The role of adaptive decision noise in exploration

Robert C. Wilson, John M. White and Jonathan D. Cohen

Everyone at the SfN conference faces the explore-exploit dilemma: do you go see posters from labs you know well (exploit) or wander aimlessly down row BBB in search of something new (explore)? Exploiting is the best way to get immediate reward, but if you don't explore you might miss out on the latest advances. Solving this problem optimally is intractable in all but the simplest settings, and so the question arises as to how humans balance exploration and exploitation in practice.

In machine learning, engineers have made great use of noise as a tool for driving exploration. This strategy works by injecting randomness into the decision process. Thus, most of the time (when the noise term is small) algorithms maximize immediate reward by exploiting, while some of the time (when the noise term is large) exploration occurs by chance. The level of the decision noise reflects the degree to which exploration or exploitation is favored.

We have previously shown that, in a very simple problem, humans adapt their decision noise in a way that is consistent with such a random-exploration strategy [1]. In the present work we tested whether this was also true in a more complicated experiment in which the potential gains for exploring were higher. In particular, we investigated the explore-exploit tradeoff in a world characterized by abrupt and unsignaled change-points. By manipulating the frequency with which change-points occurred, we were able to alter the optimal balance between exploration and exploitation and hence the optimal setting of the noise.

In our change-point task, participants made a series of choices between two options. Every time an option was chosen it paid out a reward between 0 and 100 points. The reward available for each option was constant over time except at a change-point, when it was randomly reset between 0 and 100. Because the current reward value of each option was only shown when that option was played, the longer an option remained unplayed, the more ambiguous it became, as the probability that a change-point had occurred increased.

We modeled human decisions using a simple choice rule based on the observed outcome for each option, the information available for playing it and the level of decision noise. As change-points became more frequent we found systematic increases in the decision noise. A separate analysis showed that this qualitative pattern of increasing decision noise with hazard rate is optimal in this task.

These results suggest that humans both use and adapt their decision noise to effectively manage the explore-exploit tradeoff in complex tasks.

[1] Wilson et al. Program No. 830.13. Society for Neuroscience, 2011.