Mapping social perception to social behavior using artificial neural networks

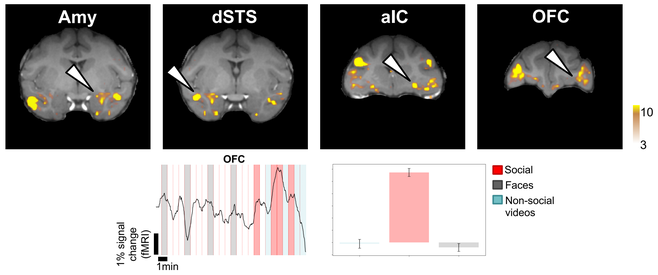
N. Dolensek[1], S. Chen[1], S. Ginosar[3,4], D. Y. Tsao[1,2]

[1]University of California, Berkeley; [2]Howard Hughes Medical Institute; [3]TTIC, [4]Google DeepMind

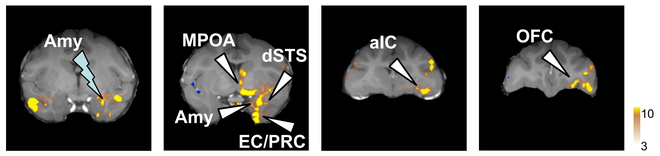
Abstract

Primates possess sophisticated social cognitive abilities, interpreting complex social cues and producing adaptive behaviors in real-time. These abilities are thought to rely on distributed circuits spanning sensory, limbic, and prefrontal brain regions. Despite some progress in identifying neural correlates of specific social features, such as gaze and social rank, a comprehensive computational framework linking social perception to behavior remains elusive; at least in part due to the high dimensionality of both the input space (complex social scenes) and the output space (diverse behavioral responses). To address this, we combine artificial neural networks (ANNs) and high-density Neuropixels electrophysiological recordings to uncover how the primate brain transforms dynamic social scene perception into social behavior. We first employ whole-brain functional magnetic resonance imaging (fMRI) to locate brain regions selective for social video in macaques, specifically identifying the dorsal bank of the superior temporal sulcus (dSTS), orbitofrontal cortex (OFC), and anterior insula (aIC). We then target these regions using Neuropixels probes, revealing strong selectivity for social stimuli on a single-unit level. By training a set of neural networks to embed and reconstruct videos from neural activity, we surprisingly observe that neural responses in OFC and aIC support astonishingly accurate reconstructions of social videos. We further extend this approach to generate optimal stimuli for neurons from each region and identify interpretable axes of neural responses like social partner angle and distance. Finally, we identify a subset of neurons predicting social behavioral responses and observe a causal role of activity in aIC and OFC in social behavior production, with electrical microstimulation evoking gaze shifts and facial movements. These results suggest that these frontal brain regions contain a surprisingly rich code for social scenes and play a critical role in transforming social perception into behavior, positioning them as central nodes in social cognition.

fMRI reveals a set of social processing regions in the macaque brain

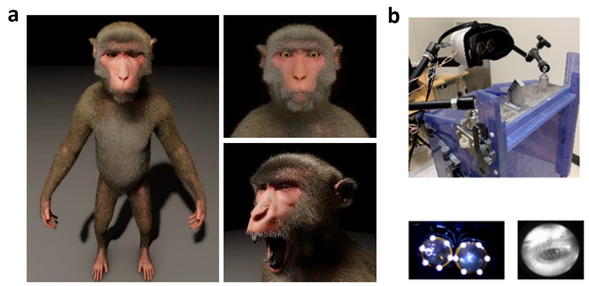


Social processing regions form an interconnected network



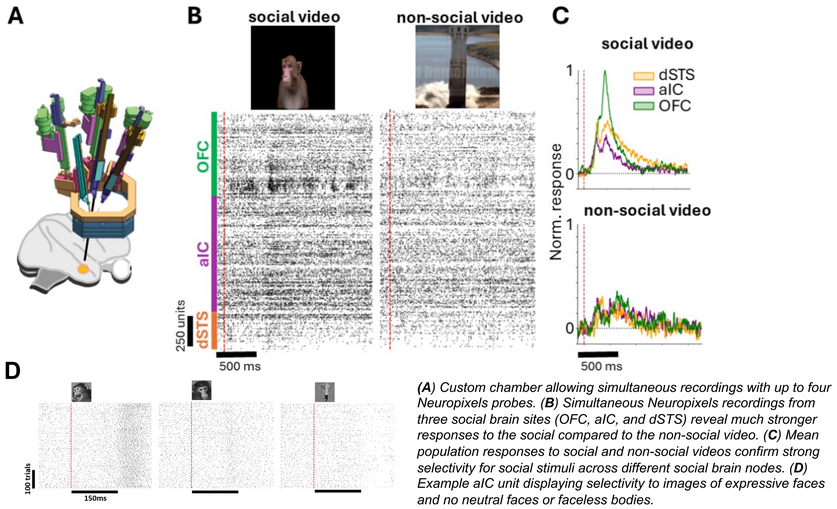
Virtual macaque avatar and Virtual Reality (VR)

Social interaction in primates is an active, bidirectional process. To study all aspects of social cognition, we developed a virtual macaque avatar (**a**) enabling real-time programmatic control of all avatar aspects (appearance, movements) and facilitating true interactive tasks in VR (**b**).



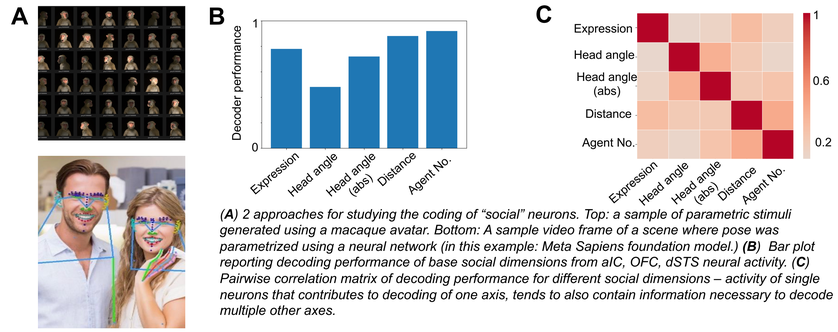
fMRI-targeted NHP Neuropixels recordings reveal selectivity for social stimuli at a single-unit level

Recording from thousands of cells across multiple social brain sites simultaneously reveals rich coding for naturalistic social stimuli and complex temporal dynamics.



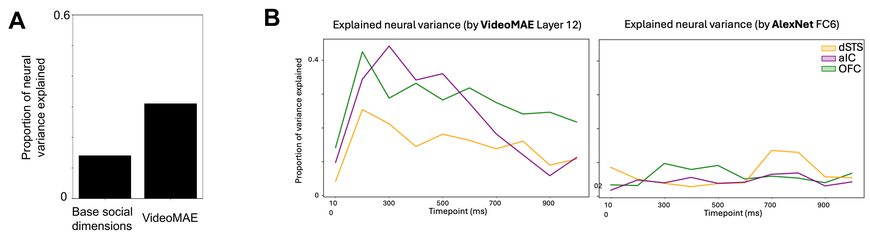
*(**A**) Custom chamber allowing simultaneous recordings with up to four Neuropixels probes. (**B**) Simultaneous Neuropixels recordings from three social brain sites (OFC, aIC, and dSTS) reveal much stronger responses to the social compared to the non-social video. (**C**) Mean population responses to social and non-social videos confirm strong selectivity for social stimuli across different social brain nodes. (**D**) Example aIC unit displaying selectivity to images of expressive faces and no neutral faces or faceless bodies.*

Units in social regions encode basic social dimensions, but exhibit mixed selectivity

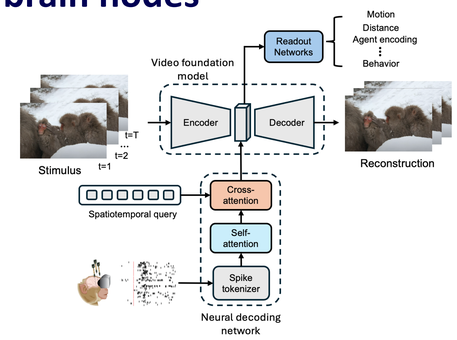


*(**A**) 2 approaches for studying the coding of "social" neurons. Top: a sample of parametric stimuli generated using a macaque avatar. Bottom: A sample video frame of a scene where pose was parametrized using a neural network (in this example: Meta Sapiens foundation model.) (**B**) Bar plot reporting decoding performance of base social dimensions from aIC, OFC, dSTS neural activity. (**C**) Pairwise correlation matrix of decoding performance for different social dimensions – activity of single neurons that contributes to decoding of one axis, tends to also contain information necessary to decode multiple other axes.*

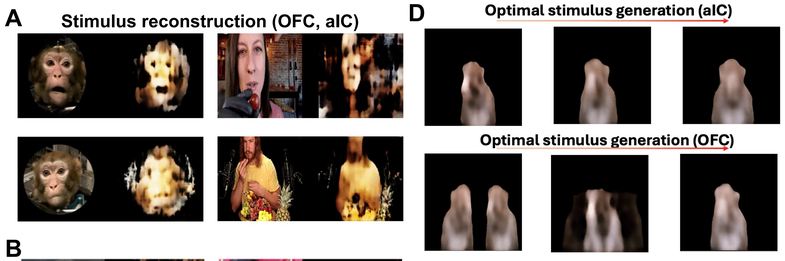
Social neuron activity aligns with video DNN features

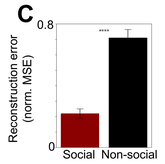


ANN approach for uncovering representations in social brain nodes



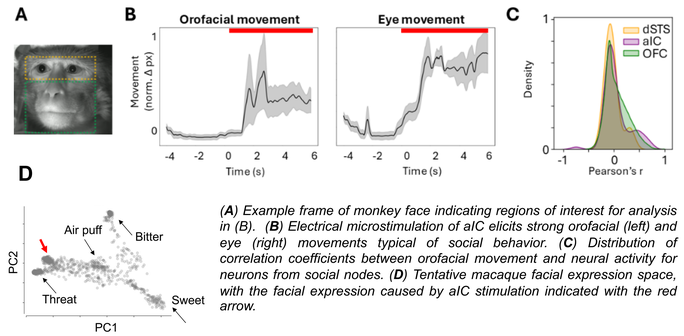
Social nodes contain detailed representations of social scenes





*(**A**) Reconstruction of social video stimuli from neural responses in OFC and aIC using our ANN approach. (**B**) Reconstruction of non-social videos. (**C**) Quantification of the reconstruction results for social and non-social stimuli by embedding of reconstructed frames in ResNet50 and measuring the MSE to stimulus ground truth in the ResNet latent space. (**D**) Optimizing stimuli by increasing values of single PCs of neural activity results in interpretable axis in the resulting reconstructed videos (for example, head-turning, changing distance between agents)*

Activity in social brain nodes is causally linked to social behavior



*(**A**) Example frame of monkey face indicating regions of interest for analysis in (B). (**B**) Electrical microstimulation of aIC elicits strong orofacial (left) and eye (right) movements typical of social behavior. (**C**) Distribution of correlation coefficients between orofacial movement and neural activity for neurons from social nodes. (**D**) Tentative macaque facial expression space, with the facial expression caused by aIC stimulation indicated with the red arrow.*