University of Zurich[UZH]

Department of Economics – Neuroeconomics Seminar

November 29, 2018 - 17:00 - 18:00

Valentin Wyart
Ecole Normale Superiore

Curious by choice or by chance? Computational noise in reward-guided learning drives behavioral variability in volatile environments

When learning the value of actions in volatile environments, humans often make seemingly irrational decisions which fail to maximize expected value. Prominent theories describe these 'non-greedy' decisions as the result of a compromise between choosing a currently well-valued action vs. exploring more uncertain, possibly better-valued actions – known as the 'exploration-exploitation' trade-off. However, we have recently shown that the variability of human decisions based on multiple ambiguous cues is bounded not by choice stochasticity, but rather by computational noise in the underlying inference process. We thus reasoned that non-greedy decisions may be caused to some extent by the same kind of noise during reward-guided learning. We derived a theoretical formulation of reinforcement learning (RL) which allows for random noise in its core computations. In a series of behavioral, functional neuroimaging and pupillometric experiments, tested over a total of 90 human participants in a canonical restless bandit task, we quantified the fraction of non-greedy decisions driven by learning noise and identified its neurophysiological substrates. At the behavioral level, we show that more than half of non-greedy decisions are triggered by learning noise alone. By describing the consistency of human decisions across repetitions of the same sequence of rewards in terms of a 'bias-variance' trade-off, we rule out the possibility that learning noise is due to a misspecification of our RL framework. At the neurophysiological level, the trial-to-trial variability of sequential learning steps and their impact on behavior could be predicted both by BOLD responses in the dorsal anterior cingulate cortex (dACC) and by phasic pupillary dilation – suggestive of neuromodulatory fluctuations driven by the locus coeruleus-norepinephrine (LC-NE) system. Together, these findings indicate that most of the behavioral variability observed in volatile environments is due to the limited computational precision of reward-guided learning.