Collective behavior of social learning agents in restless multi-armed bandit

Kazuaki Nakayama*[1] and Shintaro Mori[2],

[1] Department of Mathematical Sciences, Faculty of Science, Shinshu University
Asahi 3-1-1, Matsumoto, Nagano 390-8621, JAPAN
[2] Department of Physics, Faculty of Science, Kitasato University
Kitasato 1-15-1, Sagamihara, Kanagawa 252-0373, JAPAN

E-mail: [1]nakayama@math.shinshu-u.ac.jp [2]mori@sci.kitasato-u.ac.jp

A restless multi-armed bandit is a typical model for studying the properties of social learning[1]. We use a bandit having 1 good arm with pay-off 1 and infinitely many bad arms with pay-off 0. Good lever turns into a bad one with a probability p_c and a new good lever appears. There are N_1 agents with probabilities (p_{O1}, q_{I1}) of Observe and Innovate, and an agent with probabilities (p_{O2}, q_{I2}). It is shown that, in order for the second agent to maximize his score, he should adjust his p_{O2} so that $dp_{O1} \cdot dp_{O2} < 0$. Surprisingly, it seems that this is true irrespective of p_c.

References

[1] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, M. W. Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, K. N. Laland, "Why Copy Others? Insights from the Social Learning Strategies Tournament", Science, vol 328, 208-213 (2010).