Unconscious Perception of Scenes Reveals a Perceptual Neural Signature of Memorability

Yalda Mohsenzadeh (yalda@mit.edu)
McGovern Institute for Brain Research, MIT, 77 Massachusetts Ave
Cambridge, MA 02139, United States

Aude Oliva (oliva@mit.edu)
Computer Science and Artificial Intelligence Lab, MIT, 77 Massachusetts Ave
Cambridge, MA 02139, United States

Dimitrios Pantazis (pantazis@mit.edu)
McGovern Institute for Brain Research, MIT, 77 Massachusetts Ave
Cambridge, MA 02139, United States

Abstract:

While certain images are consistently remembered and others forgotten irrespective of subjective experiences, the spatiotemporal neural circuitry subserving this behavior is yet to be understood. Using MEG data collected while participants viewed a highly rapid serial visual presentation of images, we find that memorable images have a more sustained and stronger perceptual neural signal even though memory encoding was completely masked. Revealing the underlying visual features and neuronal processes that determine the memory fate of a stimulus can guide machine learning to evolve into more human-like performance.

Keywords: image memorability; perception; rapid serial visual presentation; visual masking

Introduction

While computer vision models have approached human performance in many tasks, they crucially fail in many domains. Given the growing overlap between computer vision and neuroscience, it has become increasingly relevant to study the neural mechanisms that afford incredible efficiency and accuracy to the human visual brain and inform the design of machine vision.

Such effort critically depends on understanding the neuronal processes that give advantage to some visual features over others, and ultimately determine the fate of a stimulus in the brain. Indeed, the human visual system does not treat all images equally; recent behavioral studies have shown that certain images are consistently remembered whereas others forgotten irrespective of subjective experiences (Isola et al., 2011). Even though this intrinsic property of visual stimuli, termed memorability, is highly replicable across individuals, the neural circuitry subserving this behavior is yet to be understood (Bainbridge et al., 2017). Here, we aimed to determine a lower boundary in the visual hierarchy for memorability-related neural signals. That is, we investigated how early in vision neuronal signals encode memorability information. Revealing the underlying visual features and neuronal processes that determine the memory fate of a stimulus can guide machine learning to evolve into more human-like performance.

Experimental design and stimuli set

To capture neuronal traces of memorability while controlling for memory effects, we used a rapid serial visual presentation (RSVP) paradigm. Each trial included presentation of 11 images with high rate of 34ms per picture. A target image, which was sampled from a set of 30 face images and 30 scene images (half with high memorability and half with low memorability scores), was embedded in the middle of the sequence. The preceding and following images were randomly sampled from 150 scene images with mid-level memorability scores (Figure1).

We collected MEG data while 15 participants viewed these RSVP streams and performed a two-alternative forced choice face detection task. High and low memorable target images were controlled for low level features (spatial frequency, color, luminance, and brightness) as well as high level categorical semantics. In order to test for memory encoding during the RSVP task, subjects were given an unanticipated memory test where target images were mixed with a novel matching set of images, and asked to report if they have seen the images in the RSVP streams with a confidence level from 1 (confidently seen) to 4 (confidently unseen).

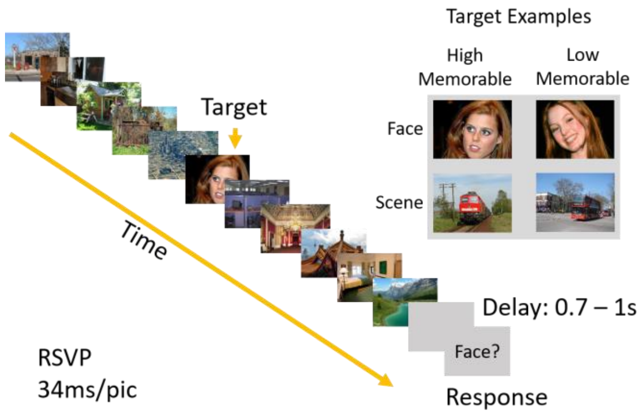

Figure 1: Experiment design. Each trial presented a rapid sequence of 11 images at a rate of 34 ms/pic. The middle image (target) was sampled randomly from a pool of 60 images (15 high and 15 low memorable faces, similarly 15 high and 15 low memorable scenes). Following a delay of 0.7-1s, participants were prompted to respond if the sequence included a face image.

Results

The behavioral performance in the subsequent memory test revealed that attended faces were remembered with a significant advantage for memorable face images ($p<<0.01$). In contrast, the report for unattended scene target images were at chance level, implying no memory encoding for unattended scenes. To study memorability without memory encoding confounds, we next focused on the scene target images.

We performed multivariate pattern analysis to extract information from MEG data (Krirgeskorte et al. 2008). The 306 MEG sensor data at each time point was arranged in pattern vectors. We used these vectors to train a SVM classifier to decode each pair of target images with a leave-one-out cross-validation procedure. The SVM classification accuracies for each pair of target images was used to populate a representational dissimilarity matrix with rows and columns indexed by the target images.

Figure 2 shows the decoding time series averaged within high memorable images in red and within low memorable images in black at each time point. MEG decoding results were similar at early time points (consistent with controlling for low level features) but high memorable images had an advantage with higher decoding starting at 149ms, until 228 ms.

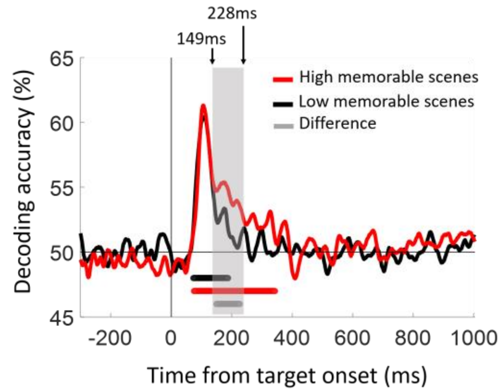

Figure 2: Time course of memorability decoding. Decoding of within high memorable and within low memorable pairs of scene images. Decoding accuracy was identical for high-memorable and low-memorable scene images early, but became consistently higher for high memorable scenes after 149 ms from target onset.

Conclusion

Our results indicated the RSVP paradigm was effective in revealing memorability-related signals early on, offering corroborating evidence that the memory fate of a stimulus is decided early in perception. Identifying the visual and neural features underlying this behavior will guide the design of vision models with neural-weights that stir closer to human-like performance.

Acknowledgments

Funded by a McGovern Institute Neurotechnology Program award and a NSF award 1532591 to D.P and A.O, and conducted at the Athinoula A. Martinos Imaging Center, MIBR, and MIT.

References

Isola, P., Parikh, D., Torralba, A. & Oliva, A. (2011) Understanding the intrinsic memorability of images. In Advances in Neural Information Processing Systems 2429–2437.

Bainbridge, W. A., Dilks, D., & Oliva, A. (2017). Memorability: A Stimulus-Driven Perceptual Neural Signature Distinctive from Memory. NeuroImage 149, 141-152.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational Similarity Analysis - Connecting the Branches of Systems Neuroscience. Frontiers in Systems Neuroscience, 2(November), 4-4. doi:10.3389/neuro.06.004.2008.