題目
リスク下における適応的な強化学習の進化：進化シミュレーションによる理論的検討

氏名　本間祥吾

指導教員　竹澤正哲

生物は、経験に基づいて自身の行動を変容させるプロセス：学習によって、自身の生存可能性を高めてきた。多くの実証研究から、学習メカニズムは遺伝的に制約を受け、進化によって形成されるということが示唆されている。しかし、多くの生物に広く見られる強化学習という学習メカニズムの進化を理論的に検討した先行研究は少ない。本研究は、リスク（選択によって得られる利得のばらつき）という要因によって、強化学習を制御するパラメータがどのような値に進化するのかを進化シミュレーションによって検討した。

シミュレーションでは、個体が Q 学習アルゴリズムにしたがって学習を進めながら、two-armed bandit 課題を行った。2 つの選択肢は期待値と標準偏差（リスク）が異なる 2 つの正規分布が設定された。Q 学習は、正の学習率（ある選択をして予測より良い結果を得た時に、その選択をより追求する傾向）、負の学習率（ある選択をして予測より悪い結果を得た時に、その選択を回避する傾向）、逆温度の 3 つのパラメータを遺伝子として持っていた。先行研究である Niv et al. (2012) から、2 つの学習率は対称的な進化を示すことが予測された。

シミュレーションの結果、負の学習率と正の学習率では非対称な進化が見られた。負の学習率は、課題が、リスク回避が有利なのかリスク追求が有利なのかによって、進化の方向が一貫していた。一方、正の学習率は、そのような課題の性質によっては一貫しない複雑な進化が見られた。以上の結果は、負の学習率と正の学習率は、分布の異なる側面に反応する非対称的な機能を持つことを示唆している。また、負の学習率は正の学習率よりも一貫して強い淘汰を受ける傾向が見られた。これより、リスク下の意思決定において進化的に重要なのは、予測より良い結果を得た時の反応ではなく、予測より悪い結果を得た時の反応であることが示唆される。今後は、学習率の非対称性を生む具体的なプロセスや、その進化の方向を予測する課題の特徴を定量的に検討する必要があるだろう。また、リスクだけではなく、より生態学的に妥当で、学習率の淘汰圧になると考えられる変動性（分布の期待値の変化）を導入したシミュレーションを実行し、パラメータが持つ機能をより詳細に検討する必要がある。本研究のような進化的アプローチによって、生物の多様な行動の背後にある学習の至近要因・究極要因の双方が明らかになることが望まれる。