Recurrence is required to capture the representational dynamics of the human visual system

Tim C Kietzmann, Courtney J Spoerer, Lynn K A Sörensen, Radoslaw M Cichy, Olaf Hauk, and Nikolaus Kriegeskorte

Presented by Hal Rockwell 1/23/21

Motivations

- Trying to elucidate the way that visual representations evolve over time and interact between areas

- Mostly about the methods, showing that they produce expected results
  - That’s my judgement, at least

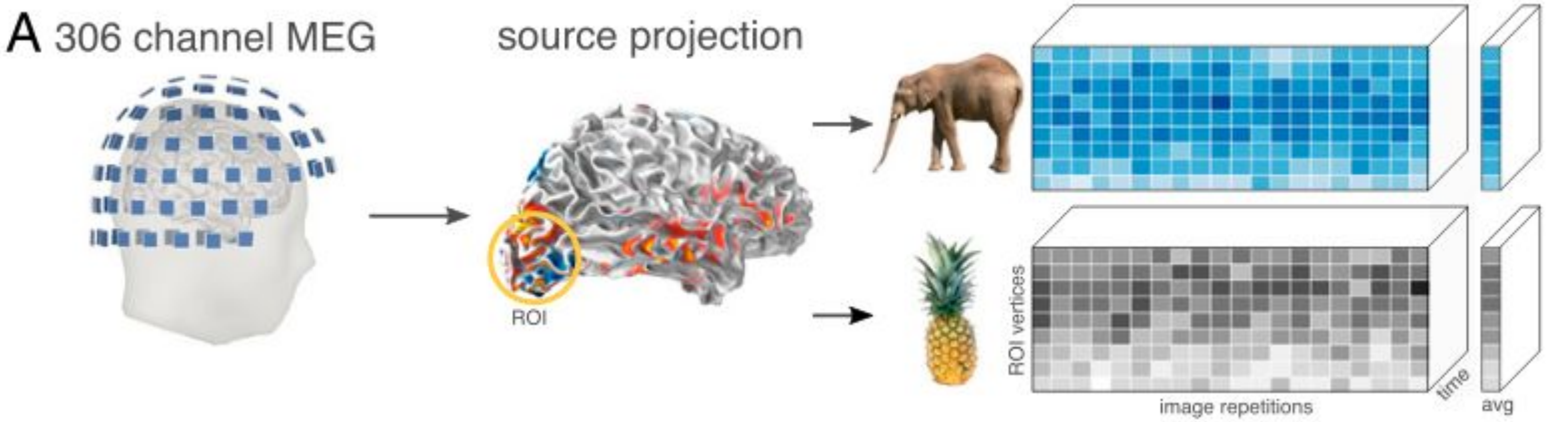
Main points

- RDA (Representational Dynamics Analysis) in multiple brain areas
    - Very cool methodology, though with some caveats

- Recurrent CNNs trained via Representational Distance Learning to replicate the representational dynamics of the neural data

Data

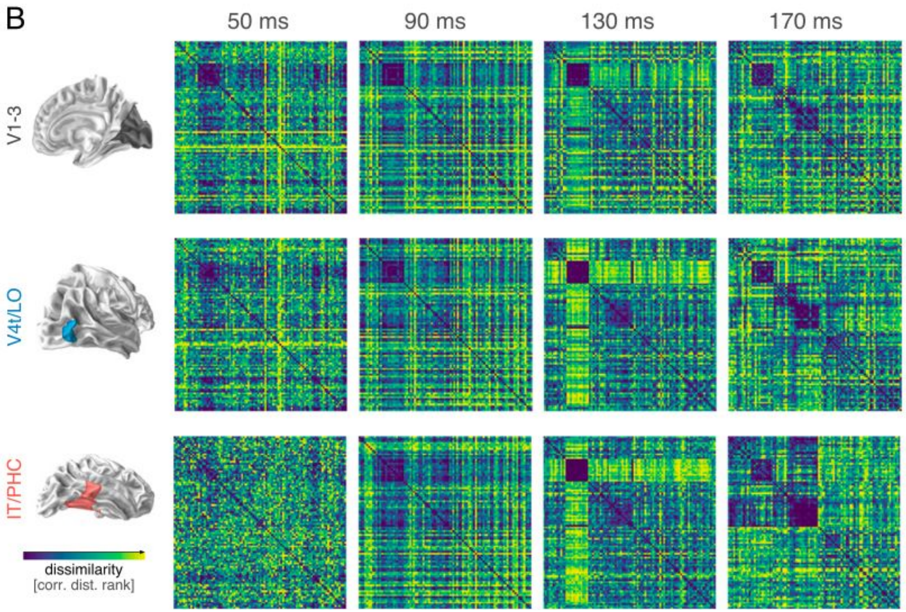
- MEG from 15 humans, 92 static image stimuli from “diverse” categories

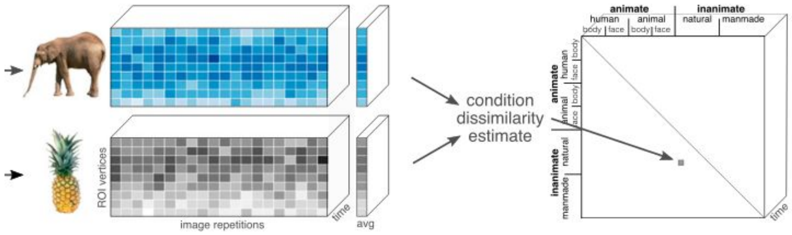


- Split areas into early, middle, high (V1-V3, V4/LO, IT/PHC)

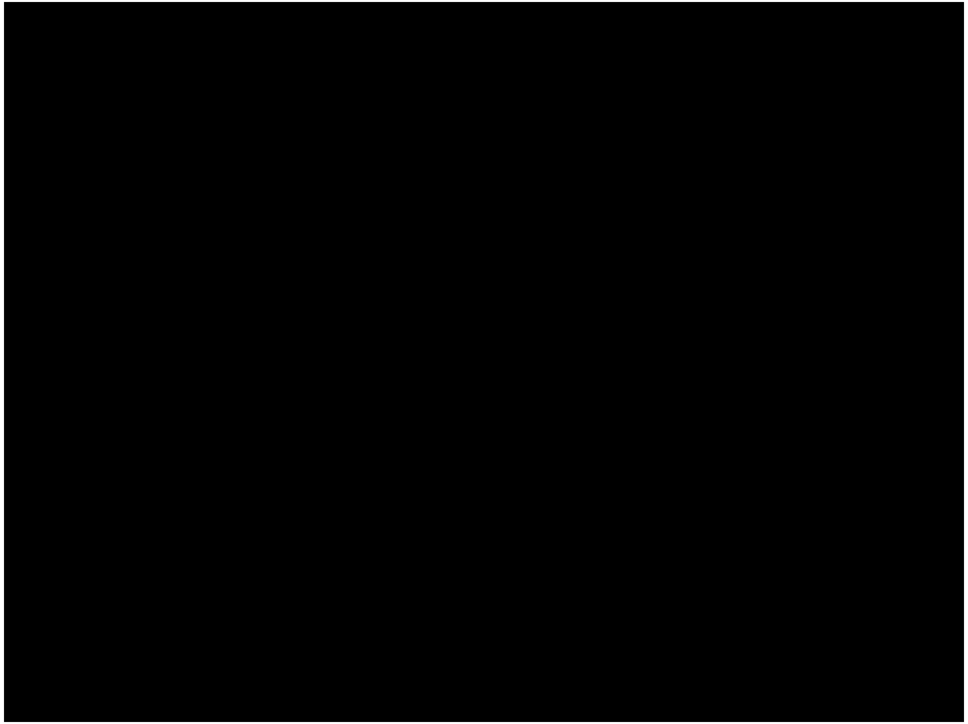
Representational dynamics analysis

- Get an RDM (representational dissimilarity matrix) for each discrete time bin



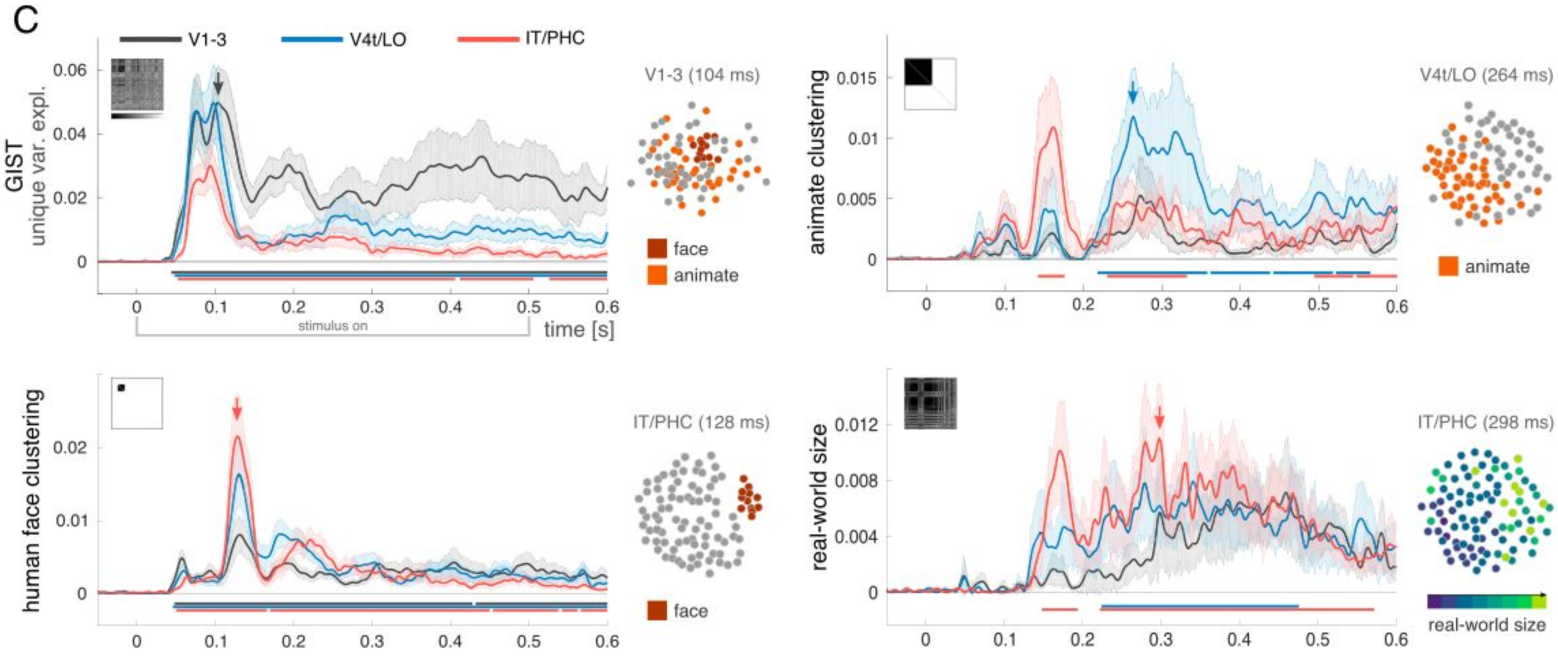


RDA movie



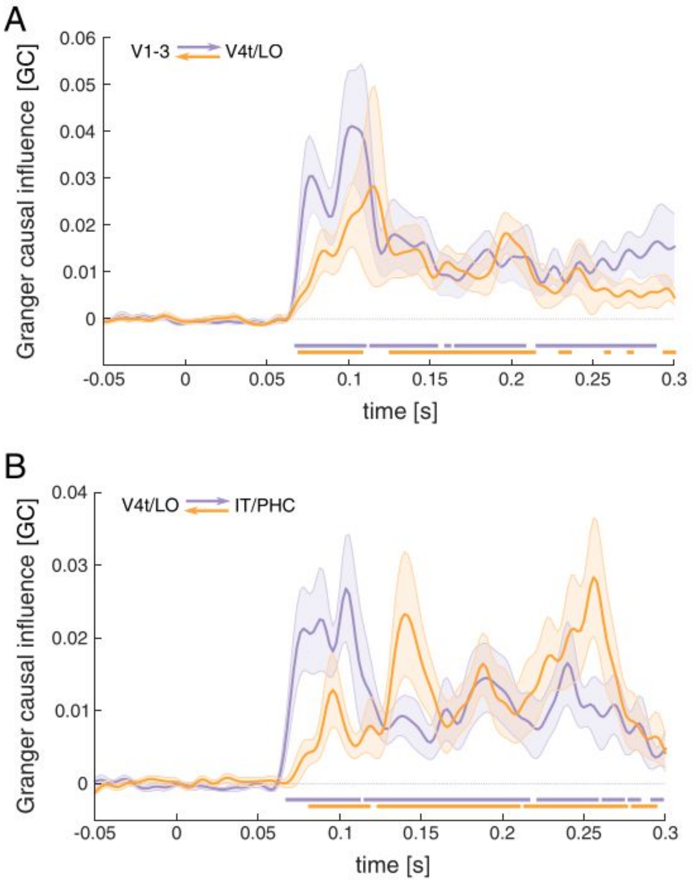
Linear modeling of RDA

- Model RDMs as nonnegative combination of component RDMs like Gabor wavelets, animacy, size, face-or-not



RDA Granger causality analysis

- How well an area can be predicted by a different area's past over its own
- Statisticians have some issues, so it's not perfectly "causality", but better than nothing
- Results are roughly as expected -- feedforward then recurrent



RDA conclusions

- Rich bidirectional information flow between areas
- Significant temporal changes in structure of representations
- Implies recurrence is very important to the basic computations of the ventral visual stream
  - Which we kind of already knew...

CNN modeling

- Intention is to further test the hypothesis that recurrence is important in developing visual representations

- Train two classes of CNN to match the RDA: feedforward and recurrent
  - Bottom-up only (B) and bottom-up, lateral, and top-down (BLT) :)
  - Parameter-matched by increasing kernel size :(

- B models have “ramping-up” for dynamics

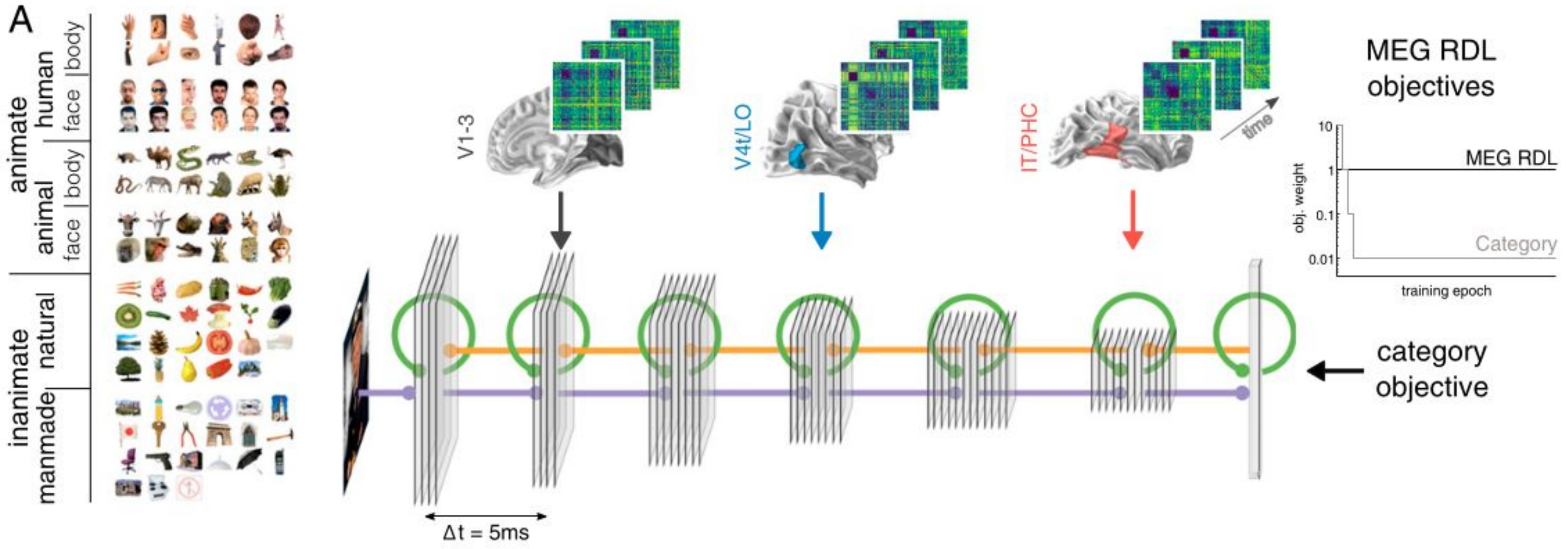
$$\boldsymbol{h}_{\tau,\mathrm{cat}} = W^{b}_{\mathrm{cat}}\, \bar{\boldsymbol{h}}_{\tau-1,N} + \omega_{\mathrm{cat}} \boldsymbol{h}_{\tau-1,\mathrm{cat}} + \boldsymbol{b}_{\mathrm{cat}},$$

CNN modeling

- RDL + categorization loss

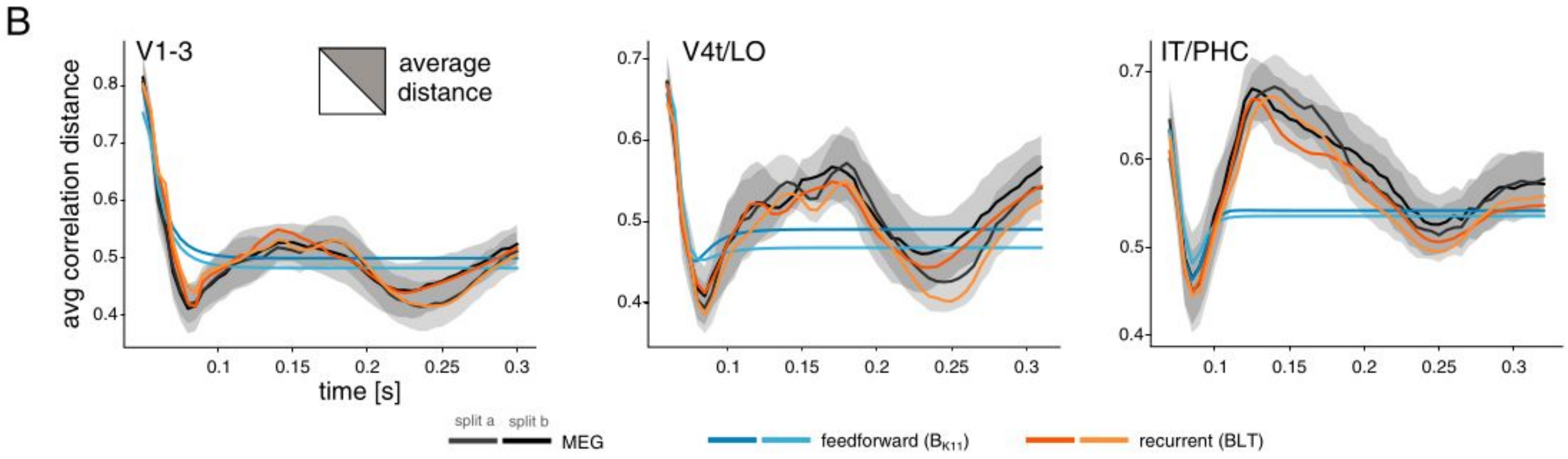
$$E_{RDL} = \sum_{n \in L,\ r \in R} \frac{1}{T} \sum_{\tau}^{T} \frac{\left(\hat{d}_{\tau,n}(x_a, x_b) - d_{\tau,r}(a,b)\right)^2}{\sigma^2_{\tau,r}}$$

$$\mathcal{L} = \gamma_{cat} \bar{E}_{cat} + \gamma_{RDL} \bar{E}_{RDL} + \lambda |\mathbf{w}|_2,$$

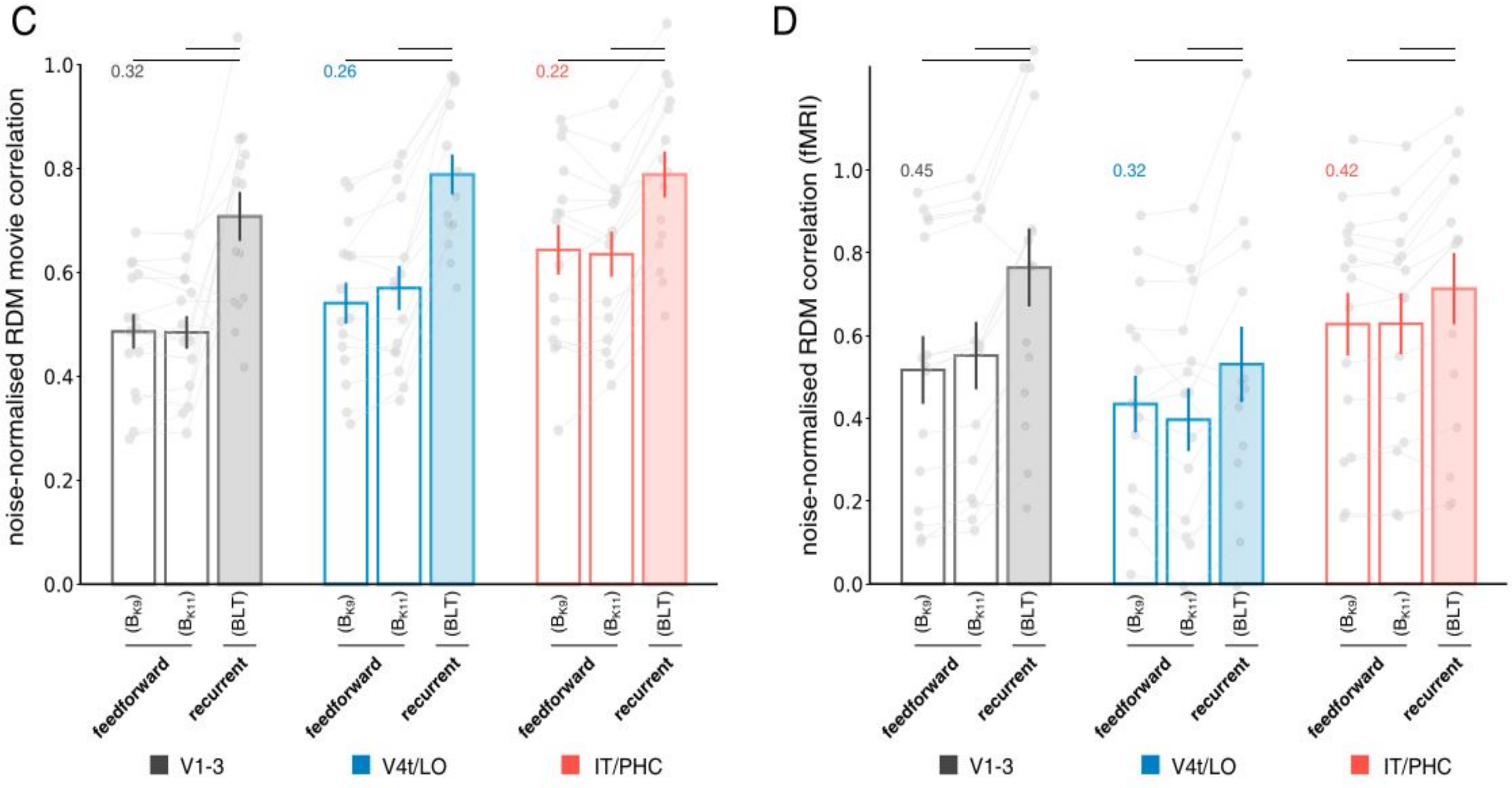


CNN results

- Average distance between patterns is well-matched to data
  - This is a pretty rough metric



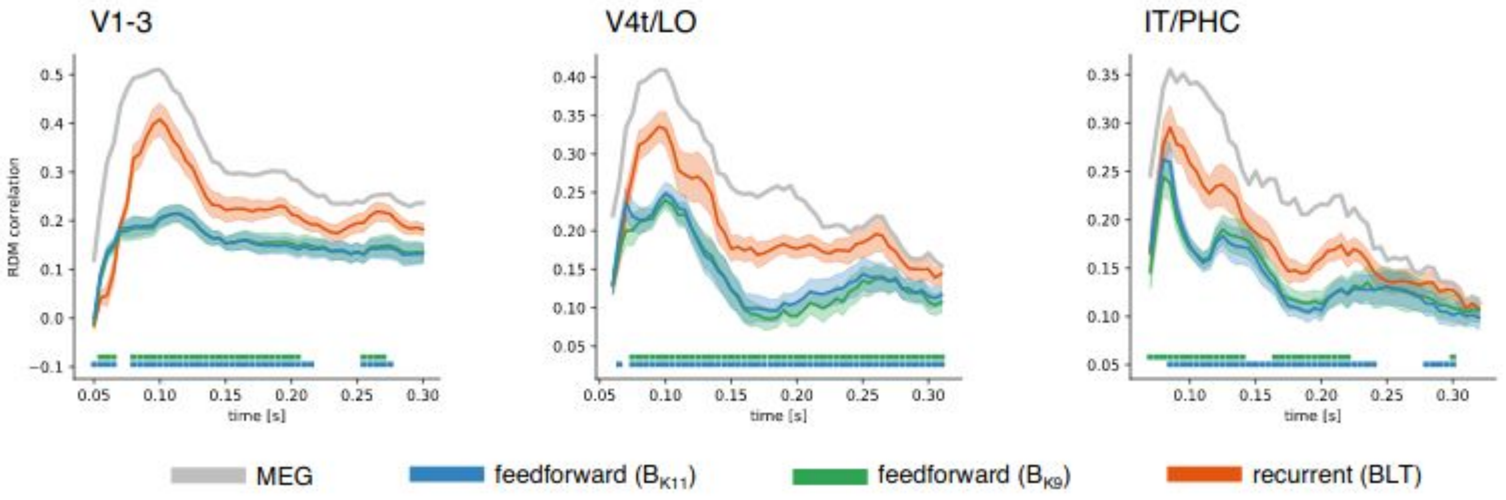
CNN results



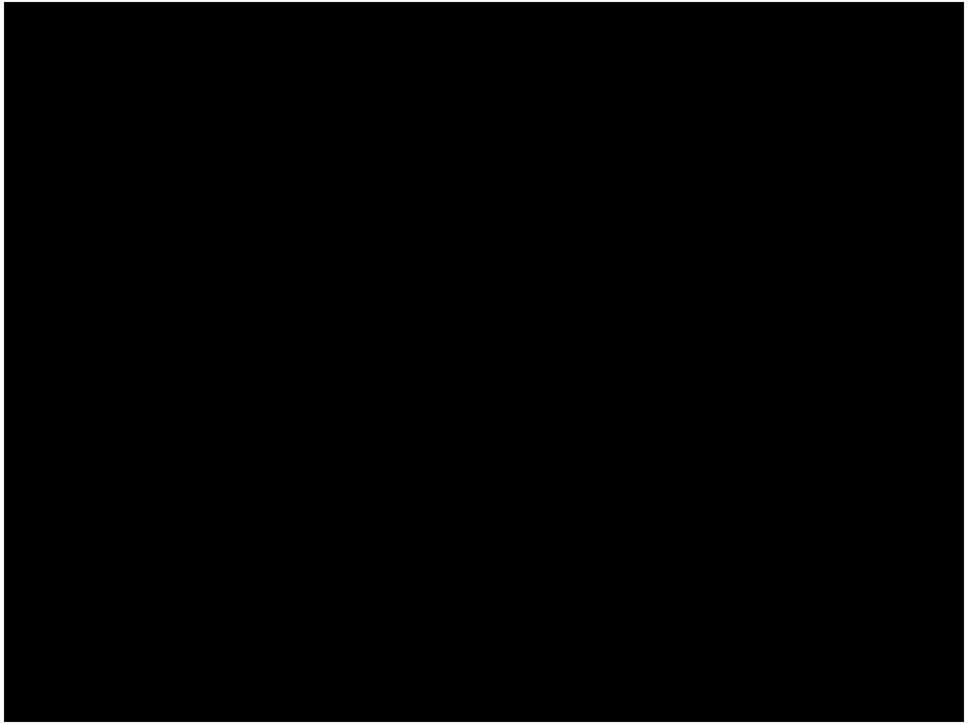
C
noise-normalised RDM movie correlation
1.0
0.8
0.6
0.4
0.2
0.0
0.32
0.26
0.22
(B K9)
(B K11)
(BLT)
feedforward
recurrent
V1-3
V4t/LO
IT/PHC
D
noise-normalised RDM correlation (fMRI)
0.45
0.32
0.42

CNN results across time

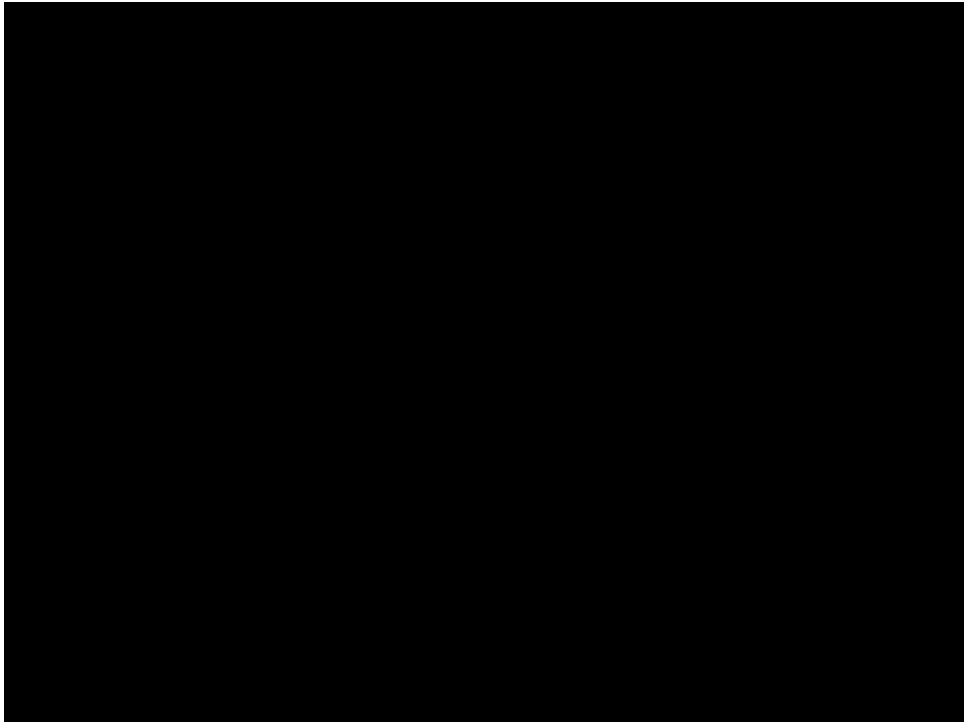
- Predictably, performance decreases
  - Just a supplementary result



BLT CNN movie

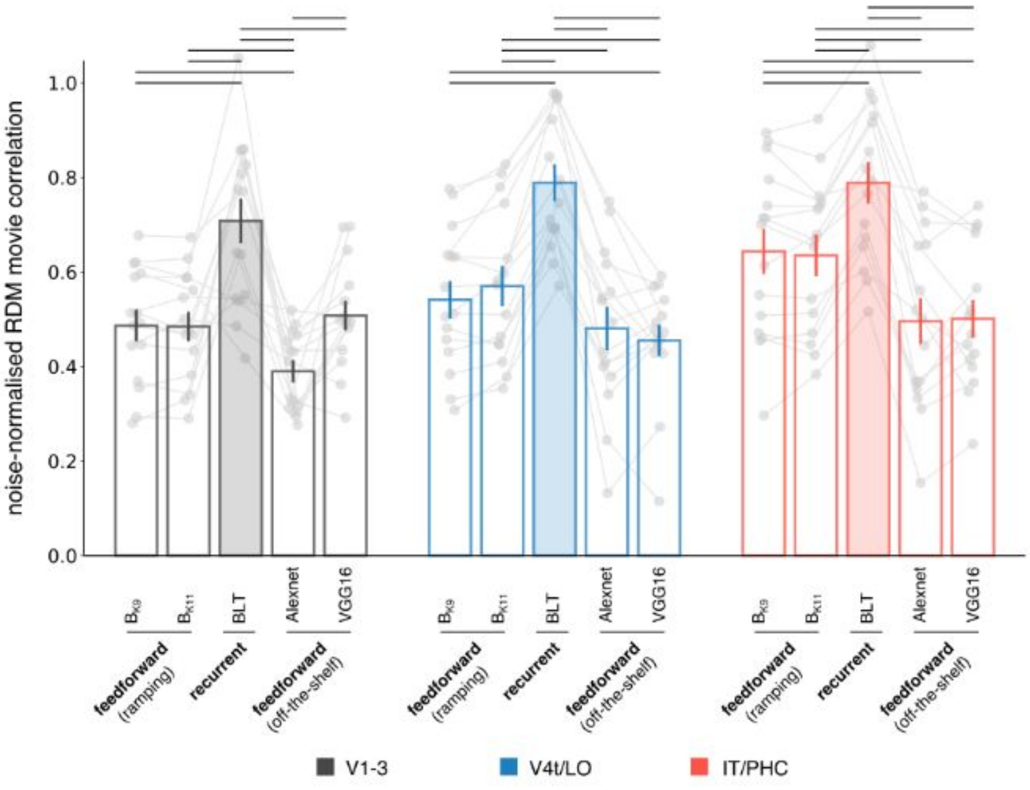


B CNN movie (worse)



CNN results -- Imagenet models

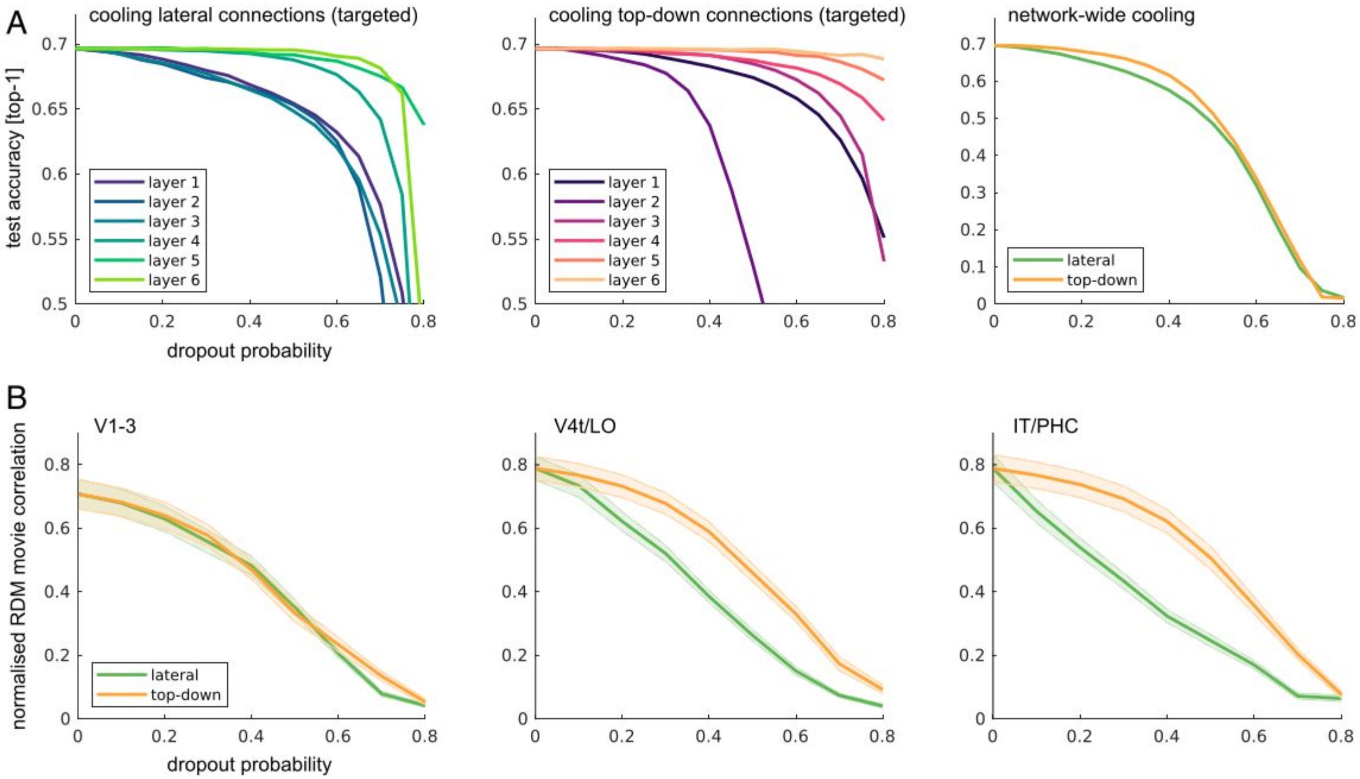
- Just a supplementary result



CNN “cooling” (dropout) experiments

- Lower layers more important for classification

- RDMs rely less on top-down connections for high areas



Overall conclusions

- Dynamics of image processing in the human stream rely heavily on recurrence both within and between areas
  - Supported by RDA and CNN modeling

- Really we already knew that, but it’s good to quantitatively check, and this helps to validate the RDA and RDL-trained CNN approaches used

- Would be interesting to try to apply this to higher-resolution data, to look at more specific computation than just e.g. “top-down between areas”