

2504PavlovAI. pdf

2025. 4 ブログ：「パブロフの複合型条件づけ研究の展開」を読んで、の詳細
(→ <http://www.1968start.com/M/blog/index4.html#2504>)

「パブロフの複合型条件づけ研究の展開」を読んで

中所武司

■このエッセイのきっかけ

人工知能学会誌の最新号の下記の解説論文の『パブロフの複合型条件づけ』が、50年以上前の第一次AIブームの時の私の卒業研究との関連で気になり、読んでみた。

- ・人工知能学会誌「人工知能 40巻2号(2025年3月)」
特集：「AIと神経科学の接点2025」
p. 214-220
強化学習をめぐる動物実験心理学から神経科学研究への変遷史：
パブロフの複合型条件づけ研究の展開を中心として
- ・(参考) 私の卒業論文「条件反射の生体工学的解析」(1969)
<https://www.1968start.com/M/bio/olduniv/soturon.htm>
- ・(参考) 学会発表「条件反射における学習機能に注目した回路モデル」
電子通信学会全国大会、242(1969).
<https://www.1968start.com/M/bio/olduniv/gakkai1969.html>

■内容の要約とコメント (→★)

- ・『驚きは知識の始まりである』プラトン

→★本文の最初に掲げられたこの句は、犬が条件反射を習得する前提として、
条件刺激(例：ベルの音)の後の無条件刺激(例：餌)に驚く必要を示す。

1. はじめに

- ・強化学習研究において重要な転換点が訪れている。
- ・強化学習の代表的なアルゴリズム Temporal Difference Learning (TD 学習) は、報酬の予測誤差 (TD 誤差) を時間的に前の状態・行動に逆伝搬させる。
脳内では、ドーパミン細胞の活動パターンがこの TD 誤差と類似していることから、ドーパミン細胞と密接に結合する大脳基底核が強化学習の座として注目されてきた。

- ・近年、この理論的枠組みに再考を迫る実験結果が報告された。TD 学習モデルでは、報酬から逆向きに、その前の状態へと順次的に行動価値が伝搬すると考えるが、実験では、複数の手掛かり（例えば光 CS1 と音 CS2）を用いた場合、先行する条件刺激（CS1）に対して学習初期からドーパミン系が即座に反応した。これは、TD 学習モデルでは説明できず、大脳基底核ドーパミン系の機能について新たな理論的説明を必要としている。

→★光の後に音という複数の手掛かりで条件付けを行った場合、音の前に初めの光だけでドーパミン細胞が反応したということかな。
この時点では唾液分泌がないはずなので、矛盾はないのでは。
光の刺激の後で餌につながる音の刺激を期待してドーパミン細胞が反応しても不思議はないと思われるのだが・・・

- ・この問題の本質を理解するには、「驚き」という概念の歴史的変遷を辿る必要がある。TD 誤差は、「予測と結果の不一致」という意味での「驚き」と解釈できる。「驚き」の概念は、TD 学習の理論的基盤を与えた Rescorla-Wagner モデルに由来し、無条件刺激（US）に対する条件刺激（CS）の予測性という意味で定式化されていた。

→★条件刺激 CS の後での、予測していなかった無条件刺激 US に対する「驚き」ですね。

- ・「驚き」の概念は Ivan Pavlov までさかのぼることができる。Pavlov が最初に記述した「驚き」は、US（無条件刺激）の予測性ではなく、新奇な刺激や環境変化に対する生得的な定位反射 (orienting reflex: OR) だった。「おや何だ反射」とも呼ばれ、生体が環境の変化や潜在的な危機に適応するための重要な機能である。

→★卒論では『新奇な信号に対して起こす定位反射（詮索反射とも言う）』と記載

- ・Pavlov は、CS と US を対提示することで条件反射が形成されると考えたが、1920 年代後半、J. Konorski は、新しいタイプの条件づけ (道具的条件づけ) を報告した。

→★卒論では、道具的条件づけ（操作的条件づけ）の以下の 3 例を記載している。

- ・スキナー箱の中のネズミは、食餌を得るレバーを連続的に速く押すようになる。
- ・一つの部屋を二つに仕切り、壁を白と黒に塗り分けておくと、床から電気刺激を与える白い部屋に入れられたネズミは全速力で黒い方へ逃げようになる。
- ・モルモットを回転籠に入れ、ブザーの 2 秒後に回転籠に電気刺激を与え、回転籠をある角度以上回せば電気刺激を回避できるようにしておくと、モルモットはブザーがなるや否や回転籠を回すようになる。

- ・1960年代の情報処理の革命期、パブロフ型条件づけは情報の予測過程と捉えられ、強化学習理論の基礎となる条件づけモデルを構築するに至った。

→★1960年代は第一次AIブームのときで、私の卒業研究（1968～1969）では、条件反射を学習機能とみて、そのモデル化と回路設計を試みた。

- ・本稿では、この歴史的な文脈を踏まえ、以下の3点について考察を進める。
 1. Pavlovの「おや何だ反射」から複合条件刺激実験による理論進展までの歴史的過程
 2. Konorskiの複合条件刺激実験を用いた新しいタイプの道具的条件づけの発見
 3. Kaminの複合条件刺激実験による「阻止」現象の発見から、Rescorla-Wagnerモデル、そしてTD学習理論への展開

2. 心理学の学習理論：きっかけはパブロフの卓見「おや何だ反射」に端を発する

- ・Pavlovは1904年、消化腺の研究でノーベル生理学・医学賞を受賞。その過程で、イヌが餌を見せる前に飼育係の足音を聞いただけで唾液を分泌し始める現象に気づき、条件反射の研究に注力し、学習心理学における条件づけ理論の基礎を築いた。
- ・Pavlovは医学・生理学の研究者であり、当初の関心は、学習現象そのものよりも、**大脳や神経系**の生理学的な機能にあり、『高次神経活動の客観的研究』（1923年）や『大脳半球の働きについて』（1927年）など、脳や神経に関する著作がある。

→★私の卒論の参考文献45件の中の【生理学、生体工学関連】23件のうち、調査メモが残っている10件は、**大脳や神経系**に関するものが多い。

（例）「脳と神経系」、「脳生理学の基礎」、「人間の大脳活動」など

（参照）卒業論文の参考文献の調査メモ

<https://www.1968start.com/M/bio/olduniv/memo/index.html>

- ・著書『大脳半球の働きについて』に、探索反射として、外界の音刺激に注意を向け、音源の方向に顔を向ける定位反射（orienting reflex：OR）が記されている。この「おや何だ反射」と呼ぶORこそ、後の「驚き」の概念の萌芽といえる（図1）。

・図1：「Pavlovの条件反射研究からRescorla-Wagnerモデルに至る研究の流れ」

- ・この著書には、**複合的な条件刺激**（CS1とCS2）を用いた条件づけ実験において、一方の刺激（CS2）の条件づけが形成されない「**隠ぺい**」という現象が報告されている。例えば光（CS1）と音（CS2）を同時に提示（複合刺激）してUSを与えた場合に、一方のCSのほうが生得的・物理的に強く知覚される（より顕著である）と、

もう片方の CS への条件づけが弱まったり生じなかったりする現象を指す。

→★卒論の参考文献（A. B. コーガン：脳生理学の基礎、1959（原著））の調査メモには、「複合刺激に対する条件反射」、「同時複合刺激」、「継時複合刺激」などの記載あり。
（参照）<https://www.1968start.com/M/bio/olduniv/memo/chu3.pdf>

- この隠ぺいという複合 CS の競合現象は後の「阻止」の研究と並んで、複合型条件づけ理論の重要な論点となっていく。

3. Konorski による道具的条件づけの発見と理論的展開

- この Pavlov の著書出版の 1927 年頃、J. Konorski は、画期的な発見をした。音刺激（CS1）と同時にイヌの右足を持ち上げる刺激（CS2）を与え、その後に、餌（US）を与えるという手続きを繰り返すと、やがて音刺激を与えただけでイヌが自発的に右足を上げるようになった。
- この学習は Type II と呼ばれ、後に Skinner により“オペラント条件づけ”とされるが、Pavlov が提唱した「刺激置換説」（音刺激が餌の代理として唾液分泌を誘発する）では説明できない。
- Skinner は、自発行動が環境に対して自発的に働きかけていく（operate する）ので、「オペラント反応」と名付け、現在は反応形成（shaping）という手続きで呼ばれる。

→★私の卒論の「1.1 条件反射」の節で、条件反射の形成法として、古典的条件付けのほかに道具的条件付けとして『スキナー箱』に言及している。

- Skinner は、心理学を実験科学として確立するには、観察可能な行動と外部環境だけを扱うことが最も実証的であるという信念を徹底し、「内的経験」（思考や感情など）を科学の対象として扱うことには慎重で、行動と環境の相関を丹念に測定・制御することで行動原理を探求すべきだとした。
- のちに認知革命が起こると、Skinner の立場は「過激すぎる」と批判されたが、同時に実験方法論の厳密さやオペラント条件づけの有用性は、現代の神経科学や応用行動分析に多大な影響を与えた。

→★私の卒論の「2.1 条件反射の学習性」の節では、道具的条件付け、弁別学習、迷路学習などの学習過程を、古典的条件付けとの比較で検討しており、『古典的条件反射も道具的条件付けも学習過程の本質は同じ』と結論付けている。

4. 認知動物心理学と Konorski

- その後のコンピュータ技術や情報通信技術の飛躍的な発展により、知的活動をアルゴリズムや情報処理という枠組みで捉える見方が広がった。また、内部表象や認知プロセスなど、行動の背後のメカニズムを研究する認知革命が生じ、行動主義が重視した「外部刺激と反応の相関」という図式を拡張し、脳内の情報処理過程を扱うパラダイムが台頭した。

→★私の卒業研究「条件反射の生体工学的解析」(1968~1969)は、この流れに対応する。

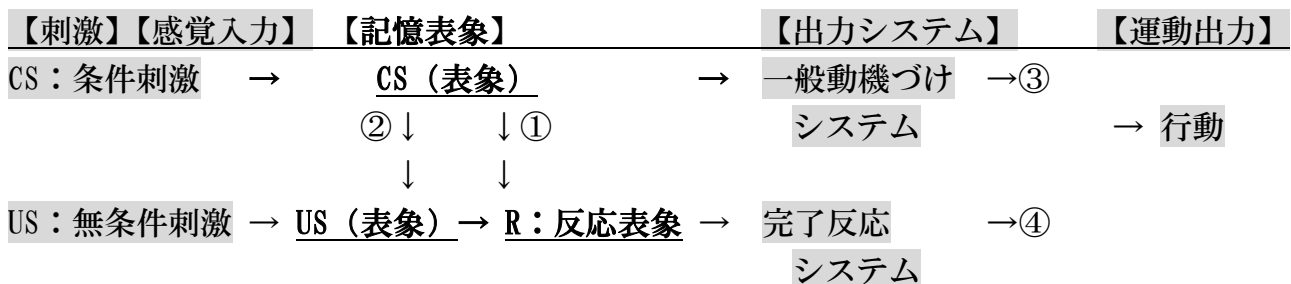
(参考) 学会発表「条件反射における学習機能に注目した回路モデル」

電子通信学会全国大会、242 (1969)。

<https://www.1968start.com/M/bio/olduniv/gakkai1969.html>

- この時期、Konorski は1967年に著書『Integrative Activity of the Brain』を出版。新たに形成されつつあった認知科学や神経科学、動物行動学などの視座から、脳と行動の関係を総合的に説明。学習理論を中心に、刺激と行動の連合関係を、認知的表象や記憶の過程としてモデル化している。

- 図2：Konorski の想定を拡張したパブロフ型条件づけの表象図式



- ① Pavlov が想定した刺激置換説の経路
- ② Rescorla-Wagner モデルにおけるCS-US 連合経路
- ③ 準備反応出力経路（接近・回避行動など）
- ④ 完了反応出力経路（唾液分泌・足の屈曲反射など）

→★この図の解釈（推測）

① パブロフは、ベルの音が直接唾液分泌を導いたと考えた。

② Konorski は、ベルの音が餌を連想（連合）させて、唾液分泌を導いたと考えた

- Konorski は、条件づけにおける内的認知過程を「表象」としてシンボリックに扱い、理論化する認知的パラダイムを取り入れ、「認知的学習理論」の基盤となる。

5. Kamin (1969 年) の実験：

複合刺激をめぐる「隠ぺい」と「阻止」の条件づけの問題提起

- Pavlov の研究で理論的に説明されていない「隠ぺい (overshadowing)」

(物理的な強度差により片方の CS の学習が弱まる現象) に着目した Kamin は、新たに最初に CS1 (光) と US (電撃) を対提示して学習させ、次に CS1 (光) + CS2 (音) の複合刺激と US を提示すると、最終的に CS2 (音) だけでは条件反応が形成されない「阻止 (blocking)」という現象を報告した。(表1)。

- 表1：[Kamin 68] の「阻止 (blocking)」の実験デザイン

群	第1段階	第2段階	第3段階	結果
阻止群	光	光+音	音	→反応なし
統制群		光+音	音	→条件反応

→★下段：「光+音」で条件付けした後に、「音」だけの条件づけが可能だった。

上段：まず「光」だけで条件付けした後に、「光+音」の条件付けが可能だったが、その後に、「音」だけの条件付けができなかった。

→★下記の疑問について、後の説明で、前の段階の条件付けを抑制していないとわかる。

* 第2段階の条件付けの前に、第1段階の条件付けを抑制しているか否か、

* 第3段階の条件付けの前に、第2段階の条件付けを抑制しているか否か。

- **阻止**では、通常なら CS と US の単純な対提示で容易に形成されるはずの条件反応が、CS1-US の連合を学習していると、後から追加された CS2 には学習が起こらなくなる。

- Kamin はこの説明のため、「**驚き (surprise)**」という認知的概念を導入した。

動物が CS によって US を予期できるようになると、US がもたらす驚きが失われる。

このため、先行する CS1 によって US が予測可能になっている状態では、

後から導入した CS2 が US 出現の手掛かりとしての情報価値を獲得できないのである。

→★表1の説明と異なるのでは？ 表1では、第2段階の「光+音」が必須に見える。

「光+音」の条件付けした後、「光」や「音」だけでは反応なしの状態と解釈したが、上記の説明では、「光+音」の条件付けもできなかったことになるのだが・・・

6. 現代動物学習心理学の理論的発展の原点：Rescorla-Wagner モデル

- 連合学習の古典的原理として、CS と US が対提示される回数が増えるほど CS-US の連合が強化されるとする「**頻度の原理**」があるが、

複合条件づけにおける「阻止」はこの原理では説明が難しい。

→★表1の第3段階での「音」だけでは反応なしのことかな？

- Rescorla-Wagner モデル(1972年)は、「驚き」仮説を数理モデルとして定式化した。USの予測誤差(実際に提示されたUSと動物が予測するUSとのずれ)を $(\lambda - V)$ と表し、「驚き」の大きさを定量化した。

$$\Delta V = k(\lambda - V)$$

λ : USの実態価値,

V : CSがもつ連合強度と定義される。

ΔV : USとの連合強度の変化量,

k : CSとUSの明瞭度を表す定数である。

- 学習の初期では V が小さく、 $(\lambda - V)$ が大きいので動物は「大きく驚き」、 ΔV がプラスに変化しやすい。
- 学習が進み V が λ に近づくにつれ $(\lambda - V)$ が小さくなるため、驚きが減少し、 V の増加量 ΔV もしだいに小さくなる。
- 最終的に V が λ に到達すると、 $(\lambda - V) = 0$ であり、これ以上学習は起こらない。

→★「学習は起こらない」というよりも「学習が完了した」ということでは。

- 阻止の実験で、第1段階で光(CS1)の連合強度がすでに λ 近くに到達しているとする。第2段階で光+音(CS1+CS2)とUSを対提示して、二つのCSの合計予測値は λ に近い。したがって $(\lambda - [V(\text{CS1}) + V(\text{CS2})])$ が0または極めて小さいため、音(CS2)の連合強度を上げる余地(=驚きの余地)がなくなる。その結果、音単独のテストでは条件反応(CR)が形成されにくい、つまり阻止が起こることを数式的に説明できるのである。

→★この説明では、表1の阻止の実験では、第1段階での「光」での条件付けの後に、その条件付けを保持したまま、第2段階で「光+音」で条件付けをしている。さらに第3段階で「音」だけの条件づけができなかったということである。ところが、第1段階を省いて、第2段階で「光+音」で条件付けをした場合は、第3段階で「音」だけの条件づけができたということである。この両者の違いの説明のために「驚き」の概念が導入されたとのことである。

- Rescorla-Wagner モデルは、誤差修正（驚き低減）の仕組みが「阻止」をはじめ、「隠ぺい」などの複合刺激効果を説明可能にした点が画期的であった。
- さらに、「共有連合原理（shared associative principle）」と呼ばれる、複数の CS（CS1, CS2）を合算して US との連合を計算する考え方を導入することで、複合 CS の競合現象を数式で記述できるようになった。

7. Rescorla-Wagner モデルのその後の展開と現代の神経科学および情報処理科学への接続

- Rescorla-Wagner モデルの数年後、Mackintosh は CS（条件刺激）側の有効性、すなわち「注意」に焦点を当てたモデルを提案した。各 CS に対する注意パラメータ α を導入し、CS が US の予測において重要であればあるほど（CS が有効な予測子であればあるほど） α が高くなり、学習が加速するという仮定を置く。

→★この「注意」は、2章でのパブロフの「おや何だ反射」（定位反射）と同じでは？
ベルの音に「おや何だ」と注意が向くから、その後の「餌」と条件付けられる。

- 一方、CS が US を予測しないとわかると、その CS への注意が低下する仕組みを導入。

→★私の卒論では、『起こるべき条件反射が起こらないという抑制作用』として、
当時の文献調査結果を詳しく引用している。

- これに対し、Pearce と Hall は逆の仮定を置くモデルを提示した。ある CS が US を完全に予測するようになると、その CS に注意を向ける必要がなくなり α が下がり、逆に CS が US を予測しない（あるいは予測が不確かな）場合にこそ、その CS への注意が高まり α が上昇すると考えた。

→★これは「逆の仮定」ではないのでは。ベルの音で「餌」を予測するようになると
「おや何だ」とは思わなくなるのは自然なことなので。

- これらのモデルでは、CS が US 到来のシグナルとなるか否かで α が動的に変動し、それを行動実験（定位反射 OR の測定など）で検証している。

- US（無条件刺激）側の驚きだけでなく、CS 側への注意が学習を調整するという視点は、パブロフ型条件づけにおける複合刺激の効果を精密に説明する学習理論へ拡張された。

→★CS への注意と US への驚きが両者の条件付けの必須条件なので、当然の視点と思う。

8. 環境予測誤差モデル

- 最近、神経科学分野で新しい概念モデル「**環境予測誤差 (EPE)**」が提案されている。内部モデルで生成された環境予測と実際に知覚された環境との差異として定式化され、環境全体の予測性を対象とする点で、より広範な予測誤差を扱うことができるが、数理モデルはまだ確立されていない。

9. 現行の強化学習モデルの課題

- 本稿では、行動心理学の研究史を振り返りながら、数理モデルの構築過程で報酬 (US) を基点とする「**驚き**」(TD 誤差) へと移行した歴史的経緯を概説した。
- 一方、人工知能の領域の深層学習モデルでは「**注意**」の概念が注目され、大きな成果を上げているが、主に言語処理や画像認識を対象としており、強化学習 (RL) における「**驚き**」や「**注意**」は相対的に議論の範囲が限られてきた。
- 強化学習は、報酬を基点に行動方策を立ち上げる数学的枠組みで発展してきたので、強化学習アルゴリズムは、各ステップで必ず行動が選択されることを前提としており、探索が強制的に行われる構造になっている。報酬を絶対的な指標とする代わりに、動物が本来もつ自発性や好奇心、警戒感など、感覚刺激を軸とした行動生成の視点が抜け落ちる。
- 複数の CS や環境の新奇性を考慮し、自発性や好奇心、警戒感を適切に捉えるには、動物の脳応答や行動生成をより柔軟に扱う実験手法と解析手法が不可欠である。動物が本来もつ多様な動機づけや刺激反応を組み込んだモデルを構築することこそ、今後の強化学習研究をさらなる段階へと導く鍵となる。

10. 結論と展望

- 強化学習の神経基盤として知られる TD 学習モデルは、US 側の情報処理に着目した Rescorla-Wagner モデルを基盤にしている。しかし、Jeong らの実験では、複数の CS (CS1, CS2) を継時的に提示した際、ドーパミン放出は予測と異なり、学習の初期から CS1 でも CS2 でも同様に起こるという結果が得られた。これは US 側からの予測誤差が逆向きに伝搬していくという TD 学習の仮定と矛盾する。
- この矛盾の要因の一つは、TD モデルが US 価値基盤の情報伝搬モデルであること。複数の CS を扱う新しいモデルでは、CS それ自体に対する「**驚き**」や「**注意**」の処理が考慮される必要がある。Mackintosh モデルや Pearce-Hall モデルのように

CS への注意を可變的とする認知的アプローチを導入すれば、ドーパミン神経活動のデータとより整合的なシミュレーションが可能になるかもしれない。

- パプロフ型条件づけの本質が「おや何だ反射」という CS の新奇性に基づく学習過程であることから，こうした試みは，今後の強化学習研究の新たな可能性を示唆する。さらに，EPE のように環境全体を対象とした予測誤差の概念を取り入れれば，複雑な環境への適応や知的行動の形成を，従来以上に包括的かつ柔軟にモデル化できる可能性がある。
- 総じて，複合型情報（CSs）を用いたパプロフ型条件づけの理論的モデルと，環境全体の予測可能性を視野に入れたモデルが融合することで，「驚くべき」発見や分野横断的な連携が今後さらに進むと期待される。
- この歴史的経緯と新展開から，動物学習心理学・行動神経科学と強化学習理論の相互作用が，脳と行動，そして人工知能の理解を深めるうえで重要になってくる。今後は，CSs の新奇性やその予測性などの多様な視点を取り入れ，複雑環境下での学習過程を包括的に解明していくことが期待される。

以上