Presenter: Kyo Iigaya

Title: Computational Principles of Value Construction

Author(s): Kyo Iigaya, Sanghyun Yi, Iman A. Wahle, Koranis Tanwisuth, John P. O'Doherty

Abstract: It is an open question how humans and other animals flexibly construct the value of stimuli, even for stimuli never encountered before. While great progress has been made toward understanding how the brain adjusts the value of stimuli through reinforcement-learning, little is known about how stimulus value arises in the brain in the first place. Here, we propose and provide evidence that the brain constructs the value of a novel stimulus by extracting and assembling common features. Notably, because those features are shared across a broad range of stimuli, we show that simple linear regression in the feature space can work as a single neural mechanism to construct the value across stimulus domains. In large-scale behavioral experiments with human participants, we show that a simple model of feature abstraction and linear summation can predict the subjective value of paintings, photographs, as well as shopping items whose values change according to different goals. The model shows a remarkable generalization across stimulus types and participants, e.g. when trained on liking ratings for photographs, the model successfully predicts a completely different set of art painting ratings, or when trained on clothing items for one gender, the model successfully predicts ratings in the other gender across contexts. Also, we show that these general features emerge through image recognitions training in a deep convolutional neural network, without explicit training on the features, suggesting that features relevant for value computation arise through natural experience. Furthermore, using fMRI, we found evidence that the brain actually performs value computation hierarchically by transforming low-level visual features into high-level abstract features which in turn are transformed into valuation. We conclude the feature-based value computation is a general neural principle enabling us to make flexible and reliable value computations for a wide range of stimuli.