



# HOKKAIDO UNIVERSITY

Title	進化・学習・規範 : 強化学習の計算論モデルによる検討 [論文内容及び審査の要旨]
Author(s)	本間, 祥吾
Degree Grantor	北海道大学
Degree Name	博士(人間科学)
Dissertation Number	甲第15987号
Issue Date	2024-03-25
Doc URL	<a href="https://hdl.handle.net/2115/92338">https://hdl.handle.net/2115/92338</a>
Rights(URL)	<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>
Type	doctoral thesis
File Information	Shogo_Homma_abstract.pdf, 論文内容の要旨



# 学位論文内容の要旨

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

### 本論文の観点と方法

人間は、動的に変化する多様な環境に適応して進化してきた。人間は、自然環境への適応を 2 つの異なるレベルで達成してきたとされている。第 1 に、学習という仕組みを通じた個体レベル適応である。人間を含む生物は、多様な環境において生存に利する適応的な行動を学習というメカニズムによって獲得してきた。第 2 に、文化／規範という制度を通じた社会レベルの適応である。人間は社会的学習による情報共有、資源分配、集団での協力を通して、自然環境が持つ不確実性に対処してきた。近年、社会神経科学や実験社会科学、進化社会科学の研究を通じて、人間の協力や規範の根底に学習というメカニズムが深く関与していることが明らかにされつつある。本論文の目的は、強化学習という計算論的アルゴリズムを軸にして、2 つの異なるレベルの適応を統合的に理解していくことにある。本論文は二部構成からなる。第 1 部では、学習というメカニズムが生物進化のプロセスから進化してきたのかに関わる問題が検討された。具体的には、学習というメカニズムを強化学習という計算論的アルゴリズムによって定式化した上で、このアルゴリズムが多様で動的に変化する不確実な意思決定状況に対して、進化的にチューニングされているという仮説の下で、一連の進化シミュレーションを行った。第 2 部では、人間の協力行動を規定する基層的なメカニズムとして、規範内面化に着目し、強化学習という計算論的アルゴリズムとの関係を行動実験によって検討した。

### 本論文の内容

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を概観し、本論文で取り組む問いを整理した。動物は進化と学習という 2 つのプロセスを通して適応を達成しているが、ボールドウィン効果に代表されるように、両者は相互に影響を与えながら適応を加速させることが知られている。何が学習可能であるかは進化によって調整され、環境で適応的な行動を効率的に獲得できるように学習バイアスが形成されている。こうした進化と学習の関係について生物学を中心としてレビューした上で、本論文における 2 つの目標と構造が設定された。第 1 部（第 2 章および第 3 章からなる）では、不確実下での意思決定という生存に重要な課題において、強化学習という個体レベルのメカニズムの進化を検討すること、続く第 2 部（第 4 章および第 5 章からなる）では、規範内面化と向社会性の基盤として強化学習に注目することが解説された。

第 2 章では、人間が生存のために直面した不確実性をリスク（選択枝の分布の分散）として定義した上で、個体が複数のリスク状況を経験する場合の強化学習アルゴリズムの進化が検討された。適応的に振る舞うためには、リスクの大小に関わらず期待値の高い行動を学習する必要がある。具体的には強化学習アルゴリズムが持つ学習率と逆温度というパラメータが、多様な環境に適応して進化したという仮説を、複数の進化シミュレーションによって検討した。その結果、進化した個体群は、リスクの大小に関わらず期待値が高い行動を多様な環境下で学習できるようになること、さらにその結果として、進化した個体群はプロスペクト理論的な振る舞いを示すことが明らかとなった。一般的に、獲得領域ではリスク回避的、損失領域ではリスク追求的行動を示すことは、非合理的なバイアスだと解釈されがちだが、本章の結果は、現実のリスク選好に見られる文脈依存的なバイアスが、進化適応的な学習バイアスの産物として理解できる可能性を示唆するものである。

現世人類が進化した自然環境は、変動性という統計的性質において特異な特徴を持つことが指摘されており、人間固有の高度な社会的学習能力は、そうした統計的性質への適応であることが理論的に指摘されてきた。第3章では、第2章のシミュレーションに変動性（すなわち期待値の時間的変化）を導入した。進化シミュレーションの結果、変動性の大きさに関わらず、第2章で得られたのと酷似した個体群の進化が観察された。この結果は、不確実性下で適応的な学習メカニズムは、同時に変動性に対しても適応的とな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だが、規範の内面化がそもそもどのようなメカニズムに支えられているか、未解明の問題である。本章では、規範の内面化が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、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）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。結果、負の学習率が高い個人ほど規範を内面化しているという関係は見られなかったものの、負の学習率と向社会的行動の間に関連が見られた。

第5章では、第4章の追試を目的として、負の学習率と向社会的な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・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ものの、正の学習率（報酬に対する感受性）が向社会的な分配、リスク追求傾向と関連していること、さらに、リスク回避傾向が向社会的な分配と関連していることが見出された。

第6章は総合考察であり、まず第2章から第5章までの成果が批判的に検討／整理された。続いて第1部（第2章および第3章からなる）、第2部（第4章および第5章からなる）のそれぞれで採用されたアプローチの限界と展望が論じられた上で、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。最後に、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る可能性が論じられた。