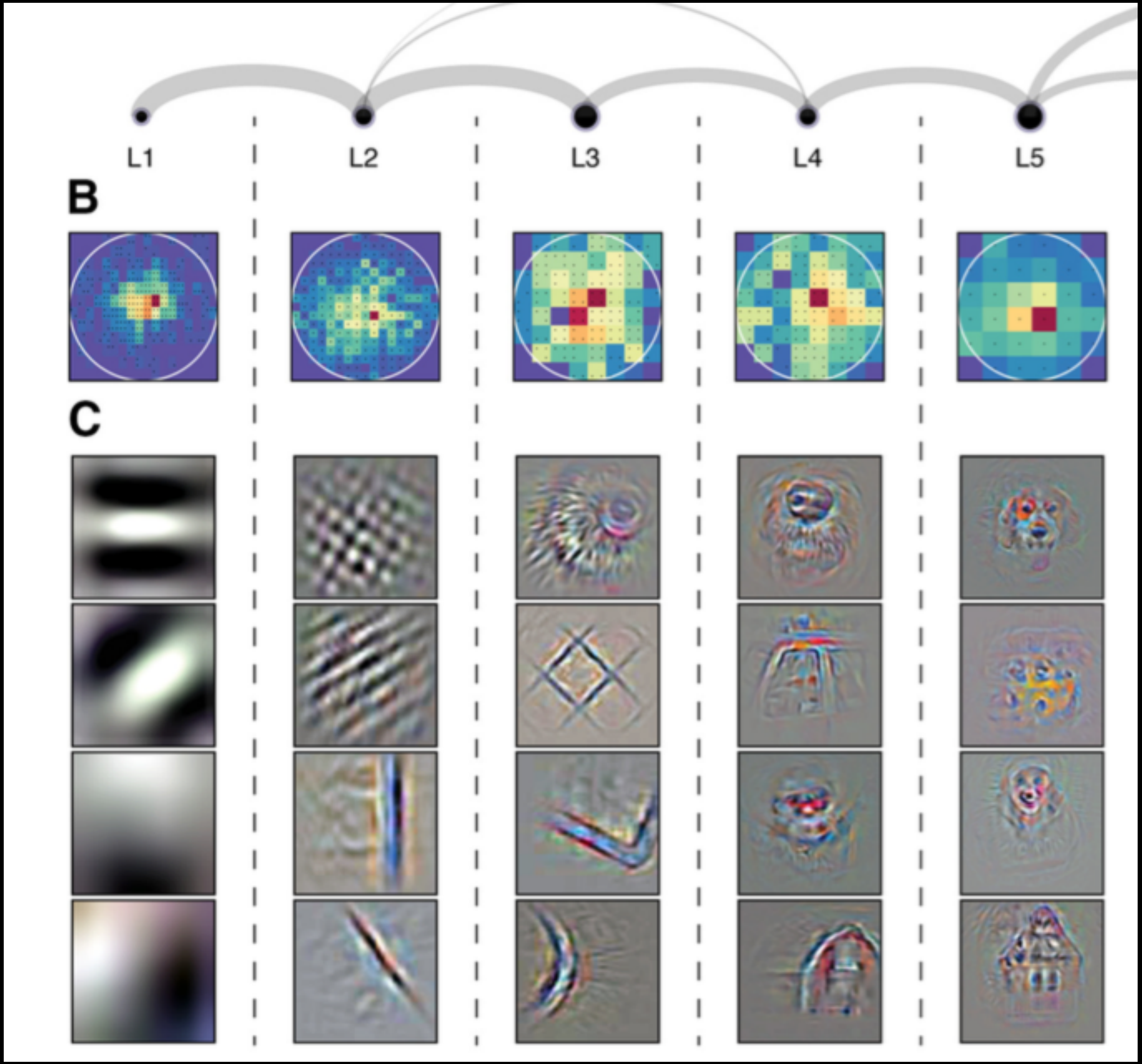
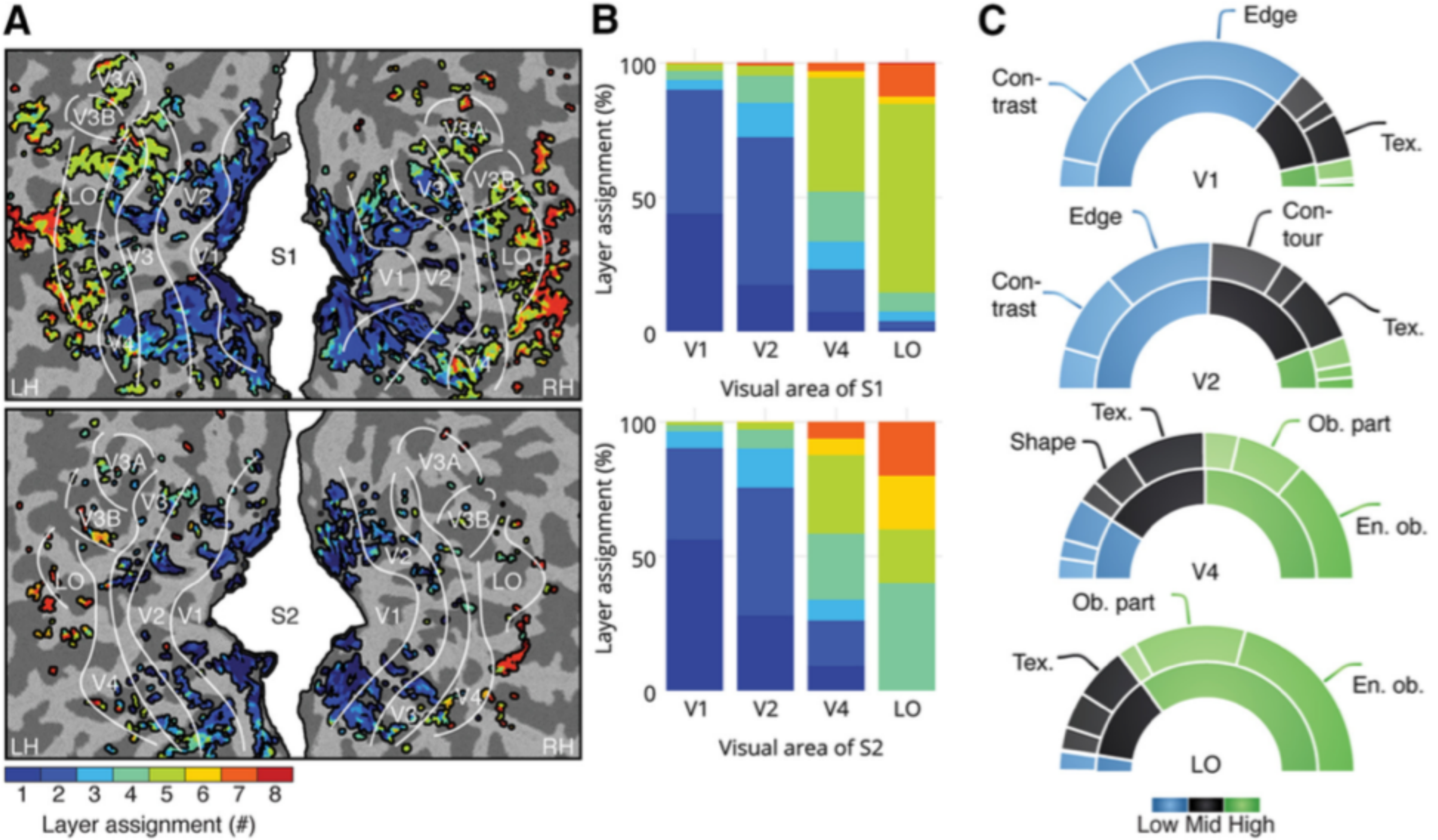


Figure 1. DNN-based encoding framework. ***A***, Schematic of the encoding model that transforms a visual stimulus to a voxel response in two stages. First, a deep (convolutional) neural networ transforms the visual stimulus (***x***) to multiple layers of feature representations. Then, a linear mapping transforms a layer of feature representations to a voxel response (***y***). ***B***, Schematic of the dee neural network where each layer of artificial neurons uses one or more of the following (non)linear transformations: convolution, rectification, local response normalization, max pooling, inne product, and softmax. ***C***, Reconstruction of an example image from the activities in the first five layers.


L1
L2
L3
L4
L5
B
C


A
B
C
V3A
V3B
LO
V3
V2
V1
V4
S1
S2
LH
RH
1 2 3 4 5 6 7 8
Layer assignment (#)
Layer assignment (%)
100
50
0
V1 V2 V4 LO
Visual area of S1
Visual area of S2
Edge
Con-
trast
Tex.
Con-
tour
Shape
Ob. part
En. ob.
Low Mid High

Anti-deep learners

RESEARCH ARTICLES

COGNITIVE SCIENCE

Human-level concept learning through probabilistic program induction

Brenden M. Lake,[1*] Ruslan Salakhutdinov,[2] Joshua B. Tenenbaum[3]

Category learning

- What is a house? What is a horse? When does a giraffe become a horse?

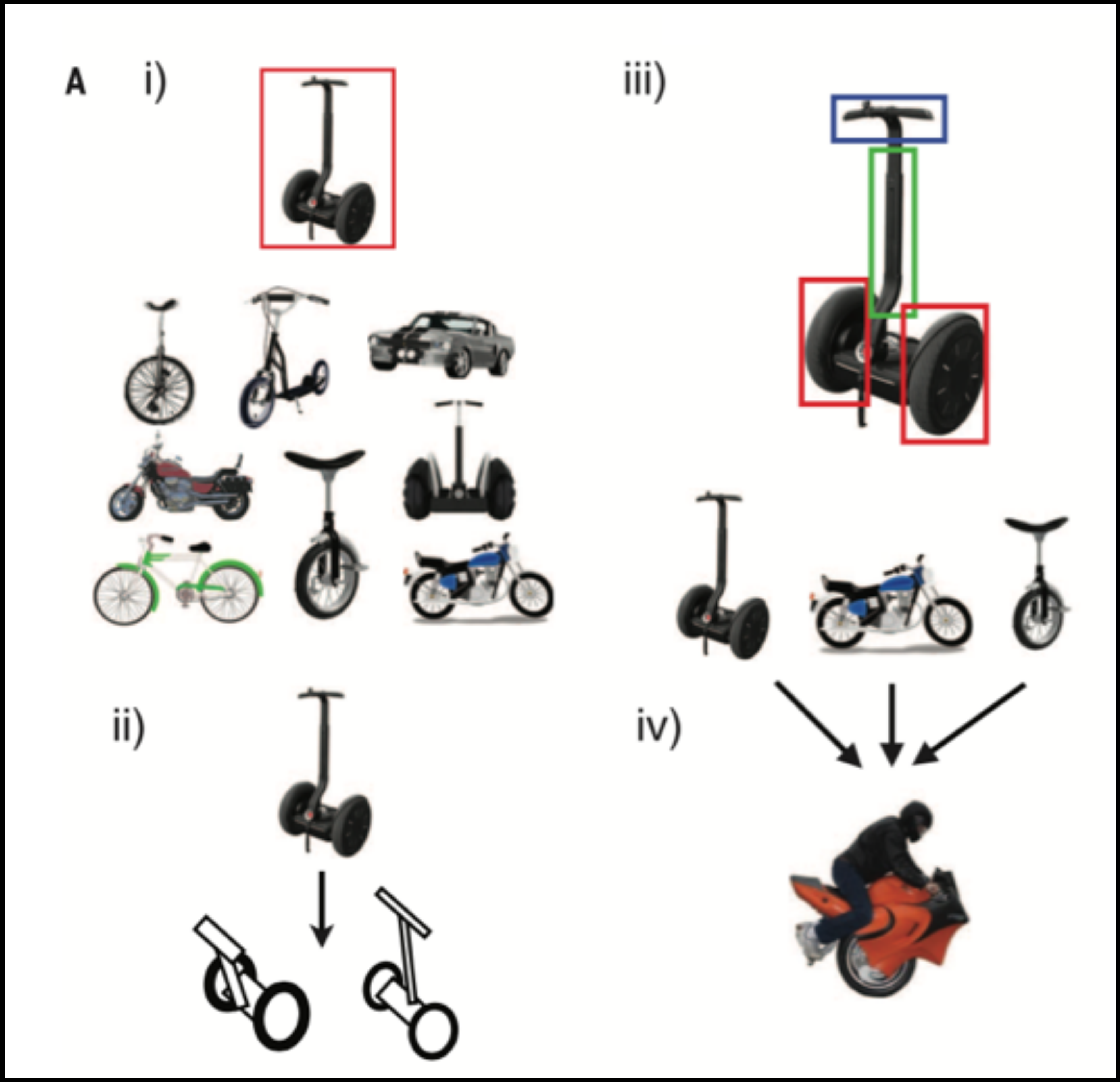
- How many giraffes do you have to show a child before it understands what a giraffe is?

  - *Sparse data*

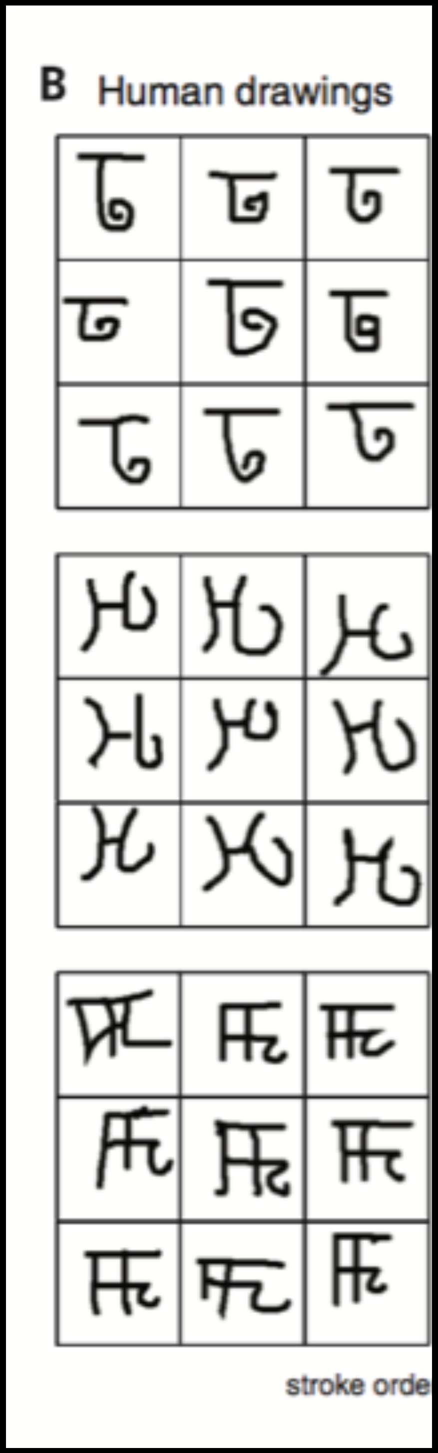
  - *Generalising*

Category learning

- Deep learning needs hundreds of examples to learn categories.
- They don't use *structural priors*

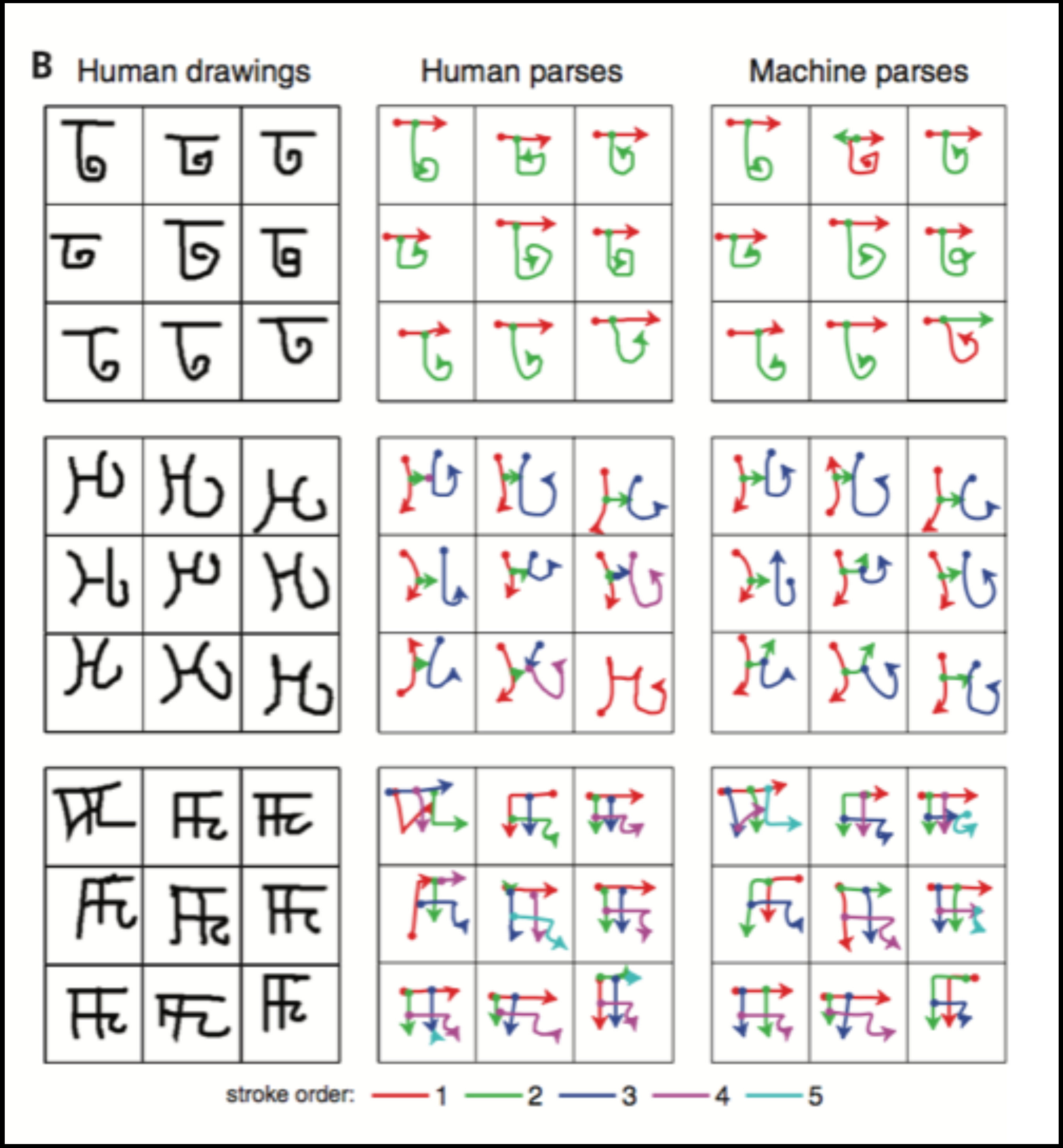


Category learning



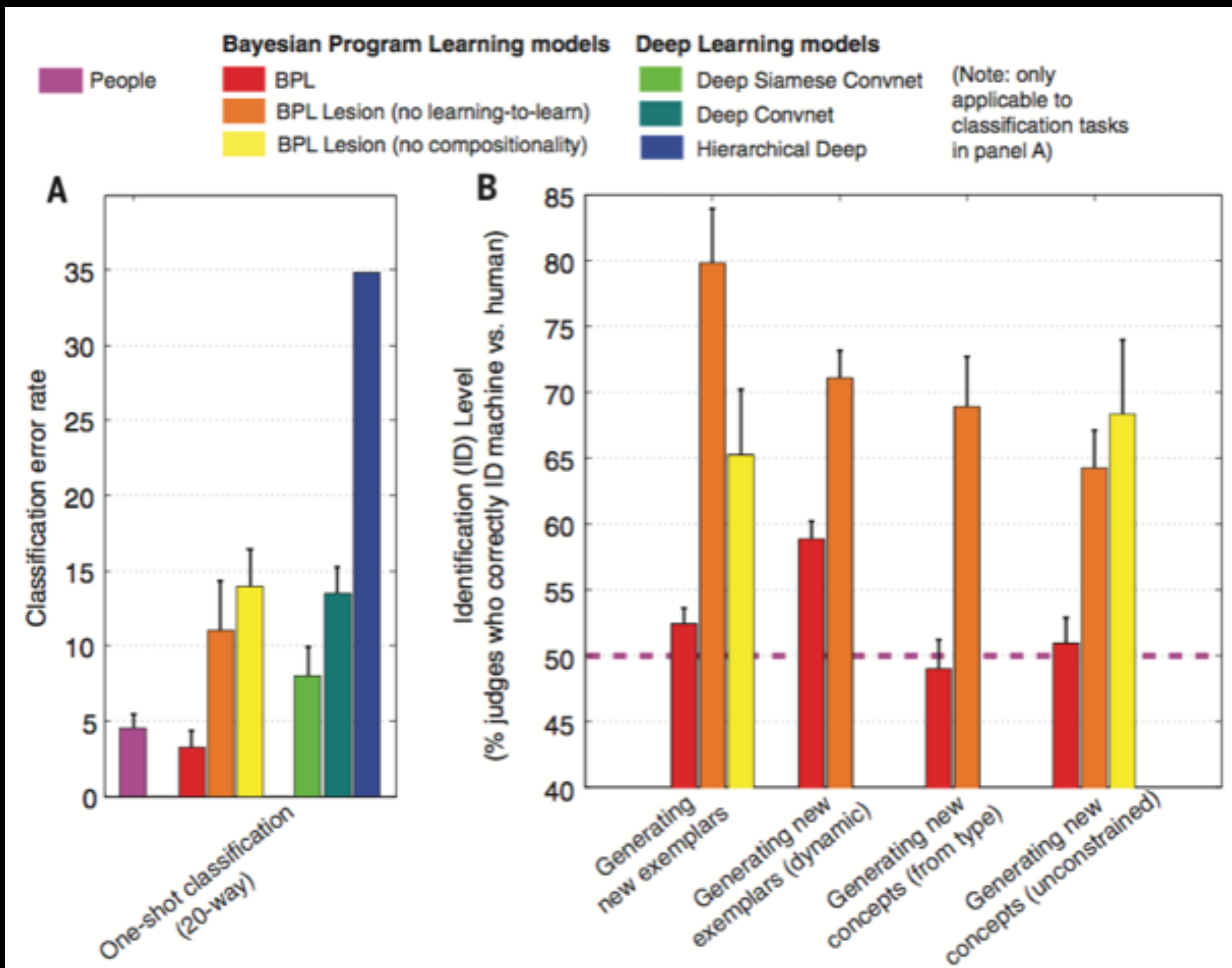
Lake et al. ,2015

Category learning



Lake et al. ,2015

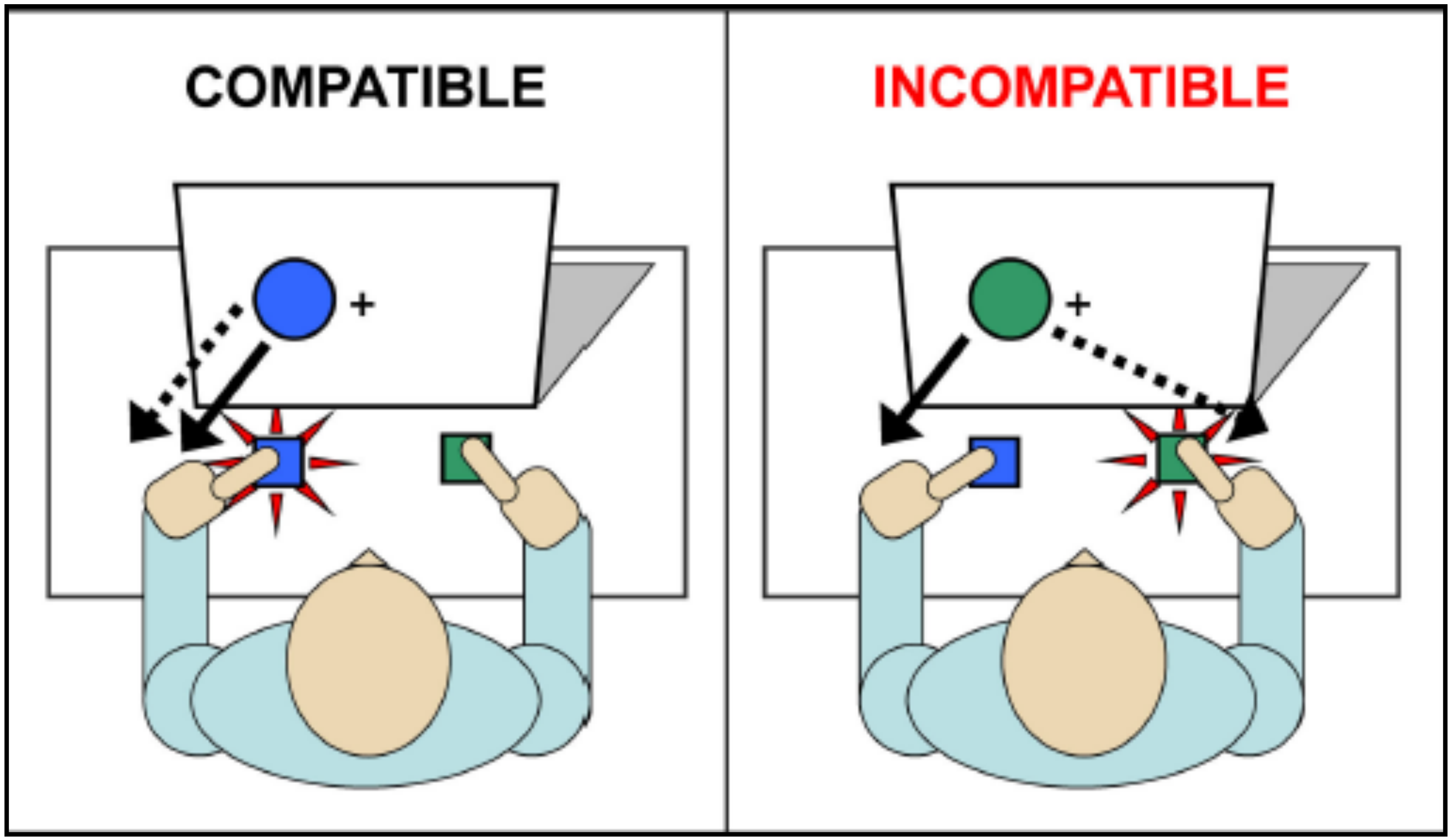
Category learning

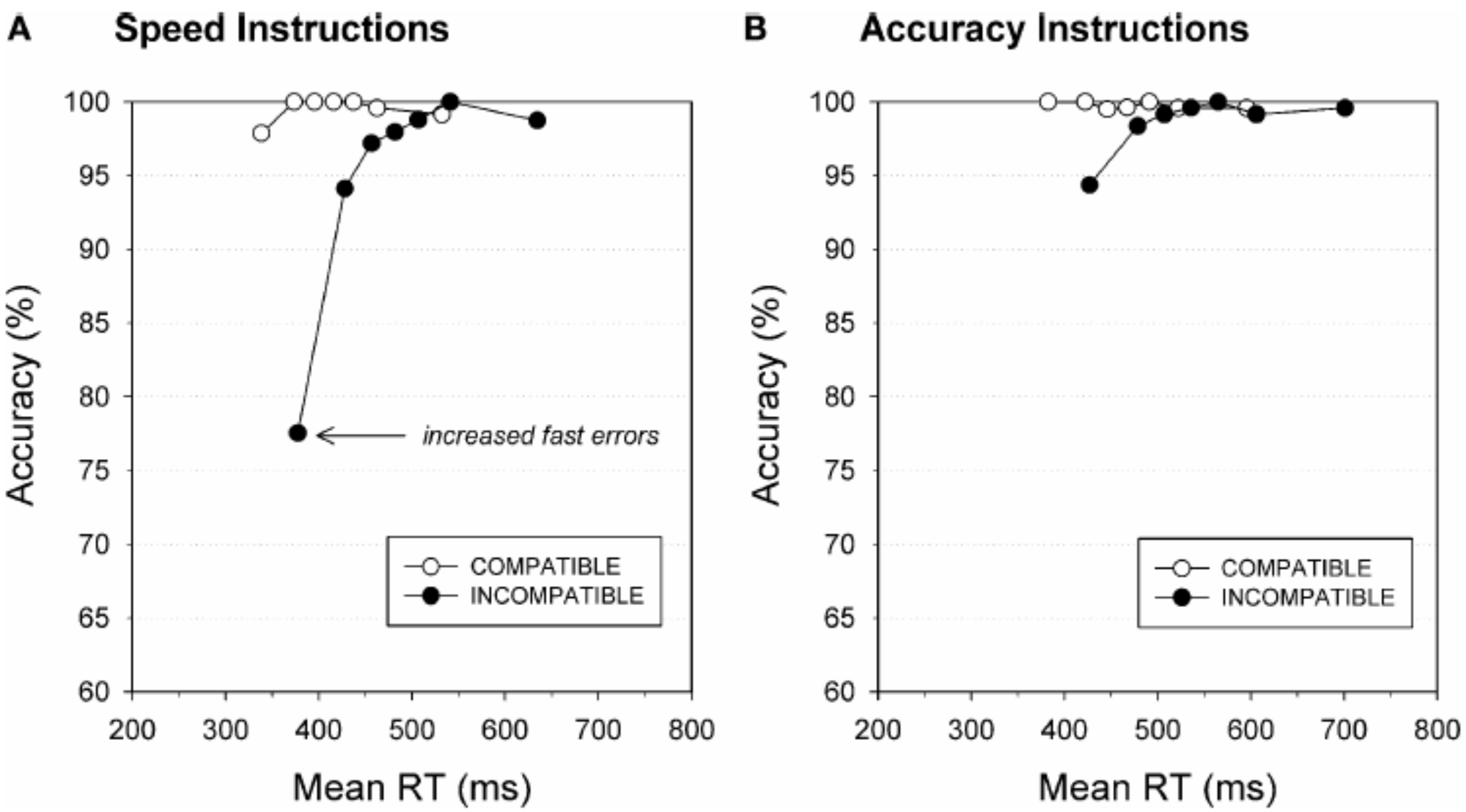


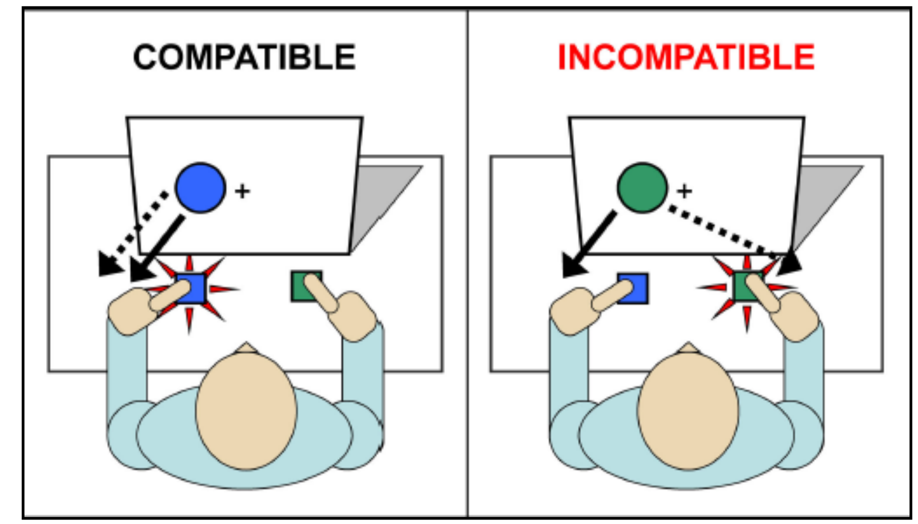
Lake et al. ,2015

Simon task

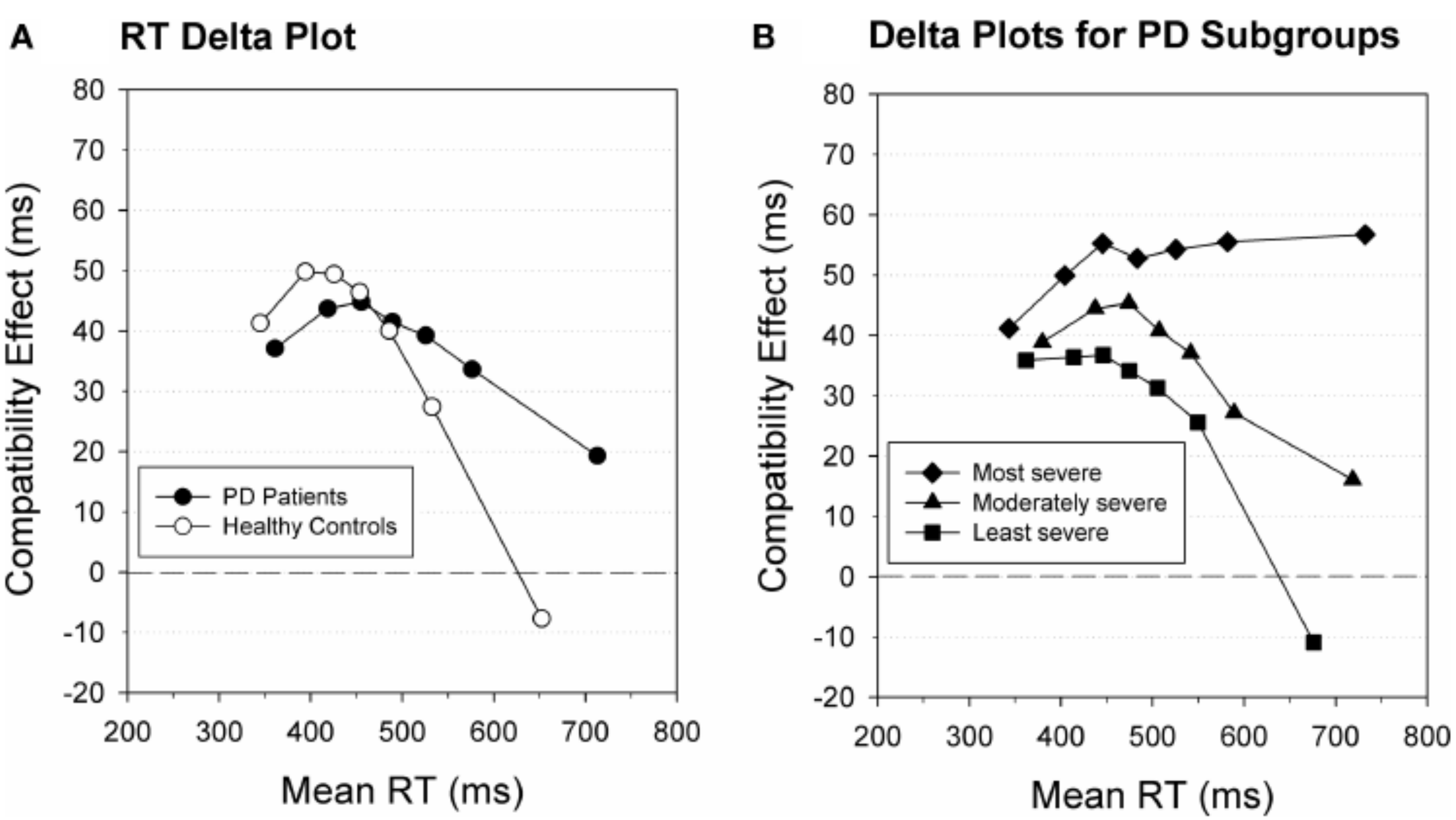
Simon task taps on the executive control function that **inhibits irrelevant response**





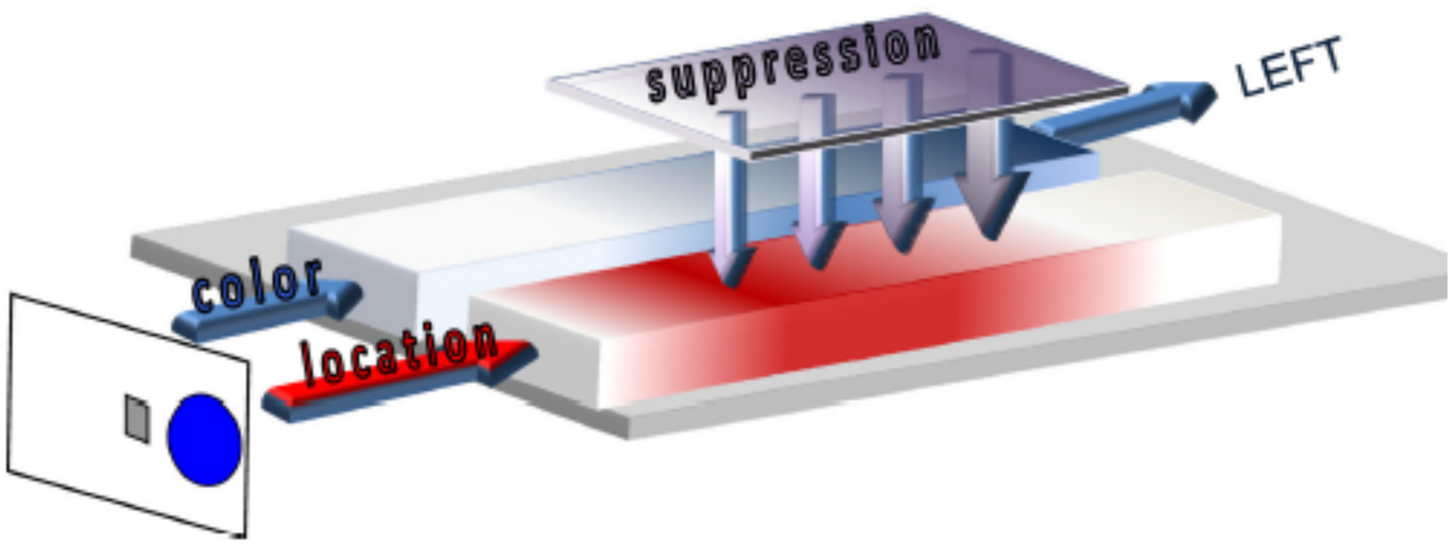


(Wylie, et al., 2009)



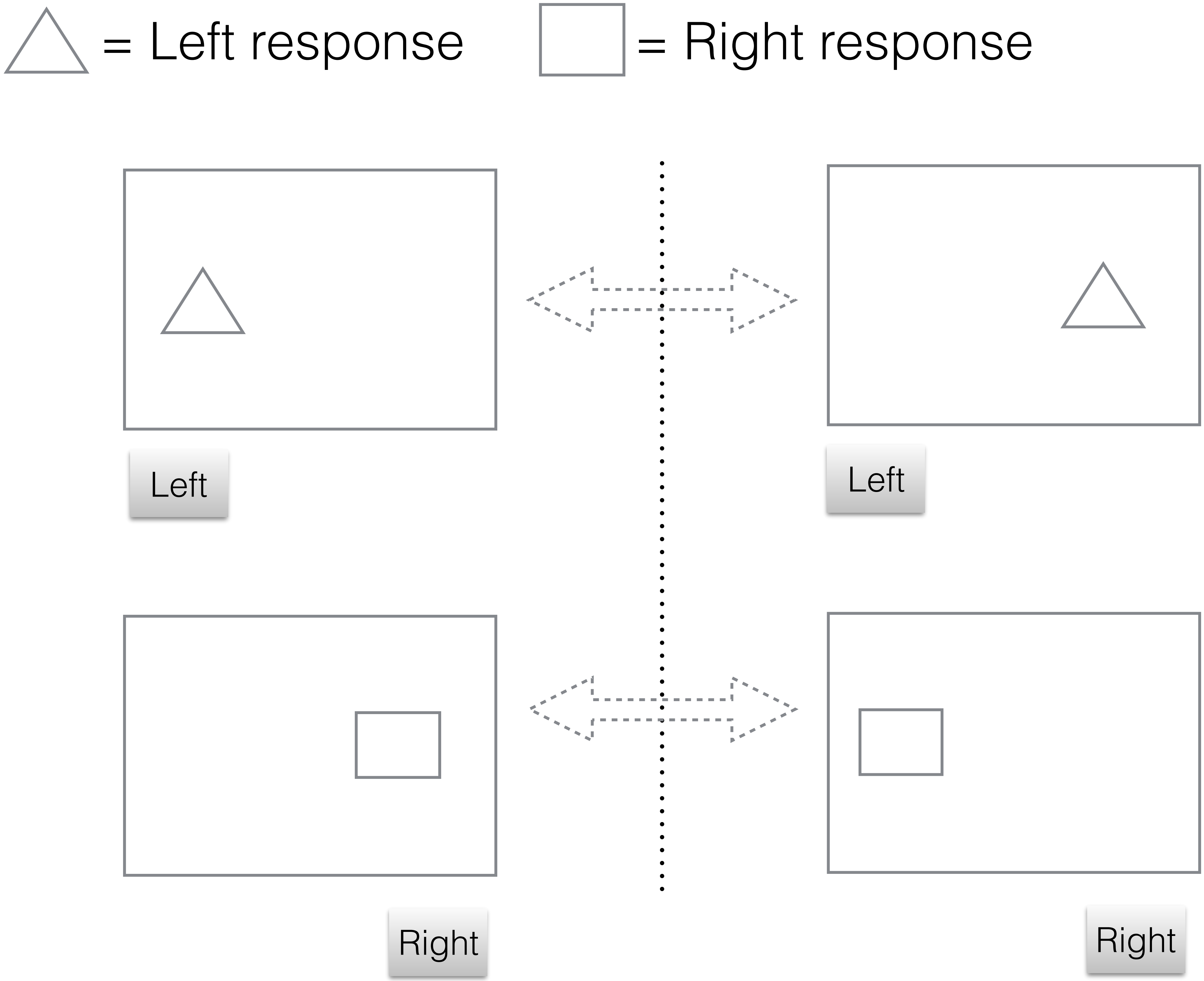
(Wylie, et al., 2010)

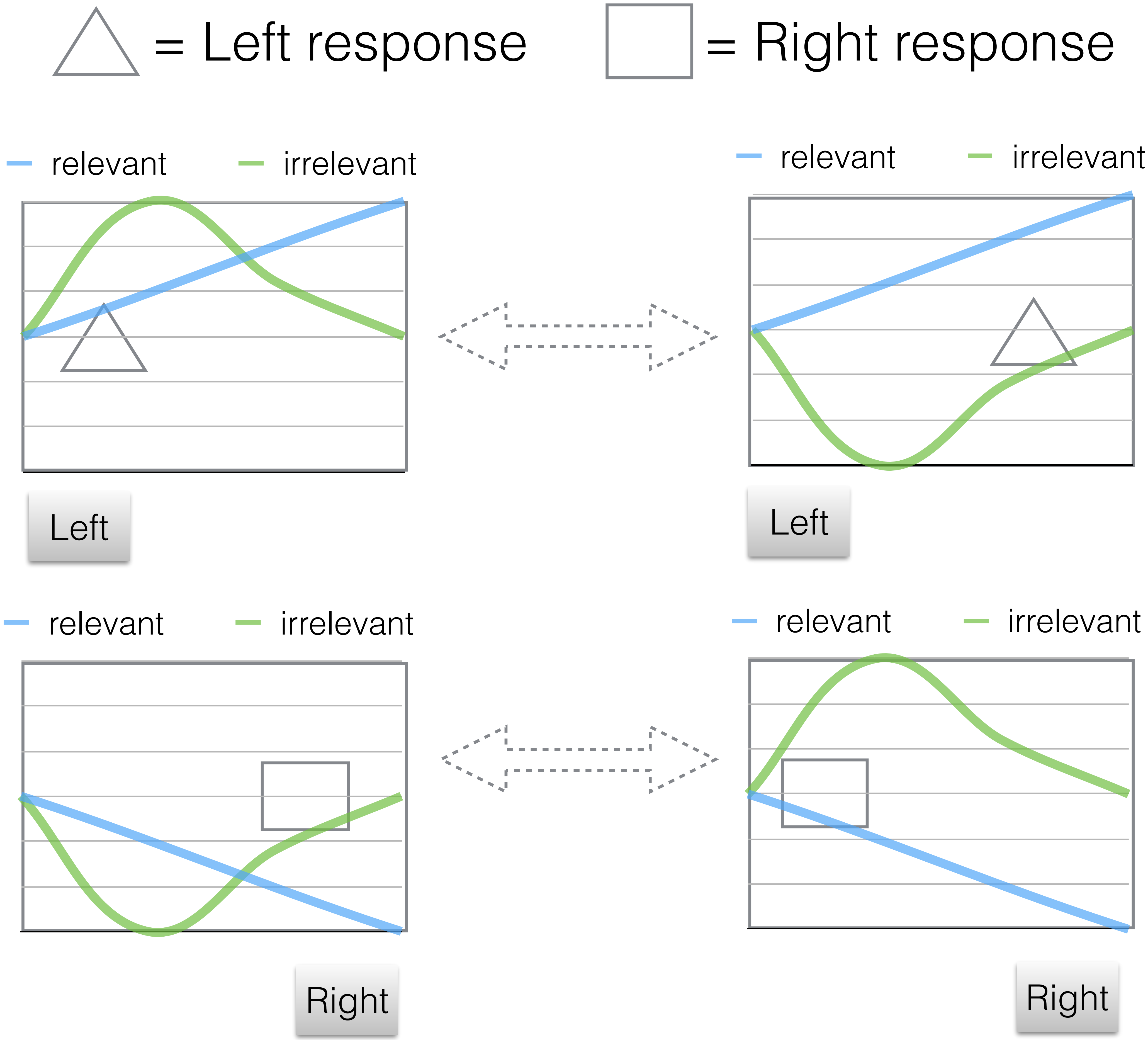
1. Activation-suppression model

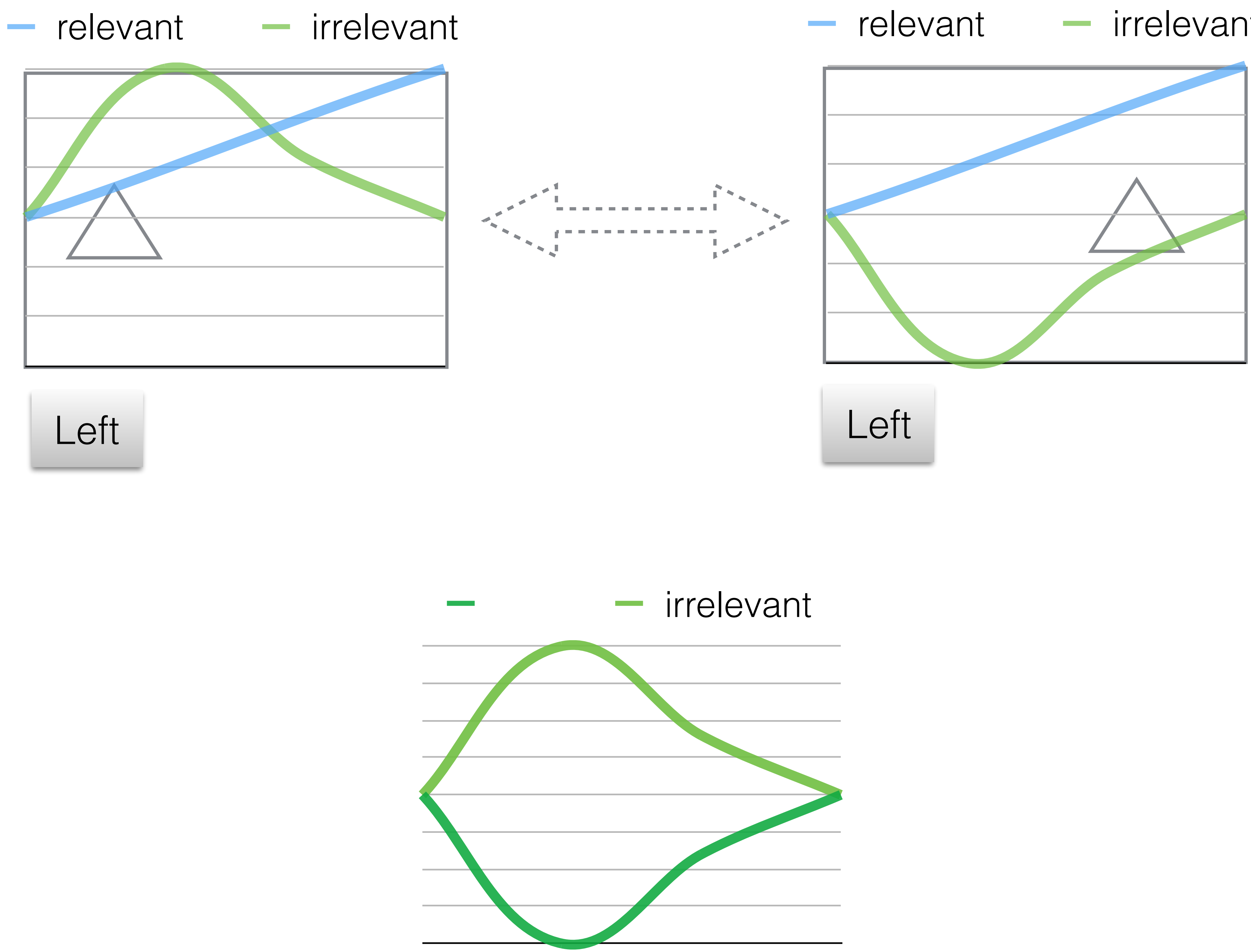


(Van Den Wildenberg, et al., 2010; De Jong et al., 1994; Ridderinkhof, 2002)

2. Passive decay of the spatial attribute (Hommel, 1994)

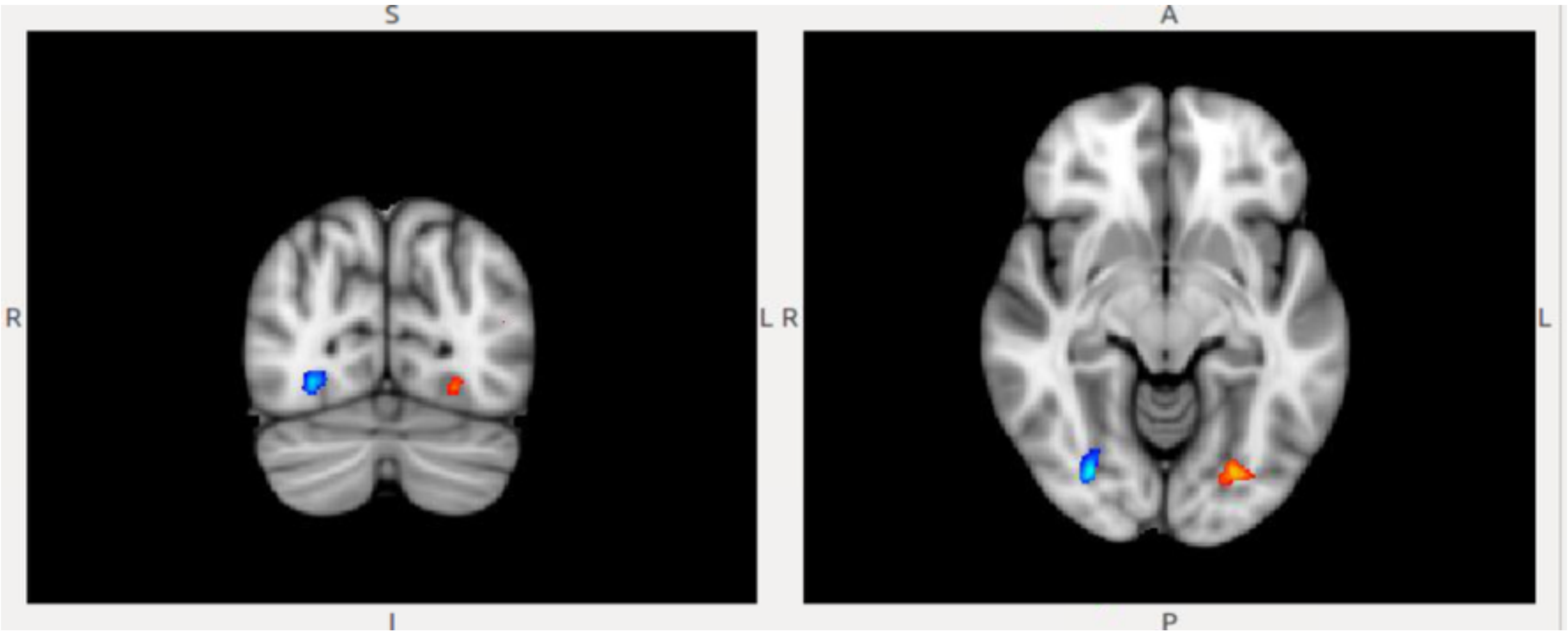

= Left response
= Right response
Left
Left
Right
Right


= Left response
= Right response
relevant
irrelevant
Left
relevant
irrelevant
Left
relevant
irrelevant
Right
relevant
irrelevant
Right



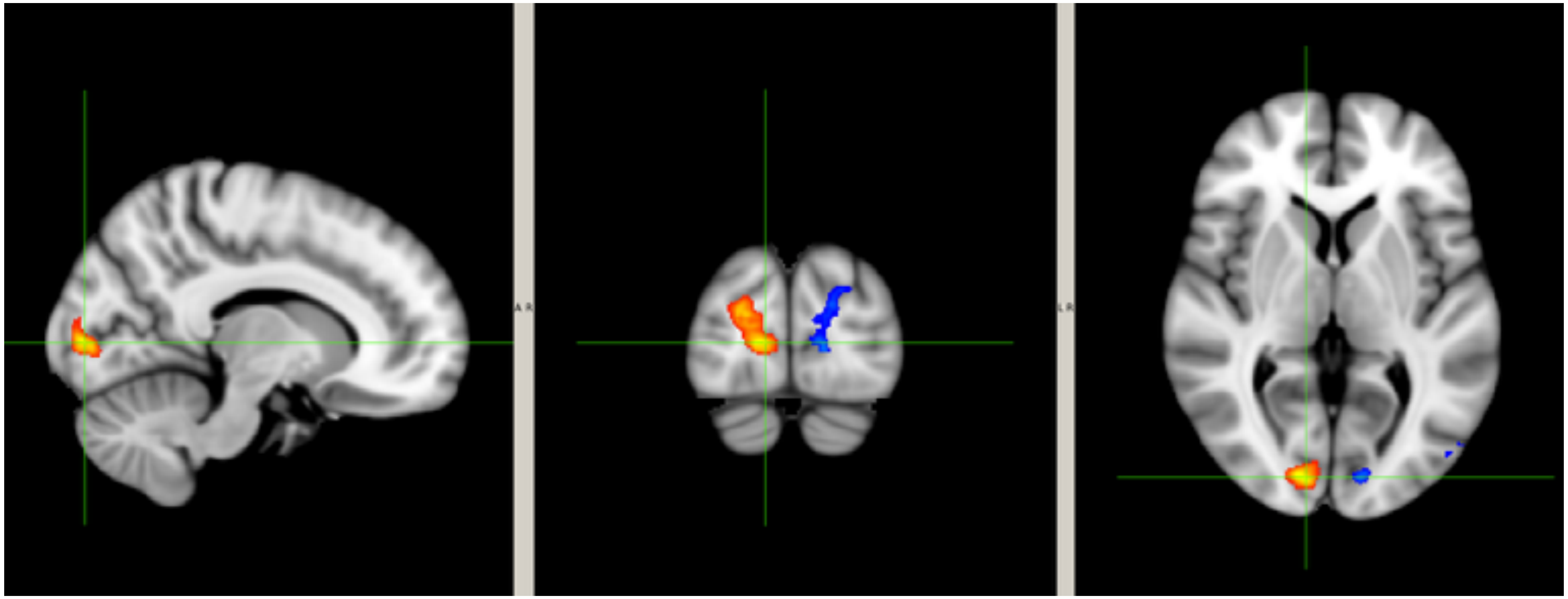
Incongruent > congruent (same response hand)

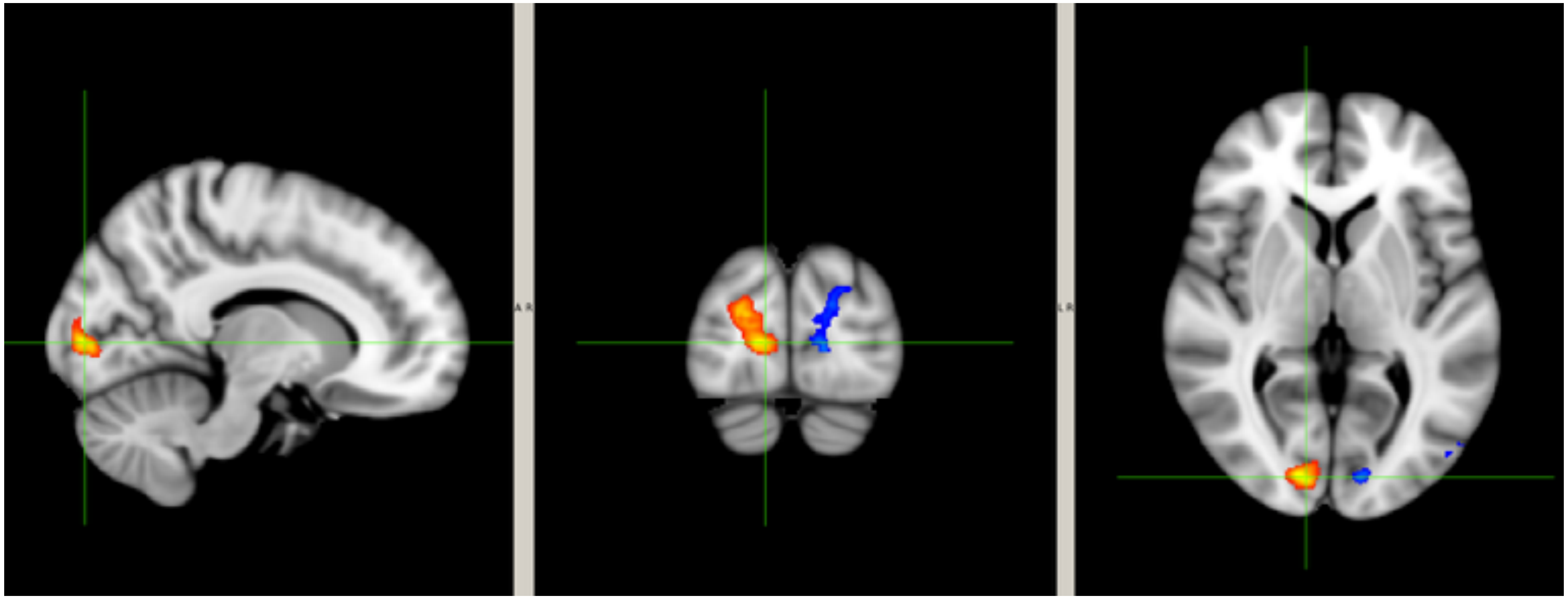
Visual

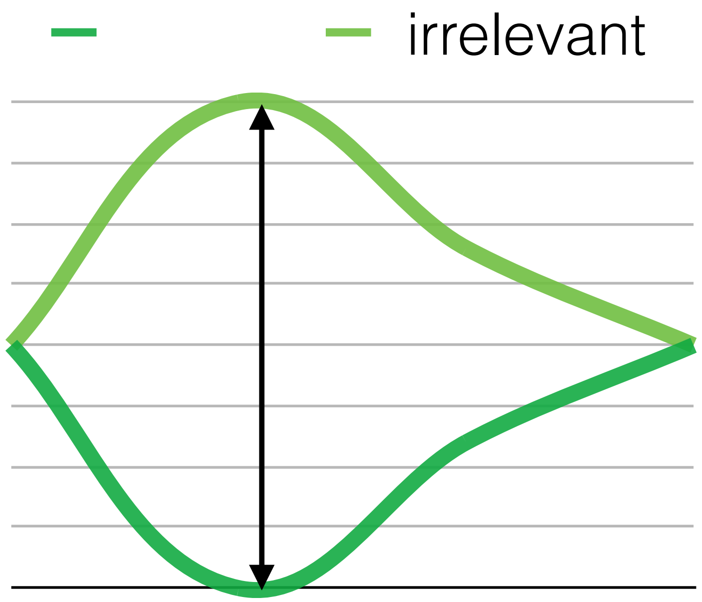


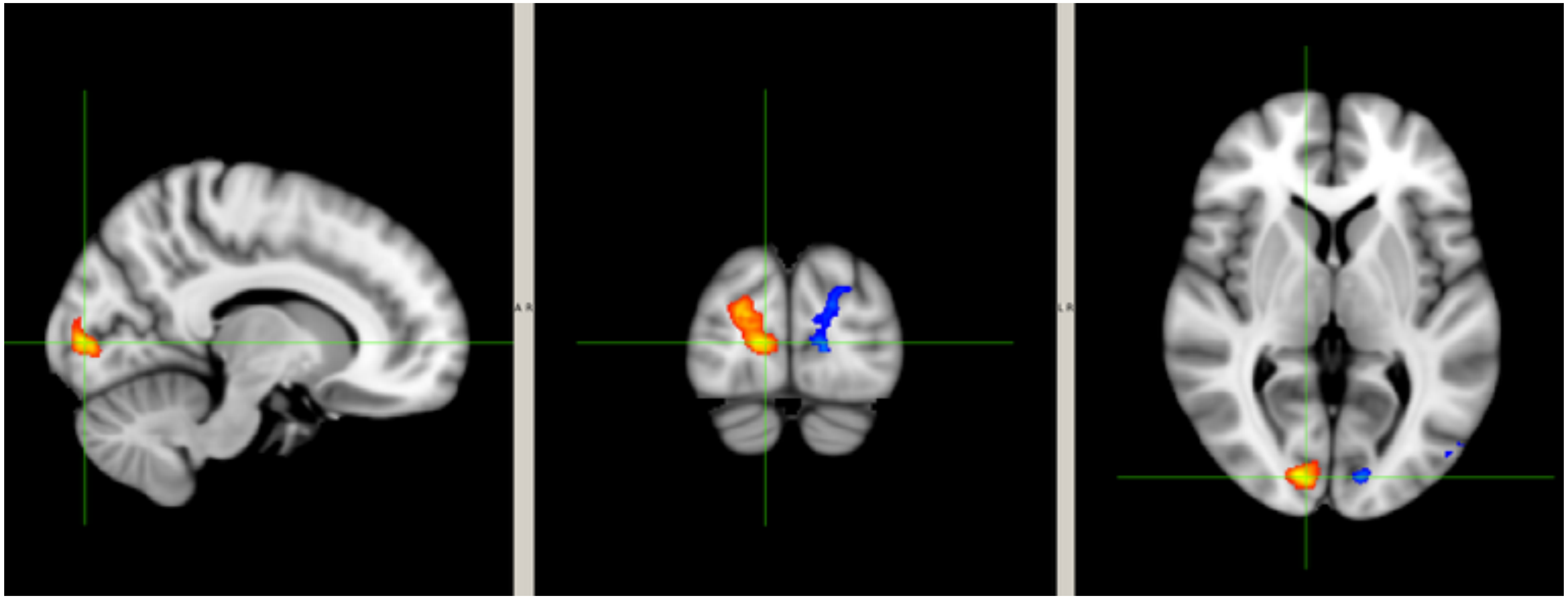
RED: Left_Congruent < Right_Incongruent (Always left response)

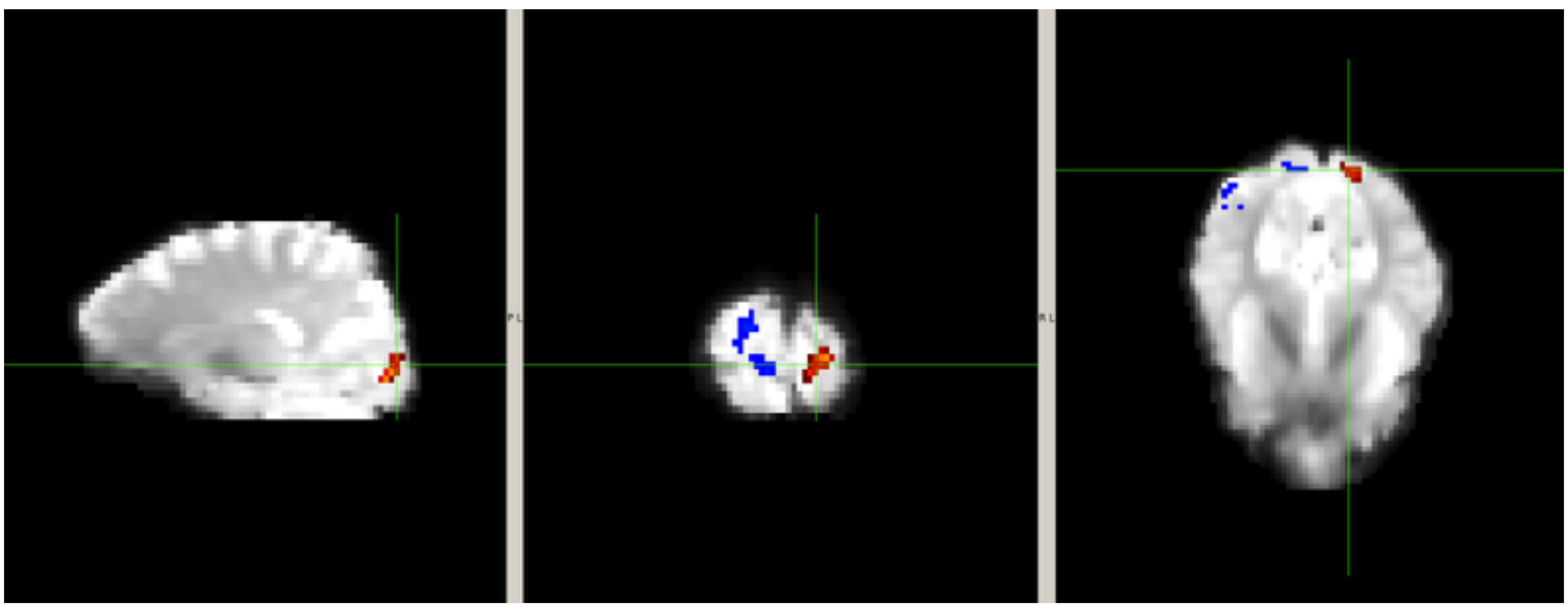
BLUE: Right_Congruent < Left_Incongruent (Always right response)

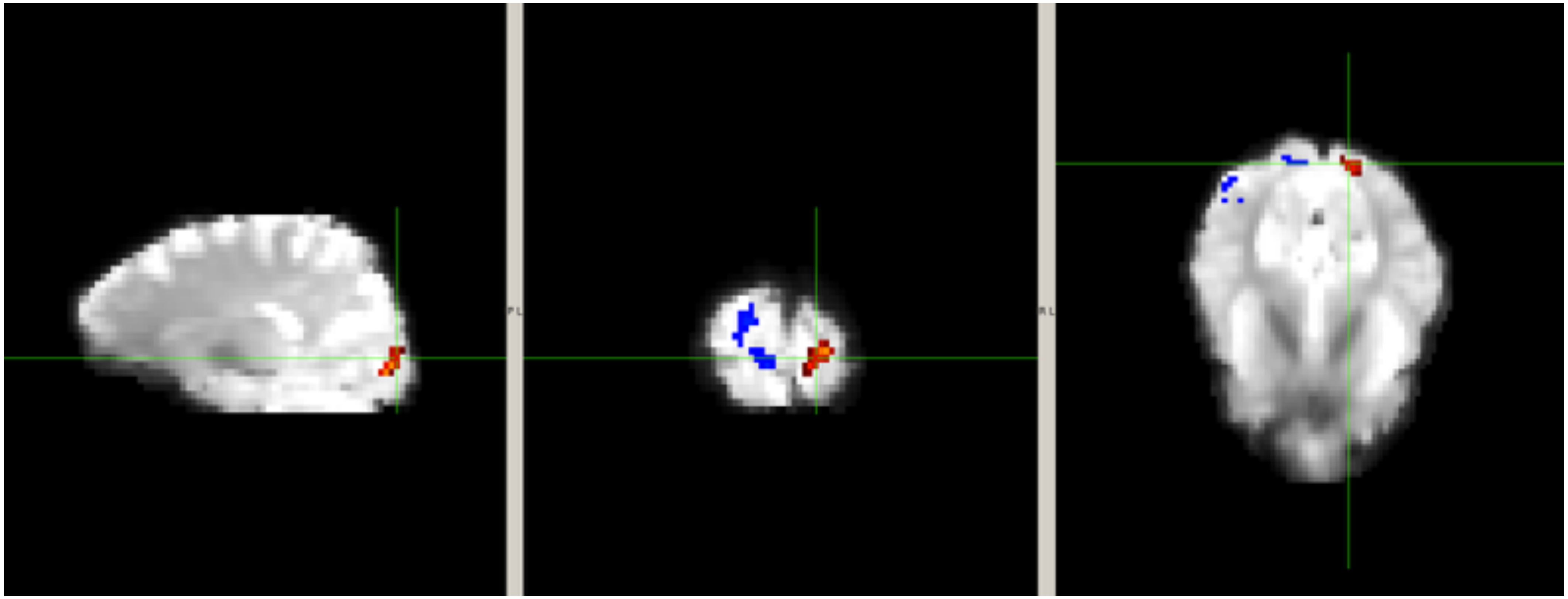


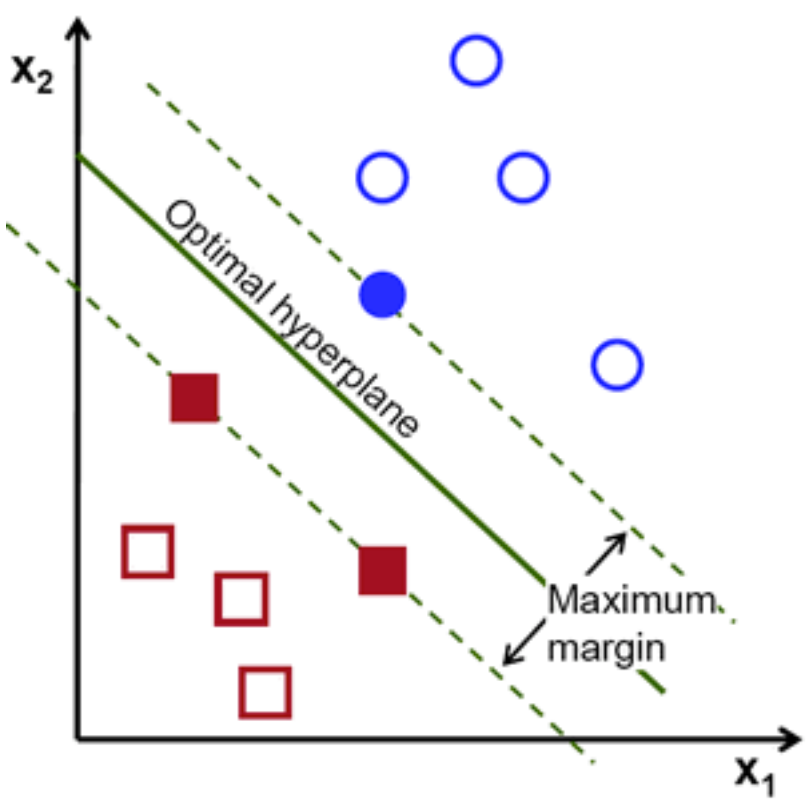


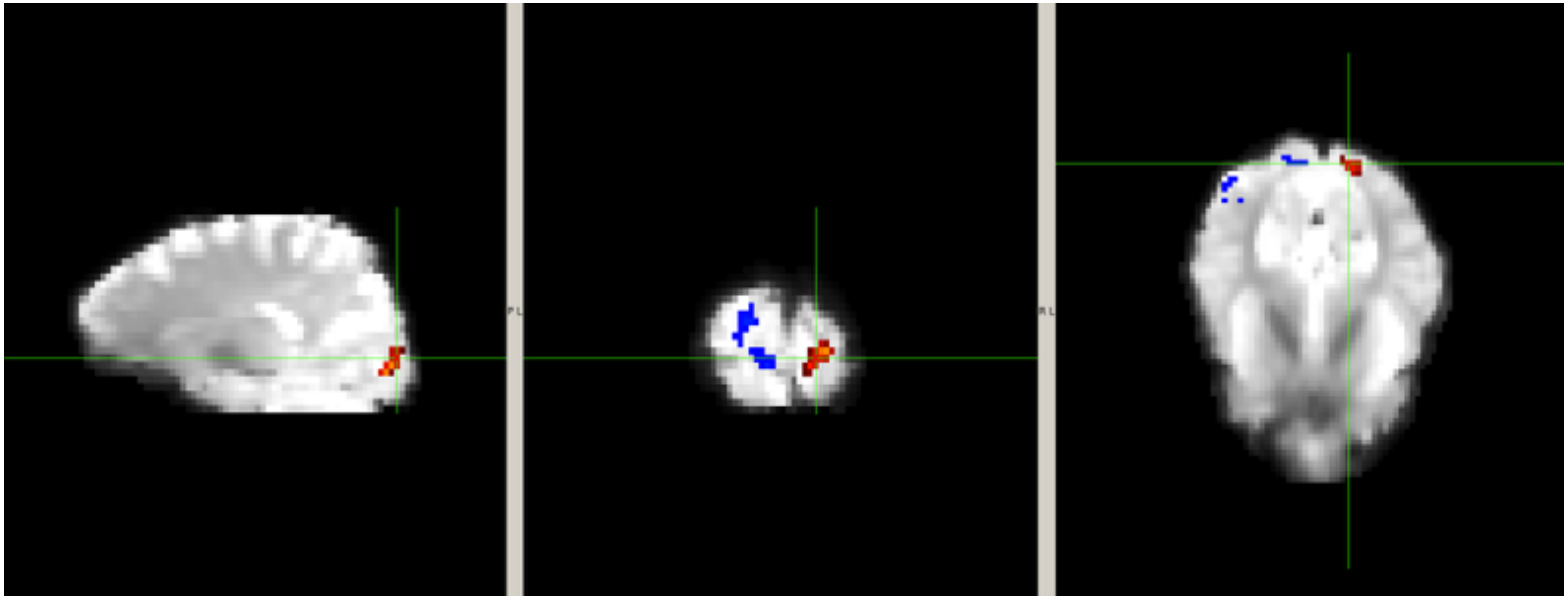

irrelevant

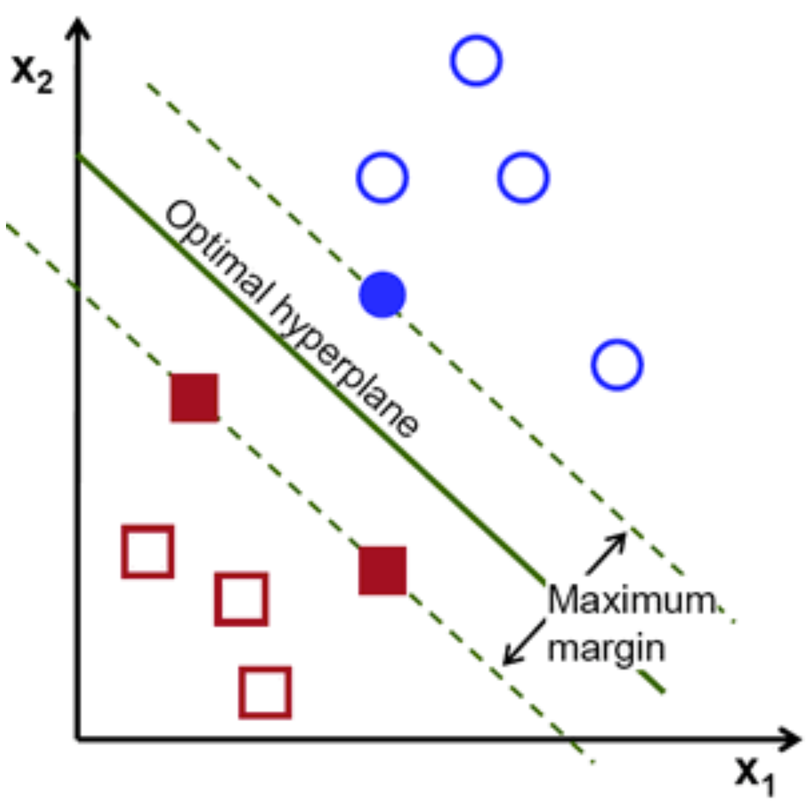






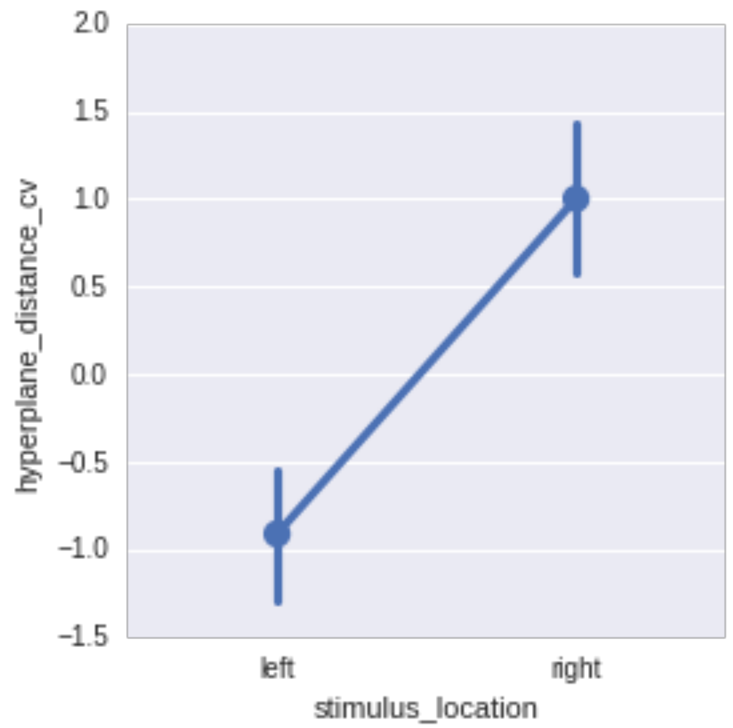

x2
Optimal hyperplane
Maximum
margin
x1



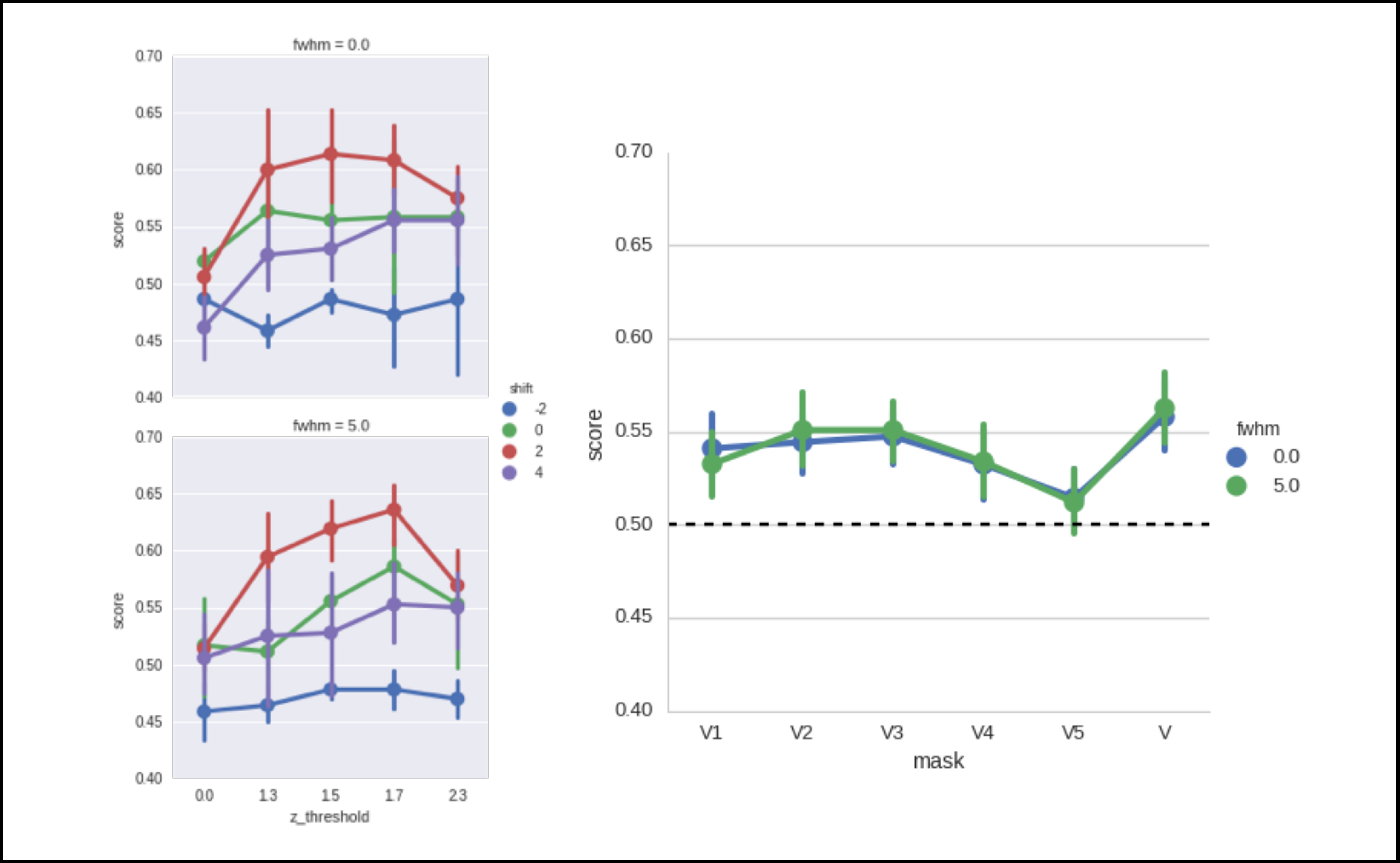


test	train	train
train	test	train
train	train	test

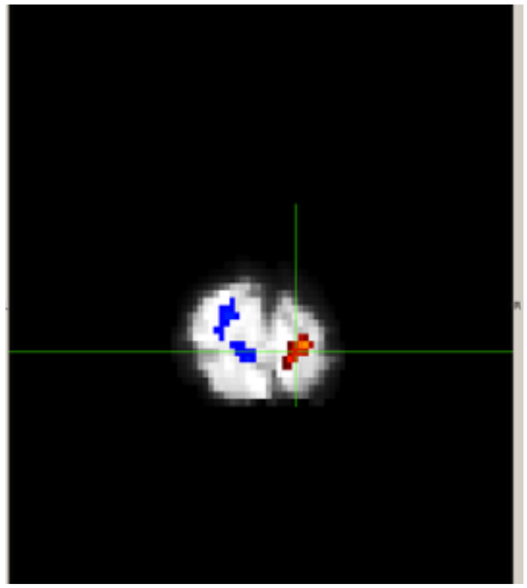
SVM hyperplane distance

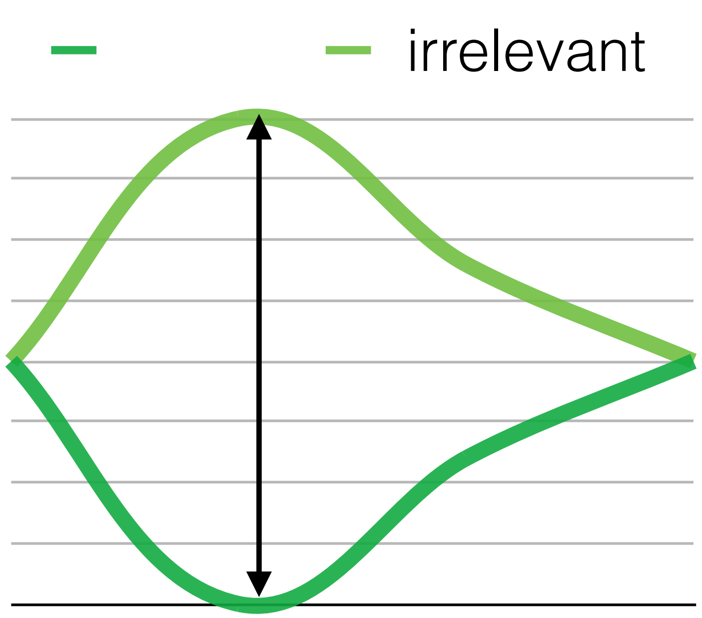


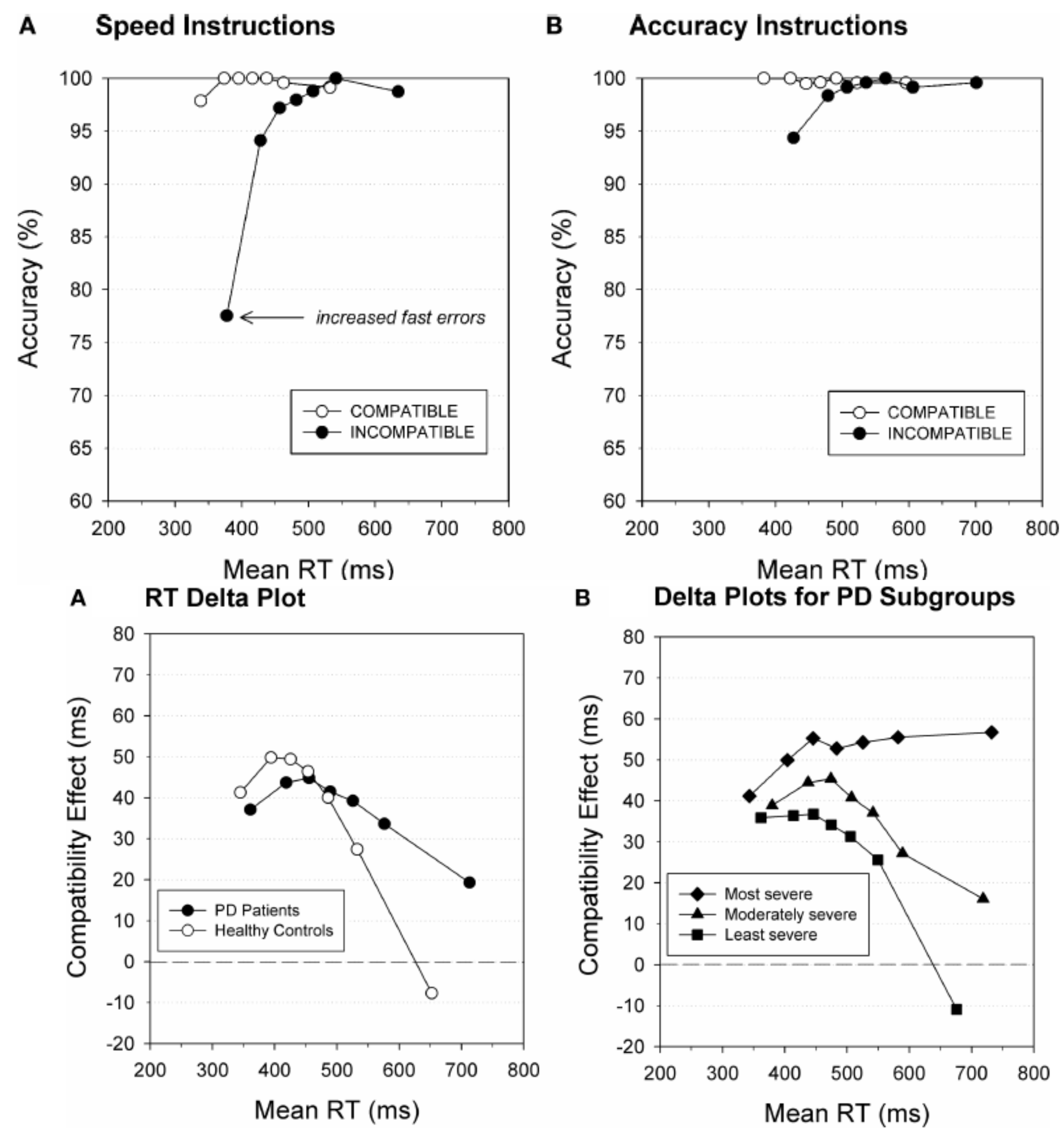
SVM hyperplane distance



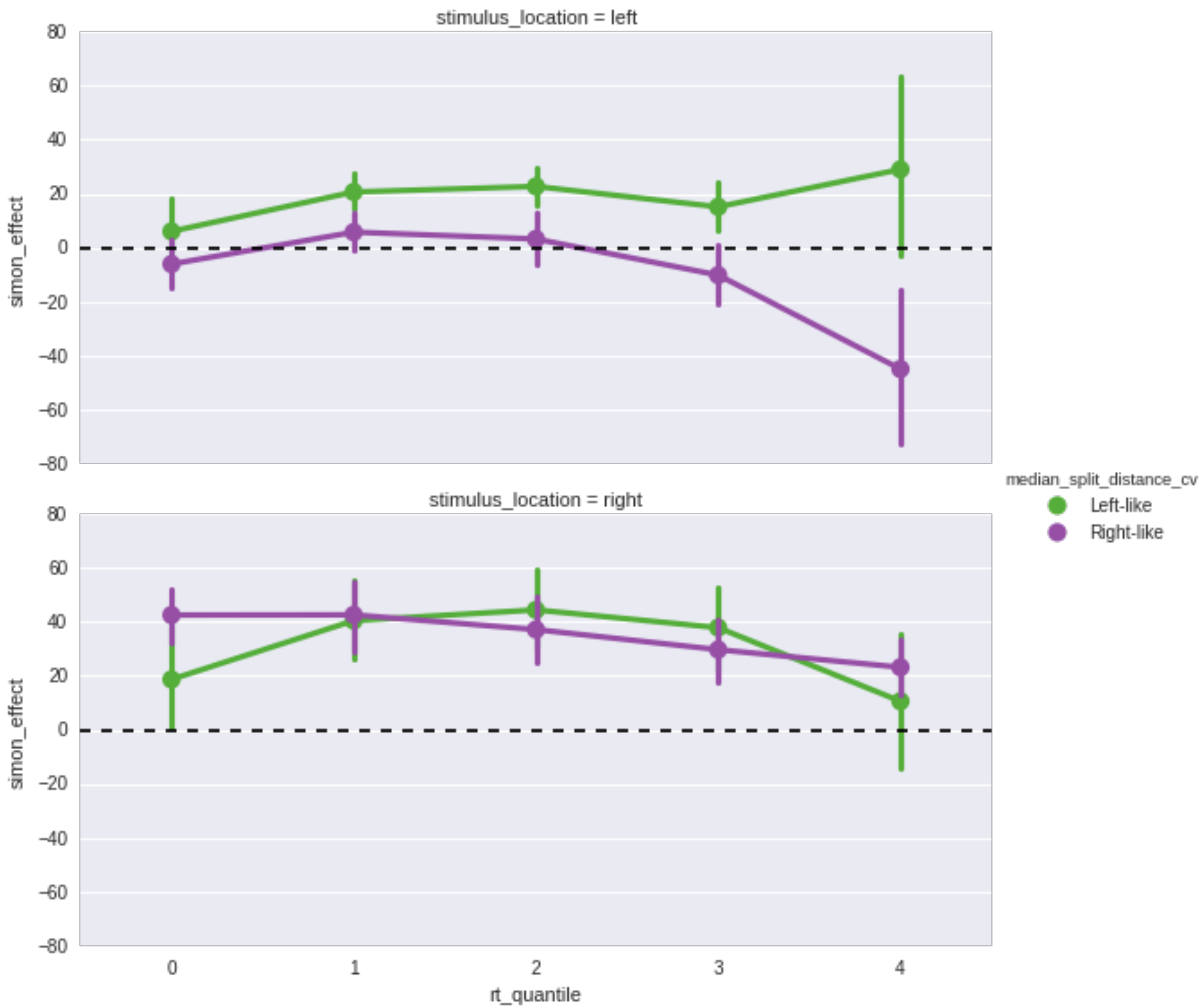
SVM hyperplane distance








stimulus_location = left
stimulus_location = right
rt
460
440
420
400
380
Left-like
Right-like
likeness
congruency
incongruent
congruent


stimulus_location = left
stimulus_location = right
simon_effect
rt_quantile
median_split_distance_cv
Left-like
Right-like
80
60
40
20
0
−20
−40
−60
−80
0
1
2
3
4

Summary

- The “irrelevant feature stimulus” feature can be detected in the visual task using MVPA
- The trial-to-trial fluctuations in this measure predict size and temporal dynamics of Simon effect

Thanks for your attention