

遅延報酬からの行動学習の脳機構

Brain mechanisms for action learning from delayed rewards

銅谷賢治(沖縄科学技術大学院大学先行研究)

ある行動により報酬が得られたらその行動を強化, つまり選択確率を高める. これは行動学習の基本だが, ではある行動が複数のステップや時間遅れを経てようやく報酬が得られる場合はどうだろう? 動物や人間の複雑な運動や社会行動などは, 遅延報酬からの学習なしには考えられない. 遅延報酬からの行動学習のための計算理論である強化学習の枠組みから,

- 1) 現在の状況や行動から将来得られる報酬を予測すること,
- 2) 得られた報酬を過去の状態や行動の履歴と結びつけること,
- 3) 予測や履歴の時間スケールを環境や課題に応じて適宜調整すること,

が重要な計算要素と考えられる. われわれは, これらにおける大脳基底核を中心とした脳回路と, ドーパミンやセロトニンなどの神経修飾物質の役割の解明に向け, 脳活動計測と神経生理学実験を行ってきた. その結果は, 大脳皮質-基底核回路には短期的/長期的な報酬予測に関わる並列的な回路が存在し, それらの働きをセロトニンが差動的に調整することにより, 目先の結果と長期的な最適性に向けた行動と学習の適応的な制御を行っていることを示唆している.