A possible function of anterior prefrontal cortex in a model-based reinforcement learning; a computational model and an fMRI study.

Wako Yoshida, and Shin Ishii

Background Natural environments surrounding humans include unobservable (hidden) states and dynamically change with time. Even in such complicated environments, humans learn the characteristics of the current environment and determine their optimal behaviors. Behavioral adaptation to an environment can be formulated as an optimal decision making problem with an on-line identification of the current environment; the environment identification involves the estimation of hidden states. In a machine learning field, the optimal decision making problem is often termed Markov decision process (MDP); reinforcement learning (RL) is a framework to deal with MDPs. If an environment involves hidden states, it is formulated as a partially-observable MDP (POMDP). In our previous paper (Ishii, Yoshida, & Yoshimoto, *Neural Networks*, 2002), we presented an RL method for POMDPs, in which hidden states were estimated based on an on-line Bayes identification of the environment. We also proposed a possible functional model in the brain for our RL method, in which the major parts of the RL scheme were involved in functions of the prefrontal cortex. According to our theoretical model, the dorsolateral prefrontal cortex (DLPF) executes maintenance and manipulation of the reward-based environmental model and action selection depending on the environmental model is done within the anterior cingulate cortex (ACC). We also suggested that the anterior prefrontal cortex (APF) is related to the estimation of hidden states. In our later study (Yoshida, & Ishii, to appear), we conducted a human imaging experiment during an MDP task, by using functional magnetic resonance imaging (fMRI). In comparison to a simple memory task, significant activations during the MDP task were observed in DLPF and ACC.

Aim of this study Although our previous experimental finding was consistent with our theoretical model, the role of APF has not yet to be examined. Using a branching task, in which the maintenance of a primary task was necessary while performing a subtask, Koechlin et al. (2000) showed that APF was activated when a subject could not predict whether the forthcoming task would be the primary task or the subtask. This result suggested that APF is involved in active switching of behavioral rules without explicit cues. Since

such switching is induced by the estimation of environmental change, we assume that APF is related to the estimation of hidden states. In this study, we identify the neural correlates to the estimation of hidden states using fMRI.

Method Normal human subjects were required to perform a three-dimensional wire frame maze displayed on the screen. The objective of a subject was to reach a goal as many times as possible within a limited number of total actions, and he/she was paid in proportion to the number of goal arrival. Before a task, a two-dimensional instruction map was displayed on the screen, and a subject was requested to memorize the spatial extent of the maze. At each step in a task, a wire frame of the current maze state was displayed and a subject was to choose an action among a forward step, a backward step, a turn right and a turn left. Since the view is limited, a subject was not necessarily able to recognize where the current position on the maze was. An experiment consisted of two task conditions. In a start-goal (SG) condition, on one hand, a subject was instructed on a two-dimensional maze indicating both of start and goal positions. If a subject successfully memorized the maze map, he/she could trace the shortest path, because the start and the goal were known. In a goal (G) condition, on the other hand, only a goal position was instructed and a start position was unknown. In this task, a subject had no clues where he/she was at the beginning of the task, and needed to explore the maze in order to know the current position, i.e., the hidden state, based on past experiences.

We used a 1.5-tesla MR scanner; functional images were obtained with a T2*-weighted EPIs, with BOLD contrasts. The volumes were acquired every 2.5sec (TR), and contained 23 slices of 5mm thickness. Data were preprocessed and analyzed with the SPM99 software.

Result BOLD activations during the G condition and the SG condition were statistically compared with respect to group random effects with a significance level $p < 0.005$. We found significant activations in APF (BA10) when subjects were conducting the SG condition.

Conclusion The maze task in the SG condition is a typical example of POMDPs, in which the environment is known but the optimal behavior cannot be determined due to the partial observability. During an SG task, a subject was required to estimate the hidden state of the environment, in order to achieve the task. Our imaging study have thus revealed that APF is related to resolving the partial observability of an POMDP environment. This result is also consistent with our theoretical model (Ishii, Yoshida, & Yoshimoto, 2002).