Klaus-Robert Müller

Title: Explainable AI for Obtaining Insight

Abstract: The talk will briefly introduce explainable AI (XAI) and its applications in Neuroscience and beyond. Given a model that has learned to accurately predict, XAI allows to extract insight from the model, even if it is nonlinear like deep learning architectures. Moreover, XAI allows to detect modeling artifacts, bugs or artifactual data thus contributing to a more accurate modeling process.

Karl Friston

Title: The Physics of Sentience

Abstract: How can we understand ourselves as sentient creatures? This presentation uses the free energy principle to furnish an account in terms of active inference. First, we will try to understand sentience from the point of view of physics; in particular, the properties that self-organizing systems—that distinguish themselves from their world—must possess. We then rehearse the same story from the point of view of a neurobiologist, trying to understand functional brain architectures. The narrative starts by casting life—or self-organization—as an emergent property of any dynamical system that possesses a Markov blanket. This conclusion is based on the following arguments: if a system can be differentiated from its external milieu, then its internal and external states must be conditionally independent. Crucially, this equips internal states with an information geometry, pertaining to probabilistic beliefs about external states. In short, internal states will appear to infer—and act on—their world to preserve their integrity. This leads to a Bayesian mechanics, which can be neatly summarized as self-evidencing. In the second half of the talk, we will unpack these ideas using simulations of Bayesian belief updating in the brain and relate them to predictive processing, sentient behavior and the emergence of agency.

Key words: *active inference · autopoiesis · cognitive · dynamics · free energy · affordance · self-organization · agency.*

Yael Niv

Title: Latent Cause Inference and Mental Health

Abstract: No two events are alike. But still, we learn, which means that we implicitly decide what events are similar enough that experience with one can inform us about what to do in another. We have suggested that this relies on parsing of incoming information into “clusters” according to inferred hidden (latent) causes. In this talk, I will present this theory and illustrate its breadth in explaining human learning. I will then suggest some ways in which latent cause inference can help understand mental health conditions and their treatment.

Christof Koch

Title: Consciousness and its Physical Substrate – Progress and Problems

Abstract: tba

Dora Angelaki

Title: Belief embodiment through eye movements facilitates memory-guided navigation

Abstract: Natural behavior is flexible and supported by abstracted away beliefs. Over the course of evolution, the brain evolved complex recurrent networks to interpret and act upon a dynamic and uncertain world, but its computational powers and mechanisms generating natural behavior remain largely a mystery. Most of our insights into neural computation are based on binary tasks with highly constrained actions that are artificially segregated from perception. Tightly controlling laboratory behavior by preventing natural, continuous movements has simplified interpretability but also hindered our ability to gain insights from natural behavioral strategies. Artificially keeping the eyes fixed, for example, has been standard in monkey studies of working memory and decision-making

To understand dynamic neural processing underlying natural behaviour, we use continuous-time foraging tasks in virtual reality. Although task rules do not require any particular eye movement, we find that where subjects look is an important component of the behavior. For example, during a simple task in which humans use a joystick to steer and catch flashing fireflies in a virtual visual environment lacking position cues, subjects physically track the latent task variable with their gaze. Restraining eye movements worsened task performance suggesting that embodiment plays a computational role. We show this strategy to be true also during an inertial version of the task in the absence of optic flow and demonstrate that these task-relevant eye movements reflect an embodiment of the subjects' dynamically evolving internal beliefs about the goal. A neural network model with tuned recurrent connectivity between oculomotor and evidence-integrating parietal circuits accounted for this behavioral strategy. Critically, this model better explained neural data from monkeys' posterior parietal cortex compared to task-optimized models unconstrained by such an oculomotor-based cognitive strategy. These results highlight the importance of unconstrained movement in working memory computations and establish a functional significance of oculomotor signals for evidence-integration and navigation computations via embodied cognition.