別表５
⑶

主　論　文　要　旨

No.1

報告番号	甲　乙　第　　号	氏　名	山田航太

主 論 文 題 名：
　学習心理学と行動分析学における数理工学的手法の実践と意義

（内容の要旨）

本博士論文では，学習心理学と行動分析学という実験心理学における学習研究の分野において，伝統的に扱われてきた現象に対し，強化学習や深層学習といった計算論的手法を用いてアプローチすることを通して，計算論的手法の導入による学習研究の可能性の拡張を目指した．実験心理学における学習心理学と行動分析学の両分野においては，それぞれパブロフ型条件づけとオペラント条件づけという 2 つの条件づけを主題として，様々な学習性の行動についての記述，説明がなされてきた．一方で，機械工学の分野を中心に発展してきた強化学習は，これら両分野に影響を受けて誕生した機械学習の一領域である．強化学習の枠組みには，上記の 2 つの条件づけ，及びそれらが関わる学習性の行動を包括的に扱う可能性が秘められている．本研究では，パブロフ型条件づけとオペラント条件づけが関わるいくつかの行動現象を，強化学習という枠組みの中で包括的に扱うことによって，実験心理学における学習研究において新たな可能性を示した．

　研究 1 においては，瞬間的な行動と瞳孔の動態を，計算論的な手法により計測および解析することを通して，瞳孔の大きさが報酬に対する予測を反映すること，そして，報酬の予測に伴う運動表出が薬物による阻害によって抑制されても，瞳孔の拡大が生じることを明らかにした．頭部固定装置を用いて，マウスにパブロフ型条件づけを行い，音刺激によって報酬が予測できる群と，報酬が予測できない群の 2 群をそれぞれ訓練し，その際のリッキングと瞳孔の大きさの変化を計測した．報酬が予測できる群では，リッキングと瞳孔の大きさは，音刺激によって上昇した．次に報酬の予測に伴って生じるリッキングが瞳孔の大きさに対して与える影響を低減させるために，ドーパミン D2 拮抗薬であるハロペリドールをマウスの腹腔内に投与することで，リッキングを抑制した．薬理学的操作の結果，ハロペリドールによってリッキングは濃

度依存的に抑制されたが，瞳孔の大きさの抑制は見られなかった．さらに，瞳孔の大きさが報酬予測を反映していることを裏付けるために，リッキングの動態を強化学習に基づいてモデル化することを通して，報酬予測の動態を推定し，その情報をもとに瞳孔の大きさの時間的な変化が予測できることを示した．

研究 2 においては，オペラント反応の消去に伴って生じる，消去バーストという現象の制御変数を，強化学習モデルを用いたシミュレーションによって同定し，マウスの実験によってその妥当性を検証した．マウスのオペラント反応の瞬間的な動態とその背後にある内部過程を強化学習の枠組みに基づいてモデル化した．さらに，提案モデルでは，強化学習における好奇心というアイデアに着目し，好奇心駆動型の強化学習をモデルとして実装することで，従来の行動分析学と学習心理学では説明できない現象であった消去バーストが生じる環境条件を明確に予測できることを示した．さらに，強化学習モデルによって明らかになった環境条件の下で，実際のマウスの行動においても消去バーストが生じることを明らかにした．

研究 3 においては，オペラント反応の局所的なバーストと休止期間によって特徴づけられる，バウト・休止パターンを強化学習によってモデル化することを通して，行動の背後のあるメカニズムを検証した．ここでは，現実の動物が従事しうる実験とは関連のない行動，すなわち他行動を，強化学習における状態として導入し，それらの間での遷移コストを仮定することにより，バウト・休止パターンという反応の時間構造を説明することを試みた．さらに，他行動の存在と遷移コストのいずれかを欠損させたモデルではバウト・休止パターンが生じないこと，そして双方を実装したモデルでは，バウト・休止パターンのみならず，これまでの動物実験で蓄積された結果を再現できることを示した．

研究 4 においては，ネットワーク科学的な視点を強化学習に導入することで，研究 3 において示した，学習と行動の理解のための他行動という概念の導入というアイデアを洗練し，オペラント条件づけと，パブロフ型条件づけの双方が関わるとされる習慣形成と呼ばれる現象に対して，新たな仮説を提唱した．ここでは，行動を反応が相互結合したネットワークとみなすことで，強化学習における状態遷移に，ネットワークの構造という，従来の行動分析学と学習心理学にない視点を導入した．この試みにより，習慣形成と呼ばれる現象をネットワークの構造変化として説明できることを示

した．

研究 1 ではパブロフ型条件づけによって獲得された反応の動態，研究 2 ではオペラント反応の消去に伴う変化の過程，研究 3 と 4 では，バウト・休止パターンと呼ばれるオペラント反応の時間構造や，オペラント反応とパブロフ型条件づけの双方が関与する習慣形成という現象を，他行動に注目して，強化学習によってモデル化した．このように強化学習は動物の学習性の行動を，オペラント条件づけとパブロフ型条件づけの双方を包括的に扱うことが可能であることを示した．強化学習は神経科学や実験心理学内の諸領域でも応用されており，今後はそうした隣接領域との接点を創出することも可能となる．個々の研究では，強化学習による機械論的，数理的な行動，内部過程の記述によって生み出される，行動分析学と学習心理学の異分野との接続性についても考察した．本博士論文においては，このような個々の実験事例を挙げながら，計算論的手法が行動分析学と学習心理学という実験心理学における学習研究の既存の文脈の中で果たす役割と，他領域との接点を作り出すという役割，つまりは行動分析学と学習心理学の深化と拡充のツールとしての意義を示した．