Disentangling Human Amygdala Activity with Artificial Neural Networks

Grace Jang (grace.jang@emory.edu)
Neuroscience Graduate Program, Graduate Division of Biological and Biomedical Sciences, Emory University
Atlanta, GA 30032 USA

Philip A. Kragel (pkragel@emory.edu)
Department of Psychology, Department of Psychiatry and Behavioral Sciences, Emory University
Atlanta, GA 30032 USA

Abstract:

Human neuroscience has revealed that neuroimaging measures and neural activity in the amygdala encodes a wide array of variables ranging from threat, salience, valence, and stimulus intensity. Although much has been learned about amygdala function, an overarching account of amygdala processing remains elusive. Here we use a combination of human neuroimaging, computational models of visual processing, and self-report measures of emotional experience to develop and validate encoding models that predict patterns of amygdala response acquired with fMRI during movie-viewing. When tested on naturalistic emotional images, we found that amygdala encoding models predicted ratings along the dimension of valence. Moreover, we found that the encoding models could be paired with deep generator networks to synthesize artificial stimuli that specifically engage the amygdala and anatomically defined amygdala subregions. These findings establish an approach for advancing our understanding of amygdala function by identifying how the amygdala transforms rich sensory inputs into low-dimensional representations relevant for behavior.

Keywords: amygdala; encoding models; artificial neural networks; fMRI

Introduction

The amygdala is a subcortical cluster of nuclei in the medial temporal lobe that receives multimodal inputs from sensory cortices, including the ventral visual stream. Due to its connectivity and influence on multiple brain systems, it is thought to function as an integrative hub involved in processing the salience, valence, and threat value of stimuli (Janak & Tye, 2015; LeDoux, 2000; Ohman & Mineka, 2001; Pessoa & Adolphs, 2010). Through its widespread connections, the amygdala is thought to detect events of biological significance and coordinate autonomic, skeletomotor, and cognitive responses to promote survival (Sander et al., 2003). Despite numerous theories on amygdala function, how exactly the amygdala accomplishes this remains largely debated.

Here we aim to further our understanding of amygdala function by isolating the sensory-evaluative components of amygdala processing. To this end, we used deep convolutional neural networks as a computational model of the ventral visual stream (Yamins et al., 2014) and examined whether high-dimensional representations from a convolutional network trained to classify emotional situations (Kragel et al., 2019) contain information sufficient to predict amygdala responses to naturalistic videos.

Methods

Development and Validation of Amygdala Encoding Models

We developed a set of linear encoding models that predict patterns of activity within the amygdala in response to naturalistic stimuli. To accomplish this, we extracted visual features from the every fifth frame of the movie from the penultimate layer of a deep convolutional neural network (Kragel et al., 2019). This layer was selected because the majority of inputs to the primate amygdala originate in inferotemporal cortex, as opposed to earlier stages of visual processing (Kravitz et al., 2013). We convolved these features to match the hemodynamic response of the BOLD data acquired of participants viewing the same movie from the Naturalistic Neuroimaging Database (Aliko et al., 2020). We then used partial least squares regression (Wold et al., 2001) to obtain regression coefficients (beta estimates) for each subject for the encoding models, with the time-matched features from the movie as the predictor variable and the observed BOLD activations masked by the voxels of the amygdala as the outcome variable.

We fit separate models for each participant and used 5-fold cross validation to estimate model performance. Voxel-wise performance was computed as the correlation between predicted activations from the encoding models and the observed activations from each subject's BOLD data. We also validated the predicted activations from our encoding models using naturalistic images from standardized affective image databases, the International Affective Picture System (Bradley & Lang, 2007) and the Open Affective Standardized Image Set (Kurdi et al., 2017). We predicted that encoding models would respond more

strongly to valenced images, as is commonly observed in human amygdala responses (Lindquist et al., 2016).

Generating Artificial Stimuli that Engage Specific Regions of Interest

Because the amygdala is not a single functional unit, we posited that our amygdala encoding models will have some degree of functional selectivity. Accordingly, we generated artificial stimuli that are optimized for the encoding models of different regions and subregions of interest. Using methods previously demonstrated in the visual cortex (Bashivan et al., 2019; Nguyen et al., 2016; Wang & Ponce, 2022), we used a deep generator network trained on ImageNet (Nguyen et al., 2016) and aimed to maximize activation in the amygdala encoding models. Responses spanning the entire amygdala (252 voxels) and amygdala subregions (the basolateral complex (LB), the centromedial nucleus (CM), the superficial (SF) group, and the amygdalostriatal transition zone (AStr); 29 to 178 voxels) were selected as the objective for activation maximization. Stimuli were also generated for encoding models of the visual cortex (VC; V1-V3) and inferotemporal cortex (IT) for each subject as controls. Optimization was performed using an evolutionary algorithm (Wang & Ponce, 2022), and artificial stimuli were generated with a random starting seed for each image.

Results

Human Amygdala Activity Encodes Valence

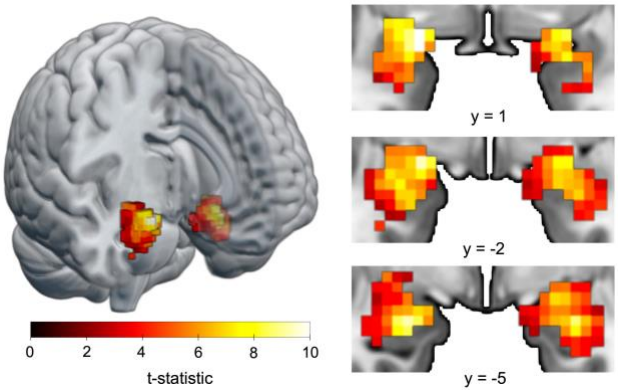

Figure 1: Predicted activation of amygdala voxels by encoding models; FDR threshold $q < .05$.

We found that the encoding models robustly predict amygdala responses to naturalistic stimuli. Voxel-wise t-tests showed that the mean performance of encoding models was well above chance in the amygdala (Figure 1). A mixed effects model revealed that predictions of the average amygdala response was above chance ($\hat{\beta}$ = .049, SE = .0053, t_{53} = 9.27, $p < .001$), and that there were marked differences in performance across amygdala subregions (ΔBIC = 23.5, Likelihood Ratio = 36.5, $p < .001$).

In response to naturalistic stimuli from standardized affective image databases, we found relationships between predicted activations from our amygdala encoding models and ratings of valence (IAPS: t = 2.86, p = .010, d = 0.64 and OASIS: t = 2.71, p = .014, d = 0.61) but not for arousal (IAPS: t = 1.35, p = .193, d = 0.30 and OASIS: t = -0.69, p = .496, d = -0.16) or the interaction between the two (IAPS: t = 1.10, p = .284, d = 0.25 and OASIS: t = 2.40, p = .027, d = 0.54), after controlling for low-level visual features.

Amygdala Encoding Models Capture Functionally Distinct Subregions.

We found that artificial stimuli selectively engaged encoding models of the targeted brain regions. A multi-way classification revealed that stimuli generated to activate each region and subregion were distinct from one another, with the exception of LB and SF (6-way accuracy = 71.7 ± 1.7%; Figure 2).

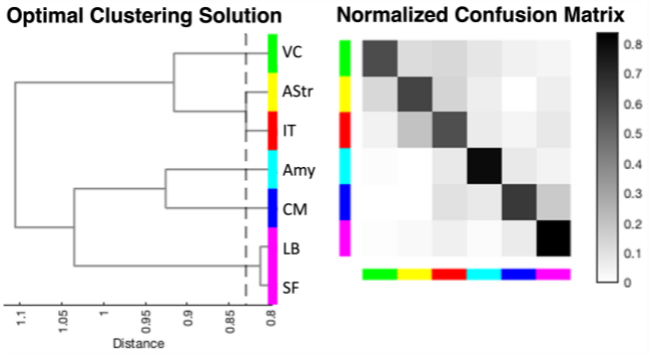

Figure 2: Hierarchical clustering and confusion matrix resulting from a 7-way classification of predicted activations in response to artificial stimuli.

Discussion

Our results show that convolutional neural networks can be used as a proxy for the ventral visual stream to accurately model and predict activity in the amygdala and its subregions in response to complex visual stimuli. Different amygdala subregions encode distinct sets of visual features, which can be recombined to predict variation in self-reported affect. These findings suggest that the amygdala functions to transform features extracted by the ventral visual stream to produce representations that predict the valence of emotional experiences.

References

Aliko, S., Huang, J., Gheorghiu, F., Meliss, S., & Skipper, J. I. (2020). A naturalistic

neuroimaging database for understanding the brain using ecological stimuli. *Scientific Data*, *7*(1), 347.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science (New York, N.Y.)*, *364*(6439), eaav9436.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Picture System (IAPS) in the study of emotion and attention. In *Handbook of emotion elicitation and assessment* (pp. 29–46). Oxford University Press.

Janak, P. H., & Tye, K. M. (2015). From circuits to behaviour in the amygdala. *Nature*, *517*(7534), 284–292.

Kragel, P. A., Reddan, M. C., LaBar, K. S., & Wager, T. D. (2019). Emotion schemas are embedded in the human visual system. *Science Advances*, *5*(7), eaaw4358.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. h

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*, *49*(2), 457–470.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The Brain Basis of Positive and Negative Affect: Evidence from a Meta-Analysis of the Human Neuroimaging Literature. *Cerebral Cortex*, *26*(5), 1910–1922.

Nguyen, A., Dosovitskiy, A., Yosinski, J., Brox, T., & Clune, J. (2016). *Synthesizing the preferred inputs for neurons in neural networks via deep generator networks* (arXiv:1605.09304). arXiv.

Ohman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews. Neuroscience*, *11*(11), 773–783.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences*, *14*(4), 303–316.

Wang, B., & Ponce, C. R. (2022). High-performance Evolutionary Algorithms for Online Neuron Control. *Proceedings of the Genetic and Evolutionary Computation Conference*, 1308–1316.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130.

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624.