Results report

1. Title of Research and Development: Testing computational models of learning from social, real, and fictive feedback in human and nonhuman primates.

2. Principal Investigator: Masaki Isoda (Associate Professor; Department of Physiology, Kansai Medical University School of Medicine)

3. Counterpart Principal investigator: Markus Ullsperger (Professor; Department of Neuropsychology, Faculty of Natural Sciences, Otto-von-Guericke Universität Magdeburg (Germany))

4. Results of Research and Development:

The goal of this consortium is to develop computational models of learning and decision making and to test them in two biological systems, humans and macaques. In fiscal year 2015, the Japanese team studied activity of single dopamine neurons in the midbrain of monkeys performing a socially-oriented, Pavlovian conditioning paradigm, which allowed us to study the neural mechanisms underlying learning from real and social feedback. In the paradigm, two monkeys facing each other were presented with a conditioned stimulus (CS) that differently signaled the probability of each monkey's upcoming reward. The reward feedback was given first to the nonrecorded animal (designated as *other*) and, 1 s later, to the recorded animal (designated as *self*). The monkeys were alternately conditioned in two contextually different trial blocks: i.e., one block in which the other's-reward probability was variable depending on the CS and the self-reward probability was constant, and another block in which the self-reward probability was now variable and the other's-reward probability was constant. Consistent with previous studies, the animals placed higher value on the CS predicting a higher probability of *one's own* reward. Interestingly, the animals placed lower value on the CS predicting a higher probability of *other's* reward. Thus, the prospect of others' reward profoundly affected the value of one's own reward. Single-unit recording was then made from presumed dopamine neurons (n = 229) in the substantia nigra pars compacta and the ventral tegmental area. About half of the neurons (n = 115) showed a significant positive correlation between the phasic CS response and the self-reward probability. Among them, the activity of 80 neurons was not influenced by the other's-reward probability, while that of 35 neurons additionally showed a significant negative correlation with the other's-reward probability as if representing the subjective value. There were few neurons that selectively encoded the other's-reward probability. Crucially, however, dopamine neurons as a population became more discriminative for the other's-reward probabilities as they discriminated more clearly between the self-reward probabilities. These findings suggest that dopamine neurons incorporate self-reward information and other's-reward information in a single value scale. This is in sharp contrast to firing properties of medial prefrontal neurons, which distinctly encoded the self-reward information or other's-reward information. Our data demonstrate that midbrain dopamine neurons and medial prefrontal neurons participate in different aspects of learning from real and social feedback.

Meanwhile the German team developed a social instrumental learning task enabling to compare learning from experienced outcomes of own actions and observational learning from outcomes observed for actions taken by another player. With appropriate simplification this task can be easily transferred to a setting allowing to study observational learning in macaques. Participants were required to learn the stimulus-outcome contingencies for an adapted three-armed bandit task. In each block, their aim was to learn over the course of 30 trials, which of three stimuli would most often lead to a positive outcome. The participants would play together, with the acting and observing player switching roles regularly. A first task version using blockwise role switches successfully demonstrated the ability of participants to learn from own action outcomes as well as from observation. However, the block structure caused difficulties in modelling the behaviour with reinforcement learning models. Therefore, role switching every 1-3 trials was introduced to engage the participant during the observation stage more fully than the previous task structure. It was also expected to ease modelling and be more suitable for a follow-up experiment using fMRI than the former design. Sixty-four channel EEG was recorded from both participants as they performed the task. An adapted Q-learning algorithm was fit to participants' choices in this task. Comparable model estimated learning rates were obtained for trials in which the same player acted consecutively, relative to when players switched from an observing to an acting role. This suggested that participants used similar computational algorithms to weight the outcomes they received from making and observing choices on each trial. Model-based single-trial EEG analysis using multiple robust regression revealed that reward prediction errors were represented in the feedback-related negativity (FRN) around 280 ms after the feedback. Furthermore, P3a and P3b amplitudes correlated with the experienced or observed outcomes. The role (actor/observer) had a sustained positive-going effect at centroparietal electrodes from 200-600 ms such that experienced outcomes elicited a generally more positive event-related potential. Moreover, the factors role and prediction error interacted significantly at the latencies of the FRN and P3b, such that both potentials were larger when participants experienced outcomes of their own actions. These findings suggest that while the general learning mechanisms are similar for own and observed action outcomes, own experiences have a stronger impact on the participants. Based on these findings the computational model will be modified to allow differential learning rates for own and observed action outcomes. In summary, the newly developed task has proven well-suited to study observational learning and can now be transferred to animal research and functional magnetic resonance imaging in humans. Moreover, it will be modified to investigate the influence of social context, for example competitive vs. cooperative settings.