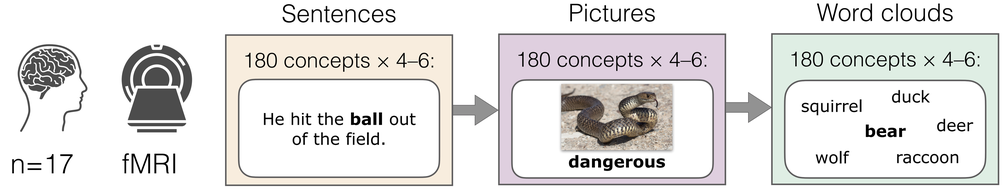
Language models align with brain regions that represent concepts across modalities

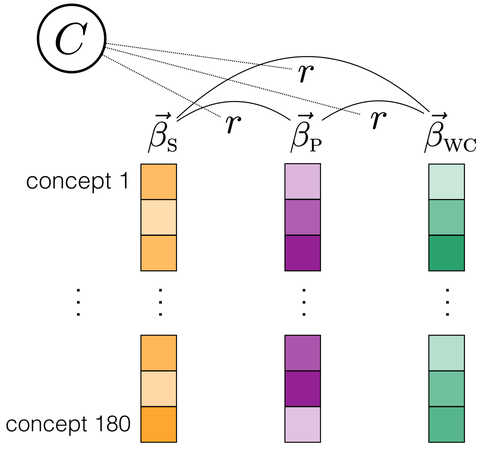
Maria Ryskina, Greta Tuckute, Alexander Fung, Ashley Malkin, Evelina Fedorenko

1. Brain responses to concepts (Pereira et al., 2018, Exp. 1)



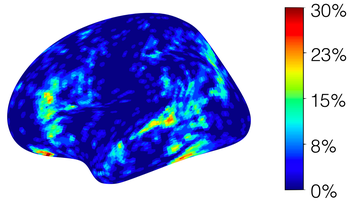
2. Semantic consistency metric

Correlation between brain responses to concepts across stimuli formats

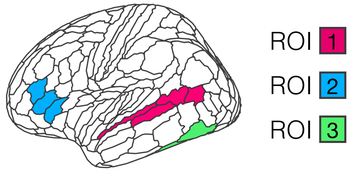


3. Consistent brain areas

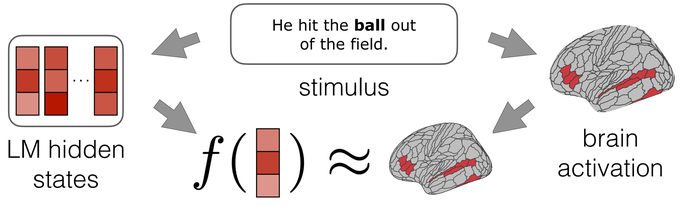
Probabilistic map:



Regions of interest:

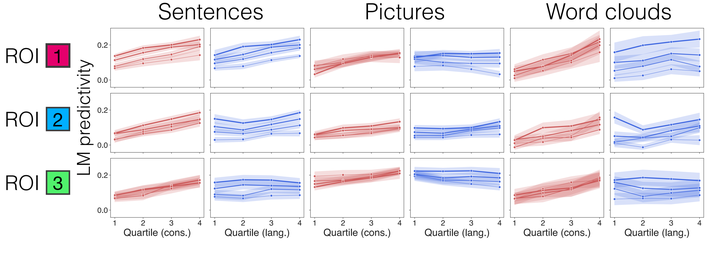


4. LMs for brain encoding



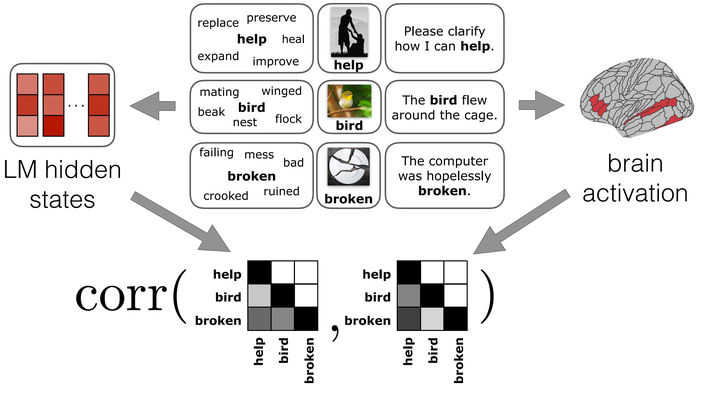
Finding 1. Across the whole brain, areas with higher semantic consistency—likely to represent concepts across modalities—are better predicted by LMs

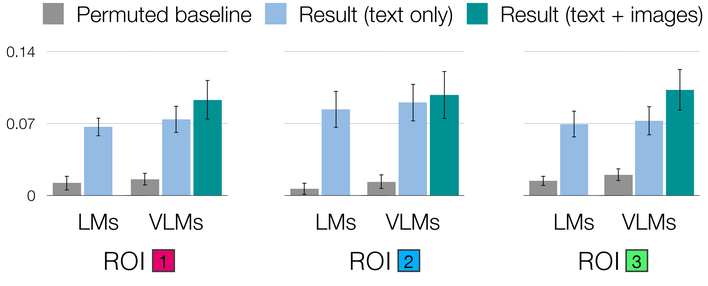
What drives LM encoding performance — semantic consistency or response to language?



Finding 2. Brain encoding quality within the chosen ROIs correlates strongly with semantic consistency, even when decoupled from language

5. Representational similarity





Finding 3. All LMs/VLMs show significant representational alignment with semantically consistent ROIs, and it further increases when both text and image stimuli are used