学位報告 4

報告番号	※甲　　第　　号

主　論　文　の　要　旨

論文題目	Artificial life investigations into the evolutionary origins of representational cognition (表象的認知の進化的起源に関する人工生命研究)
氏　　名	ARNOLD Solvi Fylgja

論 文 内 容 の 要 旨

高等な知能には，外界をモデル化し，シミュレーションする能力が不可欠である．そういう能力を「心的表象」，表象を扱う知能を「表象的知能」という．表象的知能の起源を進化論で説明するのは一般には困難である．なぜならば，行為を動機づけるものである考えや感情は直接的には進化に影響を与えないからである．この事実は，生物の高等知能や古典的人工知能において表象が重要な役割を果たしているはずなのに，遺伝的アルゴリズムや機械学習などの自然を真似た手法を用いて作成された人工知能には表象性が欠けるという問題を人工知能や心の哲学にもたらす．この問題は，知能の進化に関する我々の理解に何かが欠けていることを示す．本論文では，自然における適応プロセスがいかにして表象的知能を生み出すかという問いに答える理論を提示している．そして，人工的に表象的知能を進化させる手法を開発した上で，その理論に対する実験的検討を行っている．このような研究動機や基本方針を述べた第１章に続いて，第２章では，Herbert Spencerによる心の進化に関する理論を研究の出発点とする．彼は，認知の要素が環境の要素に一対一で対応し，心の進化がその内と外の「対応性」の拡大で進むという見方を提示した．表象は対応性の一種と見なせるので，この理論により表象と進化を強く結び付けることができる．本研究では，この対応性の概念を採用し，コネクショニズムにおける表象の欠如を対応性の「ぼやけ」と考える．そして，対応性を生得的なものと獲得されたものに二分した上で，一対一のはっきりした対応性へ導く「選択圧」について論じている．

第３章，第４章，第５章では，生得的な対応性について論じている．まず，第３章では，環境の特徴が生得認知に同化されるプロセスを考察している．どのような行動であっても，その行動を実現する「実装」が無数存在するはずだが，その中で，進化的に有利なのは単に実行効率のよい実装である．次に，環境が変化する状況を考えると，対応性がはっきりすればするほど（環境の構造と行動の実装が同調すればするほど）環境変化の際に行動をその変化に合わせることが容易になるはずである．したがって，学習能力に選択圧が加わると，それが間接的に対応性にも加わり，行動の実装においてはっきりした対応性を導く．

これを「選択圧変換原理」と呼ぶ．第４章では，人工生命モデルでこの原理を検討している．実験では，食料採集タスクを題材とし，遺伝的アルゴリズムを用いてニューラルネットワークを進化させる．その際，学習の進化が生得的対応性に与える影響を検討するために，あらかじめ定めた学習アルゴリズムを使わず，「神経修飾」を導入したニューラルネットワークを用いて，ネットワークの結合重みの更新ダイナミックスを進化の対象とし，学習能力をゼロから進化させる．実験の結果，学習を必要としないタスクの場合には，試行ごとに異なる解が得られ，また，はっきりとした対応性が現れなかったのに対して，学習を必要とするタスクの場合には，学習能力に加わる選択圧が生得的な認知構造に強く影響し，環境の基本的特徴が表現され，はっきりした対応性を創発したことが示された．第５章では，生得的対応性が，実験心理学で知られている，あらゆる動物に内在する「学習バイアス」に相当すると論じている．学習バイアスとは，学習実験での動物のパーフォーマンスがたとえタスクは同じでも刺激や報酬の違いに強く左右されうる現象を意味する．そのバイアスは種によって異なり，その種の自然環境と生き方の特徴を表して学習を容易にすると考えられている．このことは，環境の特徴が暗黙的に種の生得な認知構造に符号化されていることを表すとみなすことができ，生得的対応性に合致すると論じている．

第６章，第７章，第８章では，獲得された対応性について論じている．本論文で焦点を合わせている心的表象は，生涯という時間スケールにおいて，主に環境と作用し合いながら獲得される．第６章では，「選択圧変換原理」を適応プロセスの階層において，１レベル上位に移動して適用する．一般に，２つの適応プロセスが異なる時間スケールで相互作用する場合，速度の遅いほうのプロセスにおいて対応性は創発する．したがって，学習における対応性の創発の文脈では，速度の速いプロセスは学習の学習，つまり２次学習に相当し，２次学習への選択圧は間接的に学習における対応性にかかると考えられる．そこで，空間的表象能力を要することで知られるトールマンの迂回迷路課題を２次学習問題として概念化しうることを示して，この理論を論ずる．第７章では，前章で提示した理論を検討するために，空間的表象能力を対象とした，迂回迷路課題を使った人工生命実験を行っている．実験では，神経修飾による二次可塑性回路を形成しうるニューラルネットワークを進化させた．その際，ネズミの表象能力に関する認知神経科学の知見に基づき，ネズミの海馬を模した神経細胞グリッドを記憶装置として導入した．計算機実験の結果，迂回迷路が解けるネットワークが進化した．分析の結果，ネットワークの中心に二次可塑性回路が見つかり，二次学習能力を用いて迷路を解いていることが確認できた．さらに，学習する際，学習の表象性を示すものと解釈できる，ネットワーク内部の神経細胞グリッドに迷路の形を映す発火パターンが形成されることも確認できた．第８章では，相手の心的状況を推測する能力である「心の理論」を対象とした人工生命実験を行っている．心の理論は典型的な心的表象であるため，コネクショニズムの手法でのモデル化は困難であった．実験では，環境状態と自分の心理状態に基づいて行動するエージェントに対して，もう一方のエージェントが環境状態だけを見て，その行動を予測するという抽象的な課題を設定した．二次学習課題を用いた進化実験では，適応度と表象性が並行して増加すること，さらに，表象性が高くなるほど適応度も高くなることが示された．さらに，相手の心理状況が学習後のネットワークの重みベクトルに反映されていることも確認できた．一方，一次学習課題を用いた進化実験では，適応度が表象性とは無関係で増加し，適応度が最大値に達しても表象性が小さい値に留まることが示された．この結果は，二次学習と獲得対応性の関係を裏付けるとともに，二次学習に基づいた手法による心の理論の進化モデルの可能性を示す．

第９章から第11章では，提案した理論を多面的に検討している．まず，第９章では，提案理論に対して予想される批判とそれに対する回答の提示を次の４点に関して行っている．第一に，心的表象とニューラルネットワークによる表象の関係，第二に，２次可塑性の必要性，第三に，２次学習の不要なタスクにおける表象の使用，第四に，自然環境における２次学習に対する選択圧の存在の普遍性である．第10章では，本研究の成果を踏まえて学習の意味について論じている．コネクショニズムは，パターン認識や強化学習などのツールとして目覚しい発展を遂げたが，一方で，心に関する計算論的な研究では期待に応えられていない．この原因がコネクショニズムの行動主義にあると指摘する．行動主義者たちは普遍的な学習ルールだけで心理のすべてを説明しようとしたが，これは結局，困難であることがわかり，行動主義は消え去った．しかし，ほとんどのコネクショニズム研究は未だに普遍的な学習アルゴリズムを追究している．心のモデルとしてニューラルネットワークの技術を十分機能させるには，普遍性の追求をやめて，学習を適応プロセスの主体としてだけではなく，適応プロセスの客体ともみなす必要があると主張する．第11章では，適応プロセス間の相互作用の一般化を試みている．本論文では進化と学習の相互作用に注目したが，時間スケールが異なる進化過程の間にも似たような相互作用が起こり得ることが「進化能力の進化」に関する研究で明らかになってきている．両者の間の類似性を考察し，両者がより一般的な現象のインスタンスであるとの見解を示している．

第12章では，議論を転じて創造性に焦点を合わせている．人工知能に創造性を持たせようとする典型的な研究では，ランダム生成，あるいは試行錯誤による探索に基づいた手法が用いられ，ヒトの創造性の中心にある「アイデア」が見落とされていることをまず指摘した上で，アイデア生成能力の起源は洞察的問題解決能力の進化にあると主張している．創造性に関するほとんどの人工知能研究では，問題解決の役割が無視されている（創造的活動以外は何もしない）．人間らしい創造性を持つ人工知能を育てるには，アイデアや洞察を得る能力を引き出すためにも，あるいは創造的行動に使える意味ある経験を得るためにも，問題解決が試される環境が必要であると論じている．このようなアプローチは現在の技術で十分可能であるが，人工知能研究はここにおいてもまだ，基本的には行動主義に留まっている．それゆえ，本論文で発展させた表象的知能の進化手法が活用できると主張している．

最終章では，本論文の成果を当該分野の現状の中に位置づけて結論を述べている．技術の飛躍的発展にもかかわらず，心の理解に対する人工知能の貢献は未だに薄弱である．そもそも，ヒトにおいて知能を示す行動は機械においては必ずしも知能を示さない．我々の知能概念の焦点は行動ではなく，その行動を生み出す認知プロセスにあるからである．人工知能研究が知能の解明に貢献するためには，ボンネットの中で起きていることを問わなければならない．本論文では，従来，行動しか問わないとされる進化過程がいかにしてボンネットの中を形成するかを検討し，表象的知能の起源を探った．要約するならば，「生得的表象は学習の進化の産物であり，環境と相互作用しながら表象を獲得する能力は二次学習の進化の産物である」という理論を提案し，人工生命モデルを用いて，その理論を概念レベルで実証した．