学位論文内容の要約

博士の専攻分野の名称：博士（人間科学）　　氏名：　本　間　祥　吾

学位論文題名
進化・学習・規範：強化学習の計算論モデルによる検討

動物は絶えず変化する不確実な自然環境に生きている。中でも人間は、その進化史において特に多様で大きな不確実性を持つ環境を経験し適応してきた。人間は2つのレベルの適応によって、不確実な自然環境に適応してきたと考えられる。第 1 に、学習を通した「個体レベルの適応」である。人間を含めた動物は不確実な自然環境において、学習メカニズムによって生存に役立つ行動を獲得してきた。第 2 に、文化や社会規範を通した「社会レベルの適応」である。人間は、社会的学習を介した文化・情報の伝達、社会規範による協力・資源分配によって、集団として不確実性に対処し生存可能性を高めてきた。近年、社会神経科学や実験社会科学といった分野の知見から、文化や協力、規範といった社会レベルの適応を支えるメカニズムに学習が関与していることが明らかとなっている。強化学習は、不確実な環境において試行錯誤を通して最も大きな報酬をもたらす行動を獲得することを可能とした、動物の適応を達成する基礎的なメカニズムである。本論文は、強化学習を用いた計算論モデリングによって、この 2 つの適応のあり方を統合的に理解することを目的とする。本論文は 2 つの部で構成される。第 1 部では、個体レベルの適応に焦点を当て、生物進化のプロセスによってどのように学習が進化し形成されるのかが検討された。具体的には、強化学習アルゴリズムが進化によって最適化されるという仮説のもと、不確実性下で適応的な行動を効率的に獲得するためにどのような強化学習が進化するのかがコンピュータ・シミュレーションにより検討された。第 2 部では、人間の社会レベルを支える心的メカニズムとして規範の内面化と向社会性に着目し、強化学習がこれらのメカニズムにどのように関与しているのかという問いが行動実験により検討された。

第 1 章では、不確実性に対する適応という観点から個体レベルの適応と社会レベルの適応のあり方を概観した。そして、第 1 部と第 2 部のそれぞれで取り組む 2 つの異なる問いを整理した。動物は、進化と学習という 2 つのプロセスにより適応を達成する。両者は相互作用しながら動物の適応を加速させる（e.g., ボールドウィン効果）。進化は何が学習可能であるかに関する制約を生み出し、不確実な環境で適応的な行動を効率的に獲得できるような学習バイアスを生み出す。こうした背景のもと、第 1 部（第 2 章と第 3 章から構成される）では不確実性下の意思決定において強化学習の学習バイアスがどのように進化するのかを検討することが解説された。さらに、近年、規範の内面化や向社会性といった、人間の社会レベルの適応を支えるメカニズムには報酬から学習するメカニズム、すなわち強化学習が関与している可能性が指摘されている。第 1 章では、報酬の処理に関わる脳領域が社会的意思決定に関与しているという神経科学分野の豊富な知見を概観した上で、強化学習がこれらのメカニズムとどのように関係しているかについて実証的知見は少ないことを主張した。このような背景のもと、第 2 部（第 4 章と第 5 章から構成される）では、規範の内面化・向社会性の基盤として強化学習アルゴリズムがどのように関与しているのかを、行動実験と強化学習モデルを用いた分析により検討することが解説された。

第 2 章では、不確実性のうちリスク（選択肢の分布の分散）に着目し、多様なリスク状況下における適応的な強化学習の進化を検討した。個体が適応的に行動するためには、選択肢の期待値やリスクがわからない中で、リスクの大小に関わらず期待値の高い行動を学習によって選択する必要がある。この時、どのような学習バイアスを持つことが進化適応的なのだろうか。第 2 章では、個体が一生の間に複数のリスク状況を経験する場合に、強化学習アルゴリズムのパラメータである学習率と逆温度がどのように進化するのかを進化シミュレーションにより検討した。特に、正と負の学習率（それぞれ予期せぬ報酬・罰に対する反応を表す）の大小はリスク下の行動と関

連することが知られており、適応において重要な学習パラメータである。シミュレーションの結果、負の学習率が減少し、正の学習率が負の学習率が大きいという関係が進化した。また、進化した個体群はリスクの大小に関わらず、期待値の高い行動を獲得できるようになった。さらに、進化した個体群はプロスペクト理論的な行動も示した。本研究の結果は、獲得領域でリスク回避的、損失領域でリスク追求的な選択を行うという傾向が、多様なリスク環境における進化適応的な学習バイアスの産物として理解できることを示唆している。また、多くの先行研究で見出されてきたリスク選好の持つ文脈依存性を、進化適応的な学習を通して獲得された行動として理解可能であることも示唆される。

第 3 章では、変動性（選択肢の期待値の変化）に着目した。人間は進化史において特異な環境変動を経験した。こうした変動性は、人間の持つ高度な社会的学習能力に対する主要な淘汰圧であることが議論されている。第 3 章では、第 2 章のシミュレーションに変動性を加え、変動性の大きさに応じて強化学習のパラメータがどのように進化するかを検討した。その結果、変動性に関わらず負の学習率が減少するなど、第 2 章における複数のリスク状況を経験する場合と類似した進化が見られた。この結果から、リスクに対して適応的な学習システムは変動性にも適応的である可能性が示唆された。

第 4 章では、個人の罰に対する感受性と規範の内面化の関係が検討された。人間は外的な罰が存在しない状況でも自ら規範に従う。これは人間が社会化を通じて規範を内面化しているためであることが議論されてきた。しかし、そもそも規範の内面化がどのようなメカニズムによって生じているのかについては十分に検討されてこなかった。第 4 章では、規範の内面化が学習における罰に対する過剰な反応に支えられている可能性に着目し、規範の内面化の個人差が強化学習の罰に対する感受性（負の学習率パラメータ）によって説明されるかを検討した。実験では、規範の内面化の程度は質問紙で測定され、負の学習率は学習課題の行動データに強化学習モデルをフィッティングすることで推定された。また、内面化と関連すると考えられる利他性を経済ゲームによって測定した。その結果、負の学習率が高い個人ほど規範を強く内面化しているという関係は見られなかった。一方、探索的な分析の結果、負の学習率が高いほど経済ゲームで他者に利他的に分配するという傾向が見られた。

第 5 章では、第 4 章の結果を追試するために、負の学習率と向社会的な分配行動の関連が検討された。近年、リスク下の意思決定と分配の意思決定には共通の神経基盤が存在することが示されている。第 2 章と第 4 章の知見に基づくと、負の学習率、リスク回避傾向、他者に対する向社会的な分配の間には相互に関係が存在することが示唆される。第 5 章では、これらの 3 つの変数を行動実験により測定し、その関係を網羅的に検討した。その結果、負の学習率が大きいほど他者に対して向社会的に分配するという関係は再現されなかった。一方、探索的な分析の結果、正の学習率（報酬に対する感受性）が向社会的な分配、リスク追求傾向のそれぞれと関連を示した。加えて、リスク回避的なほど向社会的に分配するという傾向も観察された。

第 6 章の総合考察では、第 2 章から第 5 部までの結果が整理された。第 1 部の研究からは、現実の行動や学習を理解するために、学習アルゴリズムの進化というアプローチが有用である可能性が示された。第 2 部の研究からは、規範の内面化と強化学習の関係は明らかとはならなかった一方で、リスク下の意思決定と分配の意思決定が強化学習を介して繋がっている可能性が示唆された。加えて、第 1 部と第 2 部のアプローチの限界と展望を論じた。最後に、進化と学習の相互作用が人間に固有な認知メカニズムに関与しているという議論を展開した上で、進化シミュレーションと計算論モデリングが人間に固有な認知メカニズムを明らかにするのに有用なツールとなりうることが論じられた。