慶應義塾大学 湘南藤沢キャンパス 知覚・認知・モデル論
講義ノート(オンライン開講)

著者：松井　大
講義日：2021/01/12

0. はじめに

　この講義ノートは、2022年1月12日に慶應義塾大学湘南藤沢キャンパスにて (オンライン) 開講の講義「知覚・認知モデル論」で使われたものです。この講義は先端研究という枠で、1回ぶんだけ松井によって担当されました。まず、担当者の自己紹介をしておきます。松井の現在 (2022年1月時点) の所属は北海道大学 人間知・脳・AI研究教育センターで、吉田正俊研究室の博士研究員 (学振PD) をしています。元々の専門は比較認知科学です。比較認知科学とはざっくりいうと、動物の動物の行動比較を通じてその心理学的メカニズムの多様性と普遍性を描出し、ひいては行動・認知の進化の解明を目指す分野です。具体的には、カラス・ハトを用いた行動研究をこれまでは行っていました。現所属では、マーモセットを用いた神経科学的研究に従事しています。が、その話はこの講義では出てきません。

　本講義は大きく分けて4つのパートにわけてお話しようと思います。まず、学習心理学の基本事項を解説します。この部分は多くの人にとって復習に相当すると思いますので、雰囲気を見ながら飛ばし飛ばし紹介する予定です。次に、連合学習理論として代表的な3つのモデル (Rescorla-Wagnerモデル、Mackintoshの注意理論、Pearce-Hallモデル) を紹介します。これらのモデルはいずれも古典的条件づけの諸現象を説明するために提案されました。従って、モデルの考え方を関連の深い現象とともに説明することにします。3つ目の話題は、道具的条件づけの最重要現象の1つである対応法則 (matching law) について紹介します。対応法則自体は、別のコマで聞いたことがあると思います。そこで、対応法則が記述的法則であることを超えて、巨視的な強化の最大化から導かれる行動の最適化の法則であることを理解するのを目標とします。さて、古典的条件づけの理論では、予測誤差を低減するという目的をどのように実行するか試行ごとのアルゴリズムを考えます。一方で、対応法則の巨視的理論は、最大化という目的から行動がどの方向にいくかを

説明しています。最後に、自由エネルギー原理と、そこから得られる帰結である能動的推論を紹介します。能動的推論により、その都度その都度動物が実行する推論という観点から行動を最適へと導いていく過程の定式化を行います。また、予測誤差が重要な役割を果たすことを見ていきます。単に講義で話すことの台本というよりは、講義で話せなかったことや、途中の式変形などもできるだけ丁寧に記述していこうと思います。とはいえ、一回分の内容なので、本来分野的には重要な話題が抜け落ちてしまうのは避けられません。なので、そのぶん参考文献を多めに推薦することで、お茶を濁すことにします。

御託は以上で終わりにして、早速講義の中身に入っていきます。それに合わせて、口調を「です・ます」調から「である」調に変えようと思います。

1. 学習心理学、行動分析学、超入門

1.1. 学習とは何か？

そもそも、学習とはなんなのだろう？教科書的にはお決まりの定義がある。それは「経験によって生じる、比較的永続的な行動の変化」である。変化の部分は、変容ということもあるが、大きな違いはない。この定義には、「経験によって生じる」「比較的永続的」「行動の変化」という3つの要素が含まれている。それぞれに説明を要すると思う。順に見ていこう。

最初の「経験によって生じる」というのは、動物が環境とのやりとりする中で起きるということだ。例えば、子どもが一定の年齢になって反抗期に入っても、行動はドラスティックに変わる。しかし、反抗期は発達過程の中で自然に生じる傾向である。従って、反抗期になることは学習ではない。ただもちろん、反抗期になることで行動が変わり、その結果親との関係をはじめさまざまな変化が生活に起きるだろう。それによって学習されるものはあるだろうし、それは自体はなんらおかしいことではない。他にも、同じ理由で、成長に合わせて声変わりをすることも学習ではない。年老いて走れなくなることも学習ではない。そう考え始めると、学習以外にも行動が変わる要因はいろいろありそうだ。しかし、学習というときは、環境とのやりとり、つまりは経験により生じる行動の変化のことを指すということだ。「じゃあ、言葉を話すようになるのは学習なの？」と思われるかもしれない。言葉を話すようになること (言語獲得) には、周囲の大人など、子どもを取り巻く環境に接する経験が関与しているのは明らかだ。一方で、ヒトだけが言語を話すこと、発達の特定の段階で言語が獲得されることから、言語獲得には明白に遺伝的な基盤がある。このようなケースは、どう考えるべきか？心理学というより、単に国語の問題なのだが「経験によって生じる」といったときに、それは何も「遺伝は関係しない」と言っているわけではない。現代のヒトを含んだ動物の研究者で、行動が遺伝で決まっているのか環境で決まっているか、すなわち氏か育ちかを明確に区別できると考えている人は (お

そらく) いない。あらゆる行動にはなんらかの形で遺伝が関与している。そこは認めた上で「経験によって生じる」行動の変化を学習と呼んでいるわけだ。

2つ目の「比較的永続的」であるというのは、一過的な現象ではないということである。例えば、アルコールを飲んだり、過度に疲労したりしても、行動は変わる。お酒を飲み過ぎたら、普段は言えないような汚い言葉を使う人がいる。これも確かに行動の変化だが、次の日には同じような言葉を吐かなくなっていると思われる。アルコールなどの薬物による行動の変化は学習ではないということだ。ただ、例えば、学生の多くは受験勉強で多くの知識を蓄えるが、学生生活の中で大半を忘却してしまうのが世の常である。勉強で今まで答えられなかった問題に正答できるのは明らかに学習のように思われるため、「比較的永続的」というのはどの程度なのかと疑問に持たれるかもしれない。ここでは、あまり難しく考えず「比較的」という修飾語を適当に幅を持たせて解釈しておけばよい。

最後の「行動の変化」であるが、「行動」とはなんだろうか？実はさまざまな議論があって、簡単に決まるものではない。とはいえ、この点に関しても入門の段階では難しく考えず、ひとまず「動物個体が全体として行う、機能的な活動」程度に見なしておけば火傷をすることはない。しかし、「また煙に巻くのか」などと失望されても困るので、少しだけ説明をしてお茶を濁すことにする。ラットがレバーを押せば餌をもらえる場面を考えてみよう。このとき、レバーを押すことは明らかに行動のように思われる。それでは、レバーを押すときの筋肉の動きはどうだろう？あるいは、筋が動くときの電気的変化は？それを生み出すのに伴って働く、脳の活動 (活動電位、脳の血流変化、グルコースの代謝、etc...) は行動なのか？あるいは、動物の「情報処理」や「認知」は行動なのか？このように考え始めるときりがないように感じられる。そこで、「行動」と呼ぶときには動物の個体が全体として遂行する活動としておけば、この問題は回避できそうだ。勘違いしてないでほしいのだが、行動について考えるときに脳活動を考えるのは無意味であると言いたいわけではない。学習心理学が扱う行動を定義する上では、このように定義するのが混乱を招かないため、便利であるということだ。この問題についてもう少し詰めていきたいのならば、Baum (2021) に当たるとよい。Baum (2021) は、単なる便利さを超えて、必然的に行動を全体的活動として定義すべきであると積極的な主張をしている (Baumはそれを行動の存在論と呼んでいる)。

もう一点、上の定義では「機能的な」という修飾語がついているが、ここには説明が必要だろう。機能的であるというのがどういうことかというと、環境に働きかける、その内容によって行動を定義せよという意味である (Barnes-Holmes & Barnes-Holmes, 2000; Skinner, 1935)。例えば、ラットがレバーを押したら餌が出てくる実験の場合、右手で押そうが左手で押そうが餌が出てくる。手で押さなくても、身体を押し付けようが、レバーの上に乗っかることで押しても構わないわけである。このレバーは、センサーが感知できる程度にレバーが下がる圧力がかかれば、同じように餌が出てくるという結果

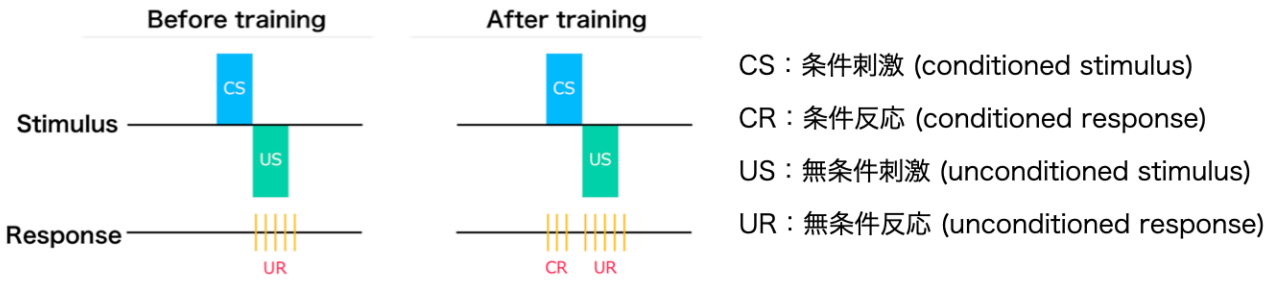

図1. 古典的条件づけの基本的な手続き

を引き起こすのである。よって、同じ結果を引き起こすどの活動も「レバー押し」行動とひとまとめにできそうだ。これが機能的に行動を定義するということである。右手で押すのと左手でレバー押すのでは、運動の形態的には全く異なる。そのことを「トポグラフィーが違う」というが、トポグラフィーが違っても、機能が同じなら同じ行動として一まとまりの行動として扱う。このようなまとめ方をした活動の集まりを「行動クラス」と呼ぶこともある。

以上で、ひとまず学習の定義が決まった。次は学習の分類である。学習には古典的条件づけと道具的条件づけがある。2つまとめて連合学習という。この連合学習における2つの学習について、簡単に見ていくことにしよう。簡単に見ていくというのは、この講義を理解する上で必要な知識だけを紹介するということだが、どちらの条件づけにも豊穣な世界が広がっていることは、心に留めていてほしい[1]。興味があれば、澤 (2021)、坂上・井上 (2018)、メイザー (2008) などに当たるとよい。

1.2 古典的条件づけ

古典的条件づけはIvan Petrovich Pavlovに発見された学習現象である[2]。既に別の講義で聞いたことがある人が多いと思うが、復習がてらPavlov (1929) の研究を元に実験手続きを紹介しよう (図1)。まず、固定されたイヌに対し、メトロノームの音を聞かせる。メトロノームは、特に生物学的に重要な刺激ではない。ゆえに、イヌは音の方向を定位する反応などは示すが、次第にそれも減少していく。Pavlovはメトロノームの音の直後に餌を提示する。餌は当然、動物にとって生存上重要な刺激である。餌へのイヌの典型的な反

[1] なにやら定型文的で、つまらない注意喚起だと思われたかもしれない。なぜこういうことを書いたかというと、私自身の (苦い？) 思い出に原因ある。学部4年生の頃、私は他大学の臨床心理学の修士学生と会話する機会があった。そこで「鳥類の行動研究に興味があります」と自己紹介したら、「え？ハトの研究なんて、もう単純な条件づけで終わってるんじゃないの？」と言われたのだ。「単純な条件づけ」というのが何を指すのかもよくわからないが、当時は気弱な学部生だったので、あまり剣呑な態度を取らず、苦笑いをするしかなかった。

[2] 正確には、Pavlovに影響を与えた先駆的な研究がそれ以前にもある。詳しくは澤 (2021)

表1.

訳語	英語	略語
条件刺激	conditioned stimulus	CS
条件反応	conditioned response	CR
無条件刺激	unconditioned stimulus	US
無条件反応	unconditioned response	UR

応は、唾液を分泌することである。私たちも、深夜にSNSを見ていたら、ステーキの画像をアップしている友人の投稿を見て、同じような経験をしたことがあるだろう。このように対に提示することを「対提示」という。また、この場合、メトロノームは条件刺激 (conditioned stimulus)、餌は無条件刺激 (unconditioned stimulus) と呼ぶ。また、無条件刺激が引き起こす反応 (餌の場合唾液分泌) は無条件反応 (unconditioned response) という[3]。

条件刺激 (メトロノームの音) に対しては、当初は目立った反応は生じない。しかし、Pavlovは繰り返し、メトロノームと餌を対提示した。結果、メトロノームの音に対して唾液分泌が生じるようになった。この反応を条件刺激に対して条件反応 (conditioned response) が獲得されたという。以上をまとめて、条件刺激 (CS)、条件反応 (CR)、無条件刺激 (US)、無条件反応 (UR) と略して書くことが多い (表1)。本稿でも、これ以降、略語を用いて話を進める。古典的条件づけ (classical conditioning) とは、このように複数の刺激を対提示することで行動が変化する学習様式である。人によってはパブロフ型条件づけ (Pavlovian conditioning) と呼ぶこともある。Pavlov以降、古典的条件づけ研究ではさまざまな刺激対が用いられ、研究が進められている。例えば足元への電気ショックなど、侵害的な刺激をUSとして用いる恐怖条件づけがある。恐怖条件づけは、神経科学でも情動や恐怖記憶の神経基盤を調べるために広く用いられている。

古典的条件づけは刺激対を提示することで生じる学習であると述べたが、そこで学習されるものは何なのか？条件づけを用いた学習を刺激反応学習 (stimulus-response learning) と呼ぶことがあるが、刺激–反応 (S-R) の連合が古典的条件づけでは獲得され

[3] 唾液分泌だけが食物への無条件反応というわけではない。Pavlovの実験の場合、イヌの身体が固定されていたというのがミソになっている。自由に行動できる動物なら、唾液分泌以外にも、そちらに向かう接近反応だったり、空腹状態によっては情動的な反応だって起きるだろう。1つの刺激が必ずしも1つの反応を引き起こすわけではない。

ているのだろうか？[4] Pavlov以降の多くの研究で、話はそう単純ではないことがわかっている。ここではその一例として、Rescorla (1973) の恐怖条件づけの研究を紹介しよう。彼は、ラットに対して、光刺激をCS、大きい音をUSとして用いた。大きい音は動物にとって嫌悪的な刺激であるため、恐怖反応を引き起こす。CSに対する条件反応が確認された後、RescorlaはUSを単独で繰り返し提示した。このとき、CSは出していない。大きい音だけをラットに聴かせるということだ。大きい音は確かに不快だが、何度も聞かされると慣れてきてしまう。馴化 (habituation) が生じるというわけだ。馴化が生じたら、ラットはもはやUSに対して恐怖反応を示さなくなる。ラットがUSに対し馴化した後、再びCSを提示する。すると何が起きるだろうか？動物の中でS–R学習が生じているならば、CSに対し反応が直接結びついているため、変わらず反応が表出されることが予想される。一方で、CSとUSの関係性についての学習、すなわちS–S学習 (stimulus–stimulus learning) が生じているのなら、USがもはや馴化してしまっているので、CSに対する反応も減じてしまうことが予想できる。Rescorlaの実験では、後者のS–S学習仮説が支持される結果となった。他の実験からも、古典的条件づけが刺激と刺激の関係性についての学習であることを支持する結果が得られている[5]。

古典的条件づけは以上で見たきたように、複数の刺激を対提示する実験手続きである。学習の測度として反応を計測するものの、学習プロセスそのもの反応は入ってきていないという点に注意しよう。どういうことかというと、Pavlovの犬はメトロノームCSに対して唾液分泌というCRを表出するが、唾液分泌をしようがしまいが、餌USは提示されるのである。つまり、USの出現はCRに依存しない。皆さんも、深夜にステーキの画像を見て涎を垂らしたところで、ステーキが手に入るわけではないということである。次に登場する道具的条件づけ (instrumental conditioning) は、反応に依存して生じるタイプの学習である。それでは、道具的条件づけについて見ていくことにしよう。

1.3 道具的条件づけ

道具的条件づけの代表的な先駆者はEdward Thorndike[6]で、彼はヒヨコ、ネコ、イヌを問題箱の中に閉じ込め、そこからの脱出する行動を測った。彼の問題箱には十数種類のバリエーションがあったが、いずれも仕掛けによって扉が開き、脱出できるようにできていた。繰り返し問題箱から脱出する経験を積むことで、不要な行動が減り、脱出につながるような行動 (例えば「留め具に脚をかける」など) が増えていき、結果として脱出にか

[4] 歴史的には行動主義の創始者のJ. B. WatsonはS–R学習を想定していたが、Pavlov自身は刺激–刺激関係の学習が生じるというS–S学習を主張していた。動物研究の歴史についてはボークスの『動物心理学史』が詳しい。

[5] ただし、S–R学習が存在しないのかというと、そうではない。

[6] Thorndikeは日本人女性で初めて博士号を取得した原口鶴子の指導者でもあった。原口は帰国後、フランシス・ゴールトンの『天才と遺伝』の翻訳などに携わっている。

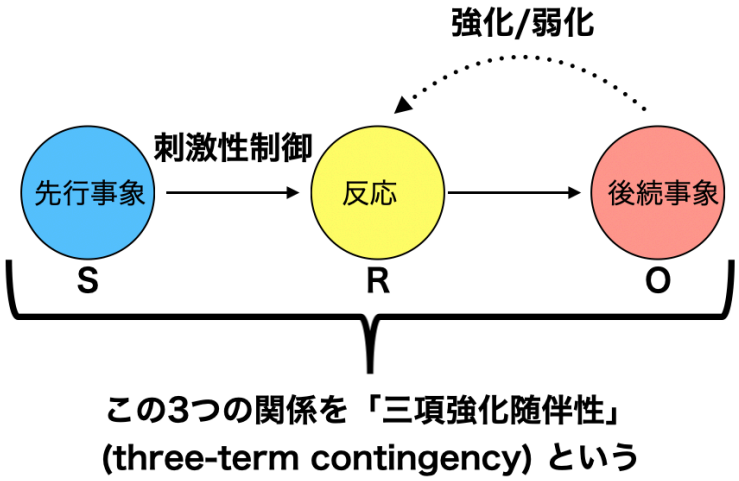

図2. 三項随伴性

かる時間は短くなった。学習が生じたということである。Thorndikeはこのような学習を効果の法則 (law of efect) と呼んだ。効果の法則は、現代では道具的条件づけと呼ばれる学習である[7]。

用語について少し補足をしよう。行動分析学で道具的条件づけに相当するのがオペラント条件づけ (operant conditioning) である。 オペラント条件づけと対になる言葉はレスポンデント条件づけで、行動分析学における古典的条件づけに相当する言葉である。行動分析学と学習心理学は、共通して扱う現象が多く、深い関わりのある分野であるが、用語法や説明の仕方、用語の運用上の射程が異なる[8]。従って、オペラント条件づけ/レスポンデント条件づけと道具的条件づけ/古典的条件づけを混在させるのは好ましくない。本稿では、後者の区分で議論を進める。

さて、古典的条件づけと道具的条件づけの手続き的な違いは何か？先述したように、古典的条件づけで提示されるUSは、行動に依存していない。反応を行おうが、行わないが、後続事象 (consequence) としてUSは来るのである。一方で、道具的条件づけの場合、事情は違う。Thorndikeの実験を例に取ると、問題箱から脱出できるという後続事象は、留め具に手をかけるという行動が起きるがどうかにかかっている。つまり、結果が反応に依存しているというわけだ。行動が起きるのに際しては、反応前の環境側に刺激も存

[7] ただし、Thorndike自身の効果の法則の理論は、現代の道具的条件づけの考え方とはかなり異なる。ThorndikeはS–R連合が、行動の結果生じる「満足」によって刻みこまれる (stamp-in) のであると考えた。逆に、「満足」を引き起こさないS–R連合はかき消される (stamp-out) されるという次第である。また、S–R連合の捉え方も独特で、身体運動が刺激と結びついた「衝動」(impulse) によって駆動されるという見方をしていた。

[8] とはいえ、この一コマではこれらの区別が決定的に重要になる現象は出てこない。詳しく知りたい人は澤 (2021) を参照するとよい。書籍を参照するのが億劫な人向けには、拙ブログで簡単にまとめてある。 https://heathrossie-blog.hatenablog.com/entry/2021/12/26/152410

在する。問題箱の場合は、留め具やその他取り付けられた器具がそれに相当する。動物は、先行刺激がある下で反応を行い、後続事象を受け取る。それにより、特定の行動を増やしたり減らしたりするということだ。このように、行動を行うに際して環境側から提示されている刺激のことは、弁別刺激 (descriminative stimulus) と呼ばれる。以上のような見方をすれば、道具的条件づけは弁別刺激–反応–後続事象の3者の関係で捉えられそうである。そこで、この3者関係を三項随伴性 (three-term contingency) という (図2)。学習心理学ではよく、S–R–Oと表記する。後続事象の提示により行動の生起頻度が増えることを「強化」(reinforcement) と呼ぶ。逆に行動が減ることを「弱化」(punishment) と呼ぶ。強化的な刺激のことを強化子、弱化を引き起こす刺激を弱化子という。ここからわかるように、道具的条件づけは後続事象と反応の関係性の学習である (R–O連合)[9]。

強化/弱化の定義は、行動の生起頻度の増減で決まっているという点が重要である。例えば、食物は動物にとって代表的な強化子で、怒鳴りつけることは子どもにとって弱化子であるように思われるかもしれない。しかし、空腹でなかったり、選好の低い食物であったりしたら、動物の行動は増えない可能性がある。ということは、定義上、その食物は強化子ではない。子どもが好ましくない行動をして怒鳴りつけた場合も同様である。子どもがそれにより、標的としている行動 (例えばいたずら) が減ったのならば弱化が生じているが、一向に行動が減らなかったり、逆に増えることもありうる。怒鳴りつけることで注目を集めて、それが強化的に働くこともあるためだ。そのとき、怒鳴りつけることは子どもに弱化子として機能していないということだ。

B. F. Skinnerはスキナー箱を開発して、道具的条件づけの研究を一気に推し進めた (Ferster, & Skinner, 1957; Skinner, 1956)[10]。典型的なラット用のスキナー箱では、反応用のレバーと強化子提示用のフードディスペンサー、刺激提示用のライトやスピーカーが備え付けられている。学習実験ではラットの他にハトもよく用いられるが、ハトのスキナー箱ではレバーの代わりにつつき反応を取るためのキーを取り付けることが多い。最近は、タッチスクリーンを用いたスキナー箱もよく用いられる。ラットのレバー押しを例に考えてみよう。レバーを押したら、強化子として餌が出てくるわけだが、どういう回数、あるいはタイミングでレバーを押せば強化を受けるかを決める規則のことを「強化スケ

---

[9] ここでは割愛したが、弁別刺激が行動に影響することは刺激性制御 (stimulus control) という。また、弁別刺激と結果事象の関係 (S–O連合) の研究も存在する。

[10] 学習心理学ではスキナー箱と呼ばれるが、Skinner自身はこの呼び方が好きではなかったらしい。行動分析学では、「オペラント箱」(operant chamber) と呼ばれる。しかし、オペラント箱だと先程のオペラント/道具的の区分の「用語の混ぜるな危険問題」に直面する。そもそも、スキナー箱は道具的行動だけじゃなくて古典的条件づけの研究にも使えるという事情もある。Skinner自身は単に「実験箱」(experimental chamber) と呼んでいたが、この呼び方は21世紀で使うには一般的すぎる。

ジュール」(reinforcement schedule) という。毎回強化を受けるような事態は連続強化 (continuous reinforcement schedule) で、必ずしも毎回強化子が提示されるわけではないスケジュールを間欠強化スケジュール (partial reinforcement schedule) と呼ぶ。間欠強化スケジュールに膨大な種類があり、それぞれで反応の出方が異なる。

強化スケジュール研究では様々な強化スケジュールで動物を強化し、反応のパターンを分析する。代表的な強化スケジュールとして、比率スケジュール (ratio schedule) と時隔スケジュール (interval schedule) がある。比率スケジュールは、決められた回数反応すれば、強化子が提示されるスケジュールである。その回数が固定の回数のときは固定比率スケジュール (fixed ratio schedule; FRスケジュール) といい、毎回異なった回数の反応が必要な場合、変動比率スケジュール (variable ratio schedule; VRスケジュール) という。連続強化は毎回1回反応すれば強化子が提示されるので、FR1スケジュールである。皆さんがパチンコを打つとして、パチンコは打ち続ければいつかは当たるが、何回目で当たるかはわからない。従って、VRスケジュールと見なすことができる。

一方、時隔スケジュールは、決められた回数ではなく、決められた秒数の経過後、1回反応すれば強化が得られる。その秒数が固定されているとき、固定時隔スケジュール (fixed interval schedule; FIスケジュール) という。例えば、毎日正確な時間にくる電車を待つ行動は、FIスケジュールと見なせる。「そろそろ来たかな」とスマホから顔をあげても、到着前であれば何度見ようと電車はまだこない。電車が来たタイミングで顔をあげて、初めて強化を受けることができる。FIスケジュールはその性質上、動物の時間知覚 (計時行動とも呼ばれる) を調べるためにも使われる (藤巻他, 2015)。時間間隔決まっていない時隔スケジュールは変動時隔スケジュール (variable interval schedule) と呼ばれる。VIスケジュールではラットの場合、何秒か経過した後1回レバーを押せば餌がもらえるわけだが、具体的に何秒後かはわからない。例えばVI30秒スケジュールというときは、平均30秒後に1回反応すれば強化を受けることができる。VIスケジュールは、VRスケジュールより平均的な反応率は低いものの、安定して反応が生起するスケジュールであることが知られている。例えば皆さんに気になる相手がいるとして、その人からの返信メッセージを待っているとしよう。スマホを小まめに確認したところで、返信が早くなるわけではない。つまり、反応の頻度を増やしても強化が早く得られるわけではない。返信が来たときに開いて初めて、強化を受けるわけだ。これがVIスケジュールである。この講義では、後ほどVR・VIスケジュールを用いた選択行動について理論的な考察を行う。

以上で、道具的行動の基本を紹介した。次節からようやく本題に移る。まずは、古典的条件づけの研究文脈で発展した連合学習理論を紹介する。

2. 基本的な連合学習理論

表2. 阻止現象の実験デザイン (抜粋)

	Phase 1	Phase 2	Test	Result
Blocking	CS1→Shock	CS1 & CS2→Shock	CS2?	weak CR
Control	-	CS1 & CS2→Shock	CS2?	CR

古典的条件づけが刺激と刺激の関係性の学習であるということを前節では見てきた。本節では、連合学習における連合の更新則について見ていく。とはいえ、いきなり大上段から理論を与えていくのは好ましくない。具体的な現象に寄り添いながら、進めていこう。

2.1 古典的条件づけと接近性

古典的条件づけにおいて、刺激と刺激の時間的な接近性 (contiguity) は学習の成立に極めて重要である[11]。例えば、痕跡条件づけ (trace conditioning) は、CSとUSの間に時間的な空白期間を挿れる。CSが出現して、それが消えてから数秒後にUSが到来するということだ。痕跡条件づけは、通常の古典的条件づけ (延滞条件づけという) よりも学習が成立しづらい、あるいは成立が遅くなることが知られている。

接近性が一見欠けているのにの関わらず学習が成立する現象もある。味覚嫌悪学習と呼ばれる古典的条件づけの現象である (taste aversion learning; Garcia et al., 1955)。CSとして味覚を用いて、USとしては塩化リチウムを注射する。塩化リチウムは嘔吐反応や腹痛を引き起こす毒物として利用できる[12]。これらのCS–USの対提示を行うと、CSとして用いた味覚刺激への選好性が減少する。問題は、この学習はCSとUSの間隔が数時間、場合によって日を跨いでも成立するという点である。先程の痕跡条件づけでは、典型的には数秒の間隔でも学習が成立しづらくなるのに比べたら、かなり長い時間が空いても学習が成立するように見える。そのためもあって、当初発見者のGarciaは味覚嫌悪学習が古典的条件づけとは別のプロセスであると考えていた。しかし、味覚嫌悪学習があまりにも長いと成立しないことなどから、現在では古典的条件づけの現象の1つであるとみなされるようになった。接近性というのはあくまで相対的なもので、用いる刺激によって適切な接近具合というのは変わってくるのである。

では、接近性が連合形成のための十分条件なのだろうか？そうであると考える理論家もいる[13]が、スタンダードな学習理論では接近だけでは不十分であると考える。それを示す

---

[11] ここでは出てこないが、空間的接近性も重要である。

[12] ただしラットでは咽頭の構造上、嘔吐は生じない。

[13] 例えば、Ralph Millerという研究者は、時間的接近性が学習の十分条件であると考えるが、実際の行動として表出される「遂行」(performance) の部分を分けて捉えている。

最たる例が、阻止現象 (blocking) という現象である。阻止現象はKamin (1968) の恐怖条件づけの実験によって発見された。実験の重要な部分を表2にまとめた。阻止現象の実験では、2つのCSが用いられる (例えば光とノイズ音)。まず、阻止群の動物にCS1と電気ショックUSを対提示する。これは通常の古典的条件づけである。CS1-USの学習が完了後、CS1とCS2の複合刺激 (compound stimulus) とUSを対提示する。一方、統制群では最初のフェーズは行わず、最初からCS1・CS2とUSを対提示する。テストでは、CS2を単独提示し、CRが生じるかを計測する。すると、阻止群の動物は統制群と比べてCRが減弱する。つまり、最初のフェーズでのCS1–USの学習が、CS2–USへの学習をブロックしたということである。これが阻止現象である。

ごくごく日常的な例で阻止現象を考えてみよう。皆さんは今日、真面目に講義に出ているわけだが、友達Aがサボっているのを観察したとする。きっと皆さんは「あいつは不真面目だなぁ」などと思うことだろう。別の日に、その友達Aと他の友達Bが講義をサボっているのを目撃したことを想像してほしい。どう思うだろうか？中には「きっと友達BはAにそそのかされたに違いない」と考える人もいるかもしれない。もし、そう思ったとしたら、友達Bへの学習 (「あいつは不真面目だなぁ」と評価すること) は友達Aへの事前学習によってブロックされたわけだ。阻止現象は明らかに刺激と刺激が時間的に接近しているのにもかかわらず、学習が妨害される。この現象を、連合によってどのように説明したらよいだろうか？

2.2 Rescorla-Wagnerモデル

Robert RescorlaとAlan Wagnerが発表した連合学習のモデルである (Rescorla, & Wagner, 1972)。Rescorla-Wagnerでは、CSとUSの間の結びつきの強さを「連合強度」(associative strength) とし、反応の強度を表す量として捉える。つまり、CSとUSの間の連合強度が強いほど、高確率で (あるはより高い強度で) CRが表出されるということだ。具体的には連合強度 V が次のように更新されることを仮定する。

$$\Delta V_t = \alpha(\lambda - V_t) \quad \cdots (1)$$

ここで、添え字 t は試行を表している。α はCSの顕著性 (salience)である。 λ はUS強度で学習の漸近値を示す。顕著性というのは、明瞭度ともいう。顕著性は学習の進む速度を制御しているパラメータになっているので、刺激が明瞭 (例えば、明るい光や大きい音) の方が一回で進む学習量が大きいことを示している。ΔV_t というのは試行 t での連合強度の増分であるので、次の試行での連合強度は次の式のようになる。

$$V_{t+1} = V_t + \alpha(\lambda - V_t) \quad \cdots (2)$$

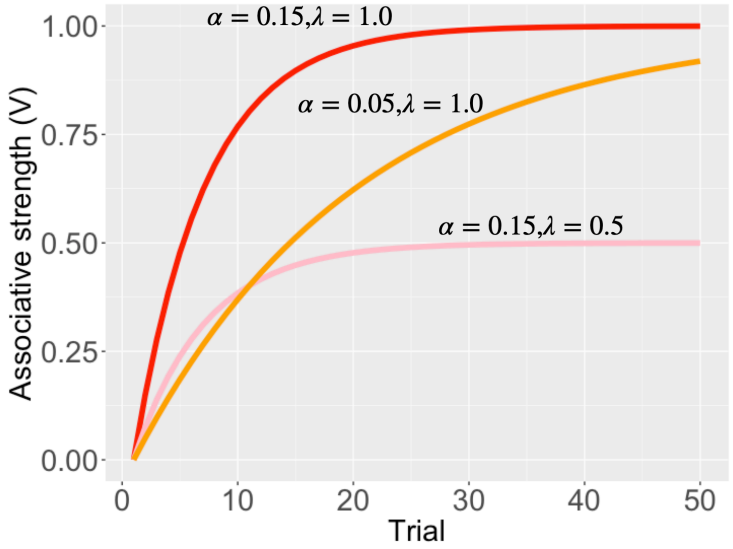

図3. Rescorla-Wagnerモデルの挙動

また、β をUSの明瞭度として別のパラメータとして用意し、次のように表すこともある。

$$V_{t+1} = V_t + \alpha\beta(\lambda - V_t) \quad \cdots (3)$$

この式を見て、どんな含意があるか直観的にわかる人は大変優秀でよろしいことなのだが、大抵の人はそうではないだろう。私もそうだった。こういうときは、まずは適当な値を入れてグラフを見てみるのが得策だ。$\alpha = 0.15$、$\lambda = 1$、$V_0 = 0$ を式 (2) に代入したものが、図3である[14]。連合強度が試行を経るごとに徐々に上昇しているが、上昇幅は少しずつ下がっていき、後半はほとんど変化していない様子が見て取れるだろう。このような曲線は負の加速度を持つ曲線 (negatively acceleralated curve) と言ったりする。収束に向かっている値は、赤線の場合1であることが縦軸を見ればわかるだろう。これはUS強度 λ の値である。オレンジ線も λ は1であるため、1に向かって学習が進んでいるように見える。一方、顕著性 α が赤線より低く、学習がゆっくりと進んでいる。赤線とピンク線を比べると、顕著性 α は同じだが、US強度 λ が異なる。その結果、連合強度の漸近値がピンク線で低くなっているのがわかるだろう。なお、値は完全に恣意的で、V に特に単位はない。連合強度という抽象的なものを反映した値であるからであるが、お望みであれば、例えばCRの生起確率と読み替えることもできる。

Rescorla-Wagnerモデルがどんな動きをするのかが把握できたところで、式 (1-3) を吟味してみよう。まず、これらの式は V_t を V_{t+1} に更新する式になっている。つまり、試行ごとの更新則となっているわけだ。どれくらい更新するのかは、$\alpha\beta$ と$\lambda - V$ の積の大き

[14] 自分でも試しに描いてみるのがよいだろう。エクセルで試すこともできる。

さ次第になっている。α と β は、それぞれCS、USの顕著性であるため、物理的に固定された定数である[15]。従って、学習の進行具合は $\lambda - V$ によって決まることがわかる。この $\lambda - V$ には「予測誤差」(prediction error) という名前が付けられている。連合強度 V は、CSとUSの結びつきの強さであるため、いわばCSがどれくらいUSの到来を強く予測しているのかを表している。一方、λ は実際のUSの強度である。そこの差分をとっているので、この量は予測がどれくらい外れたのかを表現しているのだと解釈できる。文献によっては、 $\lambda - V$ が予測誤差を反映していることを強調するために、次のように表すこともある。

$$\delta = \lambda - V \cdots (4)$$

以上の説明をまとめると、(1-4) 式は、予測誤差によって連合強度を更新し、その更新分が刺激の顕著性によって定まるということを言っているのがわかっただろう。さて、このモデルで阻止現象を説明するには、さらにもう1つ追加的な仮定が必要になる。それは複数の刺激が提示されているとき、連合強度の計算は各刺激の連合強度を用いるという仮定である。このことをあからさまに書くと、n 個の刺激について、連合強度の更新則は次のように表現できる。

$$\Delta V_t = \alpha(\lambda - \sum_{i=1}^{n} V_t) \cdots (5)$$

変わった場所は、式 (1) のV_t が$\sum V_t$ になっている点だ。例えば、阻止現象の実験のように2つのCSがあるとき、$\sum V_t = V_{1,t} + V_{2,t}$ である。ようは2つの刺激の連合強度の足し算を計算して、予測誤差を計算することになる。

式 (5) を使って、阻止現象を説明してみよう。阻止の実験では、最初にCS1とUSの対提示が行われる。CS2は提示されていないので、当然CS1のみへの連合強度が上昇する。十分な回数対提示されると、CS1はUSに対して十分な連合強度を獲得する。それに伴い、予測誤差は0に近づいていくはずである。その後、CS1とCS2の複合刺激とUSの対提示が行われる。このとき、予測誤差はどうなるだろうか？例えば $V_{1,t} = 0.9$、$V_{2,t} = 0$ であるとしよう。CS1とUSは既に対提示されているから、十分な連合強度を獲得しており、CS2はいまだ一度も対提示されていないので連合強度が0であるということを言って

[15] ただし、最近のモデルでは物理的な定数としての顕著性だけではなくて、「獲得された顕著性 (acquired salience)」という考え方も提案されている (Esber, & Haselgrove, 2011)。

表2. 潜在制止の実験デザイン

	Phase 1	Phase 2	Test	Result
Latent inhibition	CS	CS → US	CS?	weak CR
Control	-	CS → US	CS?	CR

いる。それらの総和を足すと $\sum V_t = 0.9 + 0 = 0.9$ となる。$\lambda = 1$、$\alpha = 0.1$ でCS1、CS2が獲得する連合強度 ΔV を計算すると、

$$\Delta V = 0.1 \times (1 - 0.9) = 0.01$$

である。つまり、ほとんど学習が生じないことがわかる。これがなぜ起きるかは簡単にわかる。最初の実験フェーズでCS1が十分な連合強度を獲得している。すると、CS1・CS2の複合刺激を提示しても、予測誤差がほとんど生じなくなる。予測誤差が生じなければ、学習は生じないと、Rescorla-Wagnerモデルは考える。これにより、CS2は連合強度をほとんど獲得できず、阻止現象が生じるというわけだ。このように阻止現象は、一見、連合で説明するのが難しそうな現象であるが、Rescorla-Wagnerモデルの仮定するような連合の規則であれば、自然に説明ができる。

2.3 潜在制止とMackintoshの注意理論

前節では、阻止現象を説明可能なモデルとしてRecorla-Wagnerモデルを紹介した。しかし、Recorla-Wagnerが説明できない現象も数多く存在し、その代表例として潜在制止が挙げられる (latent inhibition; Lubow, & Moore,1959)。潜在制止は、当初ヤギで発見された現象である。再び、基本的な手続きを表2にまとめた。Lubow and Moore (1959) は音 (あるいは回転ローター) をCS、電気ショックをUSとして使用した。潜在制止群は、初めにCSの単独提示を行う。USは提示せず、CSだけを動物に提示するということだ。その後、CSとUSの対提示を行う。統制群では、CSの単独提示を経験させない。すると、潜在制止群のCRは統制群よりも低い水準に留まる。このように、CSの先行経験がその後の学習を阻害する現象が、潜在制止である。

潜在制止は、おそらく多くの人が経験している。2.1節で紹介した味覚嫌悪学習について「いや、でも私は牡蠣にあたったことがあるけど、味覚嫌悪学習は起きなかったし、今でも牡蠣が大好きだ」のように感じなかっただろうか？通常、食物にあたっても直ちに嫌

いになることはあまりないだろう。これには潜在制止が関係している可能性がある。不運にも牡蠣にあたっても、それまでに何度も牡蠣を楽しんでいれば、潜在制止が働きCR (味覚嫌悪) が生じにくいというわけだ。

上述の通り、潜在制止はRescorla-Wagnerモデルでは説明ができない。式 (1) を見ると、Rescorla-Wagnerモデルは予測誤差が生じたときに学習が進むという仮定を置いてあることがわかる。つまり、CSとUSが対提示されたときに初めて、学習が生じるということだ。潜在制止は、CSを単独で提示することで生じる学習である。従って、Rescorla-Wagnerモデルの射程外となることがわかる。

潜在制止を説明する方法は複数あるが、メジャーな説明は注意 (attention) という概念を付け加えることだ。ここでいう注意とは、CSに対する注意である。代表的な注意理論に、Mackintoshの注意理論がある (1975)。Mackintoshの潜在制止の説明は、次の通りである。注意には刺激を処理し、学習を進める機能がある。潜在制止の実験では最初にCSが単独で提示される。CSは特に生物学的重要さを持たない刺激である。従って、CSへの注意が減少する。その結果、次のフェーズでCSとUSが対提示されても、CSへの注意が向けられていないため、学習が生じにくくなる。このことを式 (1) を改変することで表現してみよう。

$$\Delta V_t = \alpha_t(\lambda - V_t) \cdots (6)$$

式 (6) は式 (1) からあまり大きい変化があるように見えないが、α が α_t となっている。Rescorla-Wagnerモデルの α はCSの顕著性であったが、Macintoshのモデルでは α_t は注意量という概念になっている。添字の t がついていることからわかるように、試行ごとに注意は更新される。注意量の更新則は式 (7) で定める。

$$\alpha_{t+1} = \alpha_t + \Delta\alpha_t \cdots (7)$$

ただし

$$\Delta\alpha_t \; is \; positive \; if \; |\lambda - V| < |\lambda - V_X|$$

$$\Delta\alpha_t \; is \; negative \; if \; |\lambda - V| \geq |\lambda - V_X|$$

V はCSとUSの連合強度である。加えて、V_X という新たな文字が入っているが、これはCS以外の環境にある刺激とUSとの間の連合強度である。CS以外の刺激というのが何かというと、例えば実験箱の中の壁や床、ルームライトといった文脈刺激が相当する。t 試行目の注意の変化量 $\Delta\alpha_t$ は、CSと文脈刺激が持つ連合強度から計算される予測誤差の差

によって決まるということを、式 (7) は言っている。もし、CSの方が文脈刺激よりも予測誤差が小さいのならば注意量は上昇し、逆も然りである。予測誤差が小さいというのは、その刺激がUSを予測する力が高いということである。Mackintoshのモデルでは、CSの予測力が文脈刺激より高いとき、そのCSへの注意が上昇することを仮定している。つまり、より予測の良い刺激に対して注意を向けるというのがMackintoshの注意概念である。注意量が上昇すると、式 (6) に従ってより素早く学習が進むことが予想できる。

Mackintoshのモデルは潜在制止を説明することができる。まず、潜在制止ではCSが単独で提示される。つまり、この時点ではCSはUSの到来を予測する刺激足りえないということだ。すると、式 (7) に従ってCSへの注意量は減少する。結果、その後CSがUSと対提示されても学習が生じにくくなるというわけだ。

2.4 負の転移現象とPearce-Hallモデル

Recorla-Wagnerモデルは阻止現象を説明できるものの、潜在制止が説明できない。潜在制止の説明には注意概念が必要になることを、Mackintoshのモデルを例に見てきた。Mackintoshの考える注意は、予測力のあるCSに対して注意が向けられるというものであった。例えば、信号は道路を渡ってもよいかをシグナルし、そのような刺激に対して注意が向けられるのはある意味当然のことでもある。しかし、日常で注意という言葉を使うとき、それ以外の使われ方があることは容易に想像がつく。簡単な例としては、あなたが道を歩いていて、急に大きい音がしたとしよう。はっとなってそちらを見たら、道路の工事中であったようだ。次からその道を通るときは、おそらく注意を向けることになるだろう。このように、予期しなかった刺激に邂逅したことにより、そちらに注意を向けるという事態は日常的に存在する。これは、認知心理学の「ボトムアップの注意」に相当するような注意である。

そのような意味での注意をモデル化したのが、Pearce-Hallモデルである (Pearce, & Hall, Pearce, & Hall, 1980)。Pearce-Hallモデルは、予測誤差を大きく生み出す刺激に対してより多くの注意を向けると予測する。Pearce-Hallモデルの連合強度の更新則は次の通りである。

$$\Delta V = S\alpha_t\lambda \cdots (8)$$

ここで、S はCSの顕著性で、λ はUSの強度である。α_t が注意量であり、次式で更新される。

$$\alpha_{t+1} = |\lambda - V_t| \cdots (9)$$

$\lambda - V_t$ は予測誤差であるので、式 (9) は予測誤差の絶対値である。USが提示されたとき ($\lambda > 0$) は正の予測誤差が生じ、USが提示されなかったとき ($\lambda = 0$) のときは負の予測誤差が生じうるわけだが、いずれの場合でも、絶対値が大きいときに注意が上昇することをPearce-Hallモデルは仮定している。学習が進み、予測誤差がなくなったら式 (8) はほぼ0になるので、そこで学習は終了する。

式 (6) と式 (8) を見比べたらわかるように、Pearce-Hallモデルの注意は、ちょうどMackintoshのモデルと逆の仮定になっている。Mackintoshのモデルは、より予測力の高いCSに注意を向けると考えるが、Pearce-Hallモデルは予測誤差を生み出すような刺激に対して注意を向けると仮定している。これはトップダウンの注意とボトムアップの注意をそれぞれ反映していると考えれば、理解がしやすいかもしれない。では、Mackintoshのモデルでは説明できなくて、Pearce-Hallモデルでは説明できる学習現象は存在するのだろうか？代表的なものとしてはHall and Pearce (1979) が発見した負の転移 (negative transfer) という現象がある。

負の転移は、潜在制止の亜種の現象で、次のような手続きで生じる (表3)。負の転移群では、最初にCSと弱いUSが対提示される。Hall and Pearce (1979) で使用された弱いUSとは、アンペア数の小さい電気ショックである。条件づけの成立後、今度は同じCSと強いUSを対提示する。すると、一度も同じCSで弱いUSと強いCS両方の対提示を受けなかった統制群と比べて、CRが減弱するのである。表3を見ると、統制群は最初のフェーズでは別のCS (CS2と表記してある) と弱いUSが対提示されている。これは、強度が弱めのUSを経験する回数を負の転移群と揃えるための統制である。

日常的な例では、皆さんの英語の先生Aさんは、とても怒りっぽくてほんの小さな間違いでもすぐに声を荒げてくる先生だとしよう。なんとも嫌な先生であるが、そういう先生を想像してもらったとして、その常々怒りっぽいA先生がある日、激憤慷慨のとてつもない怒り方をしたとしよう。どうだろうか？確かに、怖いかもしれないが、「まぁ、この人は怒ってるしなぁ」と思うのではないだろうか？ここでもう一人、別の先生に登場してもらおう。皆さんの国語の先生Bさんは、決して声を荒げない穏やかな人だとしよう。その人が、ある日皆さんの誤りに怒髪天のように怒ったとしよう。きっと皆さんはこう思うこ

表3. 負の転移の実験デザイン

	Phase 1	Phase 2	Test	Result
NT	CS1-Shock	CS1-Shock!!	CS1?	weak CR
control	CS2-Shock	CS1-Shock!!	CS1?	CR

とだろう。「この先生をここまで怒らせるとは・・・きっとよほそまずいことをしたに違いない」。以上の例え話の中で、B先生に生じた学習の阻害が、まさに負の転移というわけだ。

Rescorla-WagnerモデルでもMackintoshのモデルでも、負の転移は説明できない。Rescorla-Wagnerモデルの場合、弱いUSでもCSと対提示されることで学習は進むので、むしろCRが獲得されやすくなることを予測する。Mackintoshのモデルでも、最初に弱いUSと提示されることで、式 (7) に従って注意量が上昇する。結果、Rescorla-Wagnerモデルと同様にCRがむしろ獲得されやすくなることを予測してしまう。

対照的に、Pearce-Hallモデルなら負の転移現象をうまく説明できる。なぜなら、最初のフェイズでCSが弱いUSと対提示され十分に学習された場合、強いUSと提示されたときの予測誤差が小さくなるためだ。USと一度も対提示されてこなかった刺激が初めてUSと対提示されたときは、大きな予測誤差が出てくる。これは統制群で生じることである。一方で、事前に弱いUSと対提示を経験している負の転移群では、多少なりとも連合が既に形成されおり、予測誤差の減少を招く。 Pearce-Hallモデルは予測誤差がそのまま注意量になっているため、予測誤差が小さい刺激に対しては学習があまり進まないことを予測する。その結果、負の転移現象が生じるということだ。

以上、連合学習理論の入門編として代表的な3つのモデルを紹介した。それに合わせ、各モデルと関わりの深い古典的条件づけの現象も紹介した。ここで紹介したモデルは、いずれも予測誤差を用いていて、それを低減させていく過程を表現している。一方で、連合強度とそこから計算される予測誤差という量は、やや天下りに与えられて、ひとまずそれを最小化していくことを是として話を進めてきた。次の話題である道具的条件づけの理論的な考察では、最適化していく量についてもう少し緻密に見ていくことにする。

3. 対応法則と最大化

前節では、連合学習理論の基本的なモデルを代表的な古典的条件づけの現象とともに紹介した。今度は道具的条件づけについて数理的な考察を行う。扱う現象は、道具的条件づけの最重要現象の1つである対応法則 (matching law) である。

3.1 並列スケジュールと対応法則

対応法則はHerrnstein (1961) の並列スケジュール (concurrent schedule) の実験で見つかった。並列スケジュールとは、2つ以上の強化スケジュールが同時に走っているタイプの強化スケジュールで、選択行動の実験で広く用いられる。典型的には、図4のように2つの反応キー下で反応が計測される。この図はハトのスキナー箱を描いたものなので、反応キーが2つついていて、別の色で照らされている。これらのキーそれぞれで、強

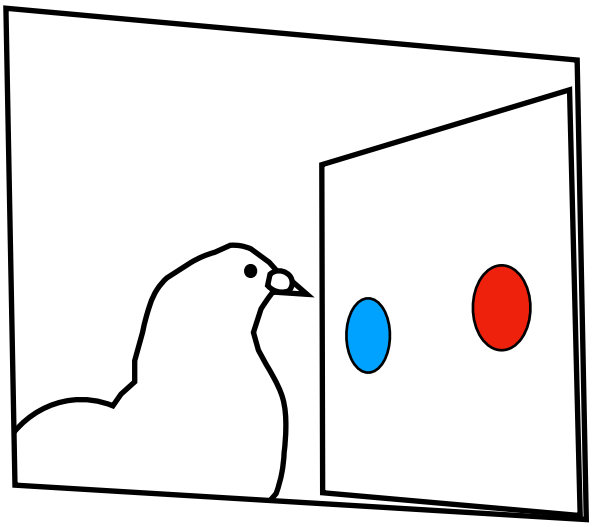

図4. 並列スケジュールを遂行するハト

化スケジュールが当てられているというわけだ。並列VI–VIスケジュールなら、左右両方のキーに別個にVIスケジュールが当てられている。例えば、右のキーにはVI30、左のキーにはVI60スケジュールが走っているといいう具合だ。この場合、右キーからは平均的に1分間に2回強化を得ることができて、左キーからは平均1回の強化が得られる計算になる。つまり、平均的に強化を得られる割合は2：1となる。

並列VI-VIスケジュールの強化率をさまざまに変えてみたら、反応の配分率はどうなるだろうか？Herrnstein (1961) は、左右のキーへの反応率がちょうど強化率の比に一致することを見出した (図5)。図5を見てみると、キー1から得られる強化の割合が50%のときというのは、言い換えるとどちらも同じVI値が設定されていることを示している。このとき、縦軸も50%になっているが、それはすなわち反応の50%を1つ目のキーに配分していることを示している。強化率の比に、反応の比が対応しているということだ。図5の対角線に当たる点線が完全なマッチングに相当しているが、実際のデータもそれに近い値をとっていることがわかる。このように、強化率の比と行動の比が対応することをHerrnsteinは対応法則と呼んだ。

具体的には、Herrnstein (1961) は反応率–強化率比の関係を次のように記述した。

$$\frac{B_1}{B_1+B_2}=\frac{R_1}{R_1+R_2}\quad\cdots(10)$$

B は反応率、R は強化率を表している。この式は前段落と同じことを言っている。反応の比が強化率比と対応しているということだ。対応法則には様々なバリエーションがあり、例えば行動の種類が3つ以上あるとき (n 個のとき)、対応法則は次のように書き直せる。

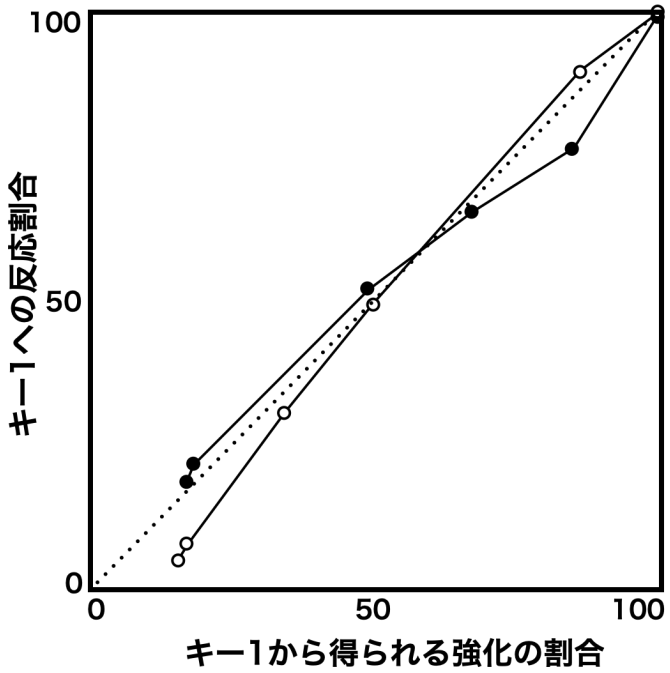

図5. マッチング実験の典型的な結果 (Herrnstein, 1961から作成)

$$\frac{B_i}{\sum_{j=1}^{n} B_j} = \frac{R_i}{\sum_{j=1}^{n} R_j}$$

対応法則が成立するのは、強化率だけではない。同じスケジュールでも、反応キーごとに強化量が異なる場合でも対応法則が成立することがわかっている (Baum, & Rachlin, 1969; Killeen, 1972)。例えば、同じVIスケジュールでも右キーからは左キーの2倍の餌が出てくるといった事態である。A を強化量としたとき、強化量の対応法則は

$$\frac{B_1}{B_1 + B_2} = \frac{A_1}{A_1 + A_2}$$

と記述できる。B は反応率であったが、行動に従事する時間 T でも対応法則は成立することがわかっている (Baum, & Rachlin, 1969; Baum, 1976)。この場合は、次のように式 (10) を書き直せばよい。

$$\frac{T_1}{T_1 + T_2} = \frac{R_1}{R_1 + R_2}$$

と、いろいろなもので対応させられることがわかっている。種間の一般性という点ではどうだろうか？対応法則は、当初ハトで発見された現象であるが、ヒト、ラット、サルでもよく研究がなされている。他の種にも拡張される試みが行われており、上述の動物以外の

哺乳類、鳥類、魚類、昆虫の研究がある。ただし、今のところ爬虫類、両生類、並びに昆虫以外の無脊椎動物では報告例がないようだ。これらの分類群を除けば、対応法則は広い範囲の種で見られる現象である。

典型的な対応法則の実験では、2つの反応キーが備え付けられた並列スケジュールを用いる。これは日常的な例で言えば、みなさんがお昼ご飯に学生食堂にいくか、ラーメン屋にいくかの選択を迫られているような状況だ。しかし、こういうあからさまな選択場面ではなくても、人生のあらゆる行動は選択であると見なすこともできる。何も自己啓発的なことを言おうとしているわけではなくて、みなさんがこの科目のレポートを書く行動をするときだって、YouTubeを見る、パートナーと出かける、昼寝をするといった他のあらゆる行動からレポートの執筆を選択しているわけだ。つまり、現在標的としている行動と、他行動との間の選択を行っていると見なすことができる。

単一のVI強化スケジュールの下で反応を遂行する動物でも、同じことが成立しているのではないかと、Herrnstein (1970) は考えた。すなわち、ハトはスキナー箱の中でキーつつきとそれ以外の他行動 (羽繕いや箱内の探索など) との間の選択を行なっているというわけだ。もしそうであれば、キーつつきと他の行動の間で対応法則が成立する可能性がある。

$$B_1 = k\frac{R_1}{R_1 + R_o} \cdots (11)$$

B_1、R_1 はそれぞれVIスケジュールが走っている反応キーへの反応率と強化率である。k は比例定数で、R_o は他行動の強化率である。R_1 は実験者が設定するものであるため、定数が入るが、他行動の強化率 R_o は動物が勝手に行う行動への強化に対応するため、未知の値である。k について少し補足すると、標的としている行動と他行動の対応法則を式 (10) 通りに書けば

$$\frac{B_1}{B_1 + B_o} = \frac{R_1}{R_1 + R_o}$$

となり、これを変形すれば

$$B_1 = (B_1 + B_o)\frac{R_1}{R_1 + R_o}$$

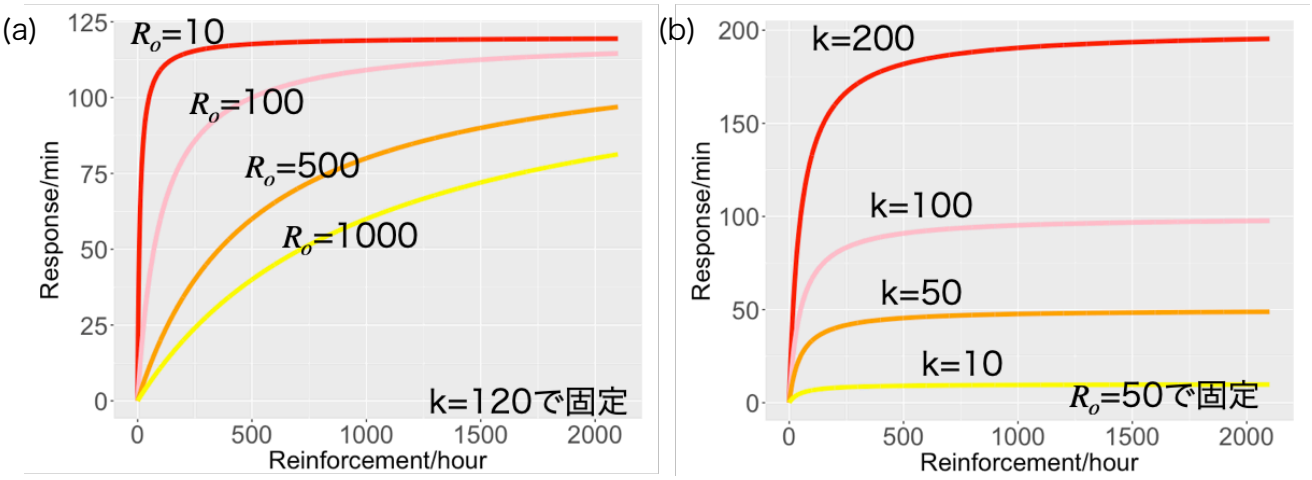

図6. 単一スケジュールの対応法則

である。$B_1 + B_o = k$ とすれば式 (11) である。つまり、k とは標的行動と他行動の総和になっていることがわかる。k を定数にするというのは、全体の行動の総和を固定していることになっている。式 (11) のパラメータを様々に振ってグラフを描いたものが図6である。図6aを見れば、他行動の強化率が高いとき、反応率が低く推移することがわかる。また、図6bは k の値を適当に振っているが、反応率の上限が変化している。VIスケジュールにおいて、実際に動物がこのような振る舞いを見せることを、Herrnstein (1970) は示した。

以上はいずれもVIスケジュールを使用したケースについて紹介してきた。VIスケジュールは強化子の提示が経過時間に依存するスケジュールであるが、反応率依存のスケジュールではどうなるだろうか？Herrnstein and Loveland (1975) はハトを対象に並列VR–VRスケジュールの反応の配分率を検証した。彼らの実験では、結果としてマッチングは見られなかった。図7の点線は完全なマッチングを表しているが、データを示す点はソニほとんど乗っかっていないことがわかるだろう。むしろ、強化率比が少しでも50%の位置からずれると、反応は一気に強化率の高い選択肢に偏って、値がほぼ0%か100%になっているのを見てとることができる。実はこの結果は直観的に理解できる。VRスケジュールは反応率依存で強化子が提示されるため、決められた回数反応しない限り、一向に強化を受けることはない。それゆえに強化率が高い方の選択肢に排他的に反応を行う方が合理的なのである。例えば、スロットの台を選ぶ場面を考えてほしい。ある程度複数のスロット台を経験して、どの台で賭けるのがよいのか掴めてきたら、当たりの確率が高い台で集中的にギャンブルに興じる方が有利であるに決まっている。並列VI–VIスケジュールの反応では見られなかったような排他的な選好が生じるのは、強化子の最大化という最適化においては合理的な行動となっているのである。

さて、強化子の最大化、あるいは最適化というキーワードが出てきた。式 (10) の対応法則は、これだけでは経験法則である。行動の配分がなぜ強化率比と一致するのかについて、式 (10) はなんの説明も行なっていない。対応法則は、行動の最適化の帰結なのだろ

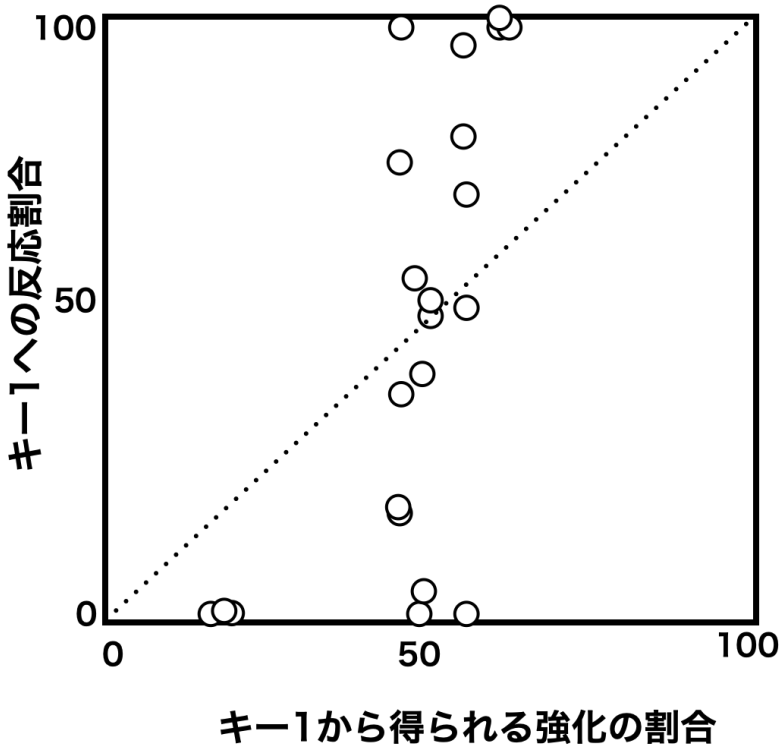

図7. 並列VR–VRスケジュールの結果 (Herrnstein, & Loveland, 1975より作成；簡便のため、複数の個体をまとめて描画している)

うか？この点をBaum (1981) が、各種のマッチング事態を対象にエレガントな説明を行なっている。そこで次節以降、Baum (1981) に乗っ取って、対応法則と最適化の関係について考えていこう。

3.2 最大化としての対応法則：並列VR–VRの場合

並列VR–VRで反応が片方の選択肢に対し排他的に配分されることが、強化子の最大化を導くというのは前節で直観的な説明を行なった。このことについて、もう少し形式的に考えてみよう。Baum (1981) ではまず、反応率と強化率のフィードバック関数 (feedback function) というものを仮定する。フィードバック関数とは、反応率に対し、時間あたりに平均的にどの程度の強化を得るかを特徴づける関数である。ちゃんと書こうとすれば、

$$R = f(B)$$

となるような関数 $f()$ である。VRスケジュールの場合のフィードバック関数は、いたってシンプルである。例えばVR30の場合、平均30回反応すれば強化子が提示される。つまり、1分あたり30回反応すれば1回の強化を受け、60回反応すれば平均的には2回の強化を受けるということだ。つまり、反応率に対して線形に強化率が上昇する。よって、VRスケジュールのフィードバック関数は、

$$R = cB \cdots (12)$$

となる。c は定数であるが、実際には強化率の逆数である。VR30なら、$c = 1/30$ である。こうすれば例えば、$B = 30$ のときに、$R = \frac{1}{30} \times 30 = 1$ となり、平均的に1分間あたり1回の強化を受けることになる。よって、c は実験者が決める値であることがすぐにわかる。動物が自ら変えられるのは、反応率 B である。

式 (12) は、単一のVRスケジュールのフィードバック関数だが、本当に調べたいのは並列VR–VRスケジュールである。並列VR–VRスケジュールの性質を調べるにあたり、1つ仮定をおこう。それは、反応数の総和は一定であるというものである。ハトのキーつつきの場合、左右のキーどちらも同じ重さであると考えれば、それほど不自然ではないだろう[16]。このことを全体の行動数を B として

$$B = B_1 + B_2$$

とする。全体の行動数 B が一定なのだから、動物が直面しているのは、行動 B_1 と B_2 それぞれにどの程度反応を配分するのかという問題である。そこで行動の配分率を

$$p = \frac{B_1}{B_1 + B_2} \cdots (13)$$

とおいておこう。すなわち、動物はマッチング実験において、 p を色々動かして、強化子の最大化を目指しているというふうに捉えていくということだ。強化子の最大化を目指すとは、2つの選択肢それぞれから得られる強化率を R_1、R_2 としたとき、$R_1 + R_2$ を最大化するということである。ここで、VRスケジュールのフィードバック関数 (12) を思い出せば、

$$R_1 = c_1 B_1$$

[16] とはいえ、左右のキーを切り替えるには、動物は移動しなければならない。よって必ずしもこの仮定が成り立つわけではない。さらに、本稿では説明しなかったが、実際のマッチング実験ではCOD (change over delay) という強化遅延を差し込む。ゆえに以降の議論は、あくまで近似である点には注意してほしい。それでもなお、現実の動物の行動を予測できるというのが重要な点である。

$$R_2 = c_2 B_2$$

であるため、

$$R_1 + R_2 = c_1 B_1 + c_2 B_2$$

である。さらに、式 (13) を使えば

$$\begin{aligned} R_1 + R_2 &= c_1 pB + c_2(1-p)B \\ &= B(c_2 + (c_1 - c_2)p) \cdots (14) \end{aligned}$$

である。

式 (14) を吟味してみよう。B は反応の総数 (あるいは、時間当たりの反応数の総和) であり、ここでは定数である。c_1、c_2 は各選択肢のフィードバック関数の傾きで、強化率の逆数であった。つまり、式 (14) は切片がBc_2、傾きが $B(c_1 - c_2)$ の一次関数であるということがわかる。それをグラフにしたものが図8である。$B = 100$ で固定しているが、これは適当な時間幅での反応の総数であると見ればよい。例えば、1分あたりの反応率が100回といった具合である。この元で、2つの選択肢への反応の配分率 p を横軸に取ってある。グラフを見たら、強化率が選択肢間で同じとき ($c_1 = c_2$；ピンク線)、傾きが0になっている。つまり、反応の配分がどうであれ、平均的な強化率は変わらないということである。強化率が選択肢間で異なるとき ($c_1 \neq c_2$；黄線・赤線)、傾きのある一次関数になって

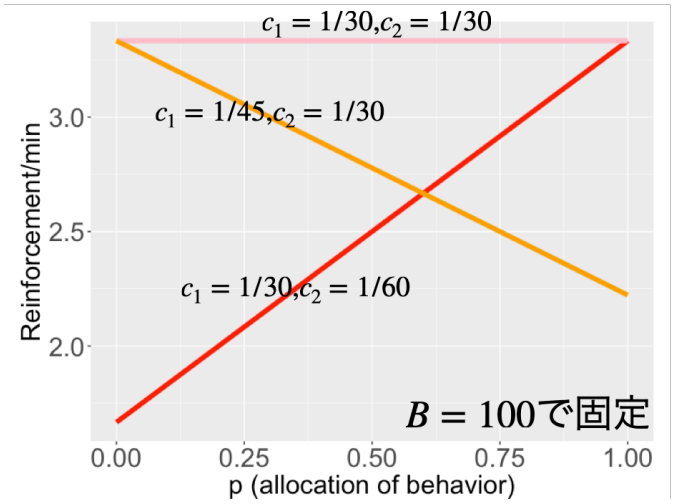

図8. 式 (14) を用いた並列VR–VRの最適化.

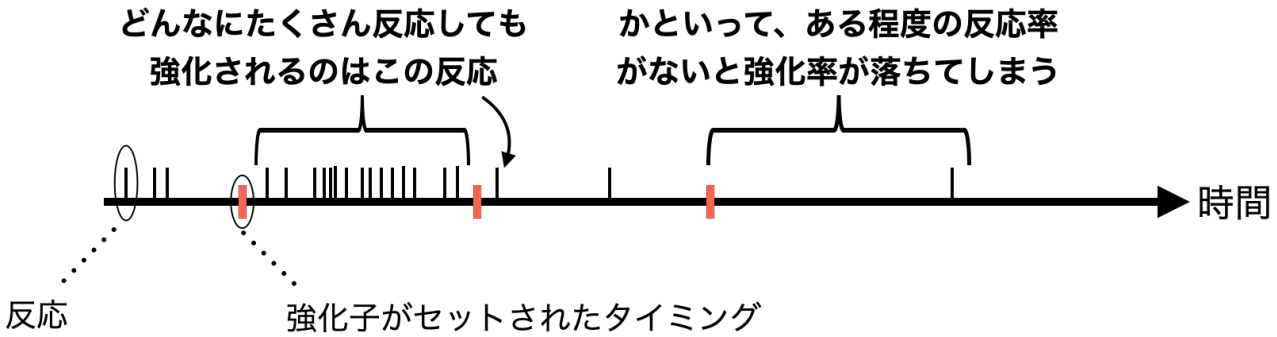

図9. VIスケジュールのパフォーマンス

いる。p は確率であるため、0から1までの値しか取れない。よって、強化率の最大値は必然的に0か1となる。すなわち、排他的な反応が最適点となるというわけだ。これは、先ほど見たHerrnstein and Loveland (1975) と整合している。並列VR–VRスケジュールでの最適な反応は、強化率の高い選択肢を排他的に選択することなのである。

3.3 最大化としての対応法則：並列VI–VIの場合

前節では、VRスケジュールのフィードバック関数を元に、並列VR–VRスケジュールにおける最適化を通じて、反応が排他的に片方にシフトすることを説明できた。この結果は、VRスケジュールのフィードバック関数が線形関数 (12) であることに起因する。VIスケジュールの場合、どうなるだろうか？VIスケジュールはVRスケジュールほど単純な形にはならない。VIスケジュールでは、反応率が低すぎたら強化率が下がってしまうが、反応率がある程度以上に高くなると、ほとんど強化率が上がらなくなる。

このことを図9を元に説明しよう。この図は黒線が反応で、赤線が強化子がセットされたタイミングを示している。VIスケジュールの場合、ランダムな時間経過後、反応すれば強化が得られるので、赤線の直後の反応が強化されるというふうに捉えればよい。縦線が密になっている箇所は、反応率がとても高い期間を表している。しかし、反応率がいかに高かろうと強化子がセットされるまでは決して強化されないので、ほとんどの行動は無駄になる。一方、縦線が疎な期間を見てほしい。その結果として、強化子がセットされてから反応までの間隔が空いてしまい、強化を得るまでに遅延期間が生じてしまう。つまり、反応率が極端に低いと強化率が下がってしまうのである。以上が、VIスケジュールの特徴となっている。

以上を考慮すると、VIスケジュールのフィードバック関数はある程度までは反応率に依存して強化率が上昇するが、それ以上は反応率を増やしても強化率がほとんど変化しなくなるような関数である。Baum (1981) は、そのような関数として、

$$R = \frac{1}{\tau + E}$$

を用いた。τ はスケジュールの強化率で、VI値の逆数を入れる。E は反応が τ を超えてしまった平均遅延時間を表している。E をどのように定式化するのが問題であるわけであるが、これは反応率が低いと大きくなり、高いと小さくなるような関数になればよいので、Baum (1981) は、

$$E = \frac{1}{B}$$

を用いている[17]。結果として、VIスケジュールのフィードバック関数は

$$R = \frac{1}{\tau + \frac{1}{B}} \cdots (15)$$

となる。この関数を描いたものが図10である。例えば赤線はVI30スケジュールなので、平均的に30秒に1回強化されるのが最大値であるはずである。つまり、1分あたりでは2回なので1時間あたりでは120回である。実際、赤線は反応率が上昇するに従って120に近づいていってるのがわかる。しかし、反応率が25回/分を下回ったあたりから急速に下がっている。これは、図9で説明したように、反応率が低いと、強化子がセットされてから反応までの間隔が空いてしまい、強化に遅延が生じるためである。

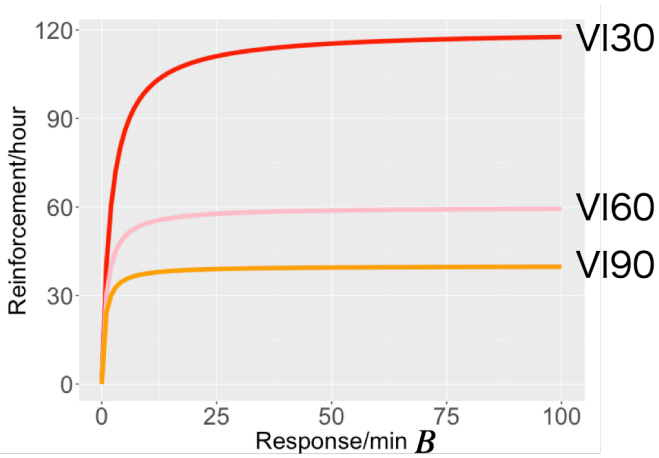

図10. VIスケジュールのフィードバック関数

[17] この関数は反応が時間的に一様に分布していることを仮定している。ただし、実際の動物の反応はそうはならない (Shull et al., 2001)。

単一VIスケジュールのフィードバック関数が式 (15) で決まったので、並列VI–VIスケジュールについて議論を進めてみよう。並列VR–VRスケジュールと同様に、2つの共立の総和を取れば、

$$R_1 + R_2 = \frac{1}{\tau_1 + \frac{1}{B_1}} + \frac{1}{\tau_2 + \frac{1}{B_2}}$$

となる。式 (13) の p の表記を用いて書き直せば

$$R_1 + R_2 = \frac{1}{\tau_1 + \frac{1}{pB}} + \frac{1}{\tau_2 + \frac{1}{(1-p)B}} \cdots (16)$$

であり、これが最大化すべき関数である。並列VR–VRと比べて、見た目が少しおどろおどろしいので、まずはグラフを描いてみよう。図11は並列VI20–VI60スケジュールを式 (16) に代入して描いたグラフである (赤線)。ピンク線、黄線は各選択肢のフィードバック関数に対応している。上に凸の関数になっているため、最適点を p が0から1の範囲のどこかに求めることができるようだ。そこで、式 (16) をp で微分し、解を0とおいて最大値を調べてみよう。

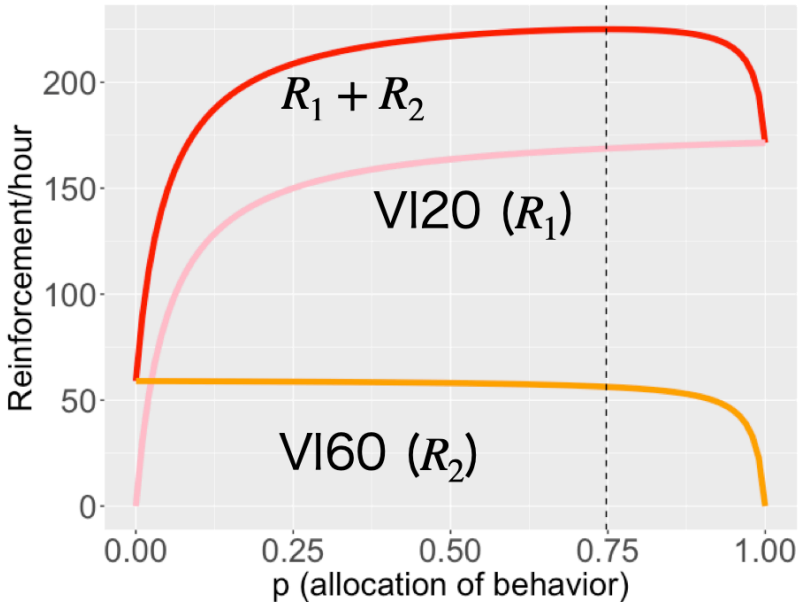

図11. VIスケジュールの最適化。点線が最適な反応配分を示す。

$$\frac{\partial(R_1+R_2)}{\partial p}=\frac{R_1^2}{p^2B}-\frac{R_2^2}{(1-p)^2B}=0$$

この等式を変形していくと、

$$\frac{R_1^2}{p^2B}=\frac{R_2^2}{(1-p)^2B}$$

$$\frac{p}{1-p}=\frac{R_1}{R_2}$$

$$\frac{B_1}{B_2}=\frac{R_1}{R_2}$$

$$\frac{B_1}{B_1+B_2}=\frac{R_1}{R_1+R_2}$$

となる。最後の式は、式 (10) の対応法則と全く同じ形をしている。VIスケジュールのフィードバック関数 (15) を繋いで獲得強化子の最大化を行なったら、対応法則に帰着するのである。Herrnstein (1961) の対応法則 (10) は、純粋に経験法則である。これ自体は、「行動がこうあるべきだ」ということを述べているわけでないということだ。しかし、フィードバック関数を式 (15) に設定し、最適な反応配分を求めた結果、対応法則に至る。しかも、実際に対応法則が並列スケジュールの反応において生じるというのが、なんとも面白い点だ。

3.4 最大化としての対応法則：並列VR–VIの場合

最後に、並列VR–VIスケジュールの反応配分について計算してみよう。議論の流れは前節までと同様である。VRスケジュールとVIスケジュールのフィードバック関数はそれぞれ式 (12) と (15) である。R_1 をVRスケジュールから得る強化とし、R_2 をVIスケジュールから得る強化としたら、獲得強化子の総和は

$$R_1+R_2=cpB+\frac{1}{\tau+\frac{1}{(1-p)B}}\quad\cdots(17)$$

となる。並列VR30–VI40で計算した式 (17) のグラフが図12である。図12の赤線を見ると、VRスケジュールに偏った点が最適点になっているのがわかる。実際、式 (17) を p で微分すると

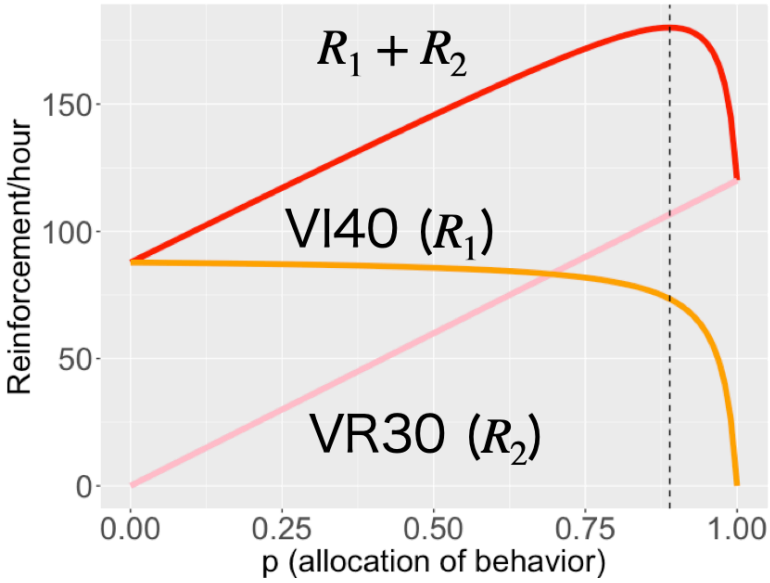

図12. 並列VR–VIスケジュールの最適化。点線が最適な反応配分を示す。

$$\frac{\partial(R_1+R_2)}{\partial p} = cB - \frac{R_2^2}{(1-p)^2 B} = 0$$

$$\frac{p^2}{(1-p)^2} = \frac{R_1^2}{cR_2^2}$$

$p = \frac{B_1}{B_1+B_2}$ であるため、最終的には

$$\frac{B_1}{B_2} = \sqrt{\frac{1}{c}}\frac{R_1}{R_2} \cdots (18)$$

となる。これは、$\sqrt{\frac{1}{c}}$ だけバイアスのかかったマッチングが起きていることに相当する。

c はVRスケジュールの強化率の逆数になっているため、常に $c > 0$ である。よって、式 (18) はVRスケジュールに偏った反応の配分を予測する式になっている。実際のデータでは、完璧にこの予測通りにはならないのだが、VRの方に反応が偏ったマッチングが実際に成立することが報告されている (Baum, 1974)。

4. 能動的推論から捉え直す行動の法則

前の2節では、古典的条件づけと道具的条件づけの理論的展開を扱ってきた。それぞれの節では、連合学習理論と対応法則について説明した。2節の連合学習理論では一連の予測誤差理論を紹介した[18]。紹介した研究は、試行ごとの連合の更新則をモデル化している。動物が刺激に直面して、学習を行う際のアルゴリズムを記述しているといってもよいだろう。一方、3節の対応法則における強化子の最適化は、試行ごとの学習則ではない[19]。第3節では何をしたかというと、反応率–強化率の相関性をフィードバック関数で特徴付けて、そこから強化の最大化を導いたのであった。巨視的最大化とは、行動が平均的には強化子最大化の方へと進むという方向性の予測である。

すると、さらなる問題設定は次のようになるだろう。行動と学習を最適化として定式化しながら、そのプロセスを試行ごと、あるいは適度に短い瞬間瞬間の学習則をアルゴリズムとして記述できないだろうか？そのような試みは、学習研究内でも存在する[20]が、ここでは自由エネルギー原理 (free energy principle) とそこから導かれる能動的推論 (active inference) からの定式化を紹介しよう (Friston, 2010; Friston et al., 2010)。そのためには、数学的準備が必要になる。詳細に立ち入ることは筆者の力量を上回ってしまうが、必要な概念について簡単に解説していくことにしよう。

4.1 数学的準備：情報量、エントロピー、カルバック・ライブラー情報量[21]

情報とはそもそもなんだろうか？「情報」という概念は、Shannon (1948) 以前は曖昧な言葉であった。Shannonの論文で初めて明瞭に定義されたらしい。実際Shannonによると

> “A basic idea in information theory is that information can be treated very much like a physical quantity, such as mass or energy.”
> – Caude Shannon, 1985.

---

[18] ただし、予測誤差修正を学習のプロセスとする試みはRescorla-Wagnerモデル以前からあった。また、この種のモデル化以外の方向性も連合学習理論には存在する。詳しくは今田・中島 (2003) を参照するとよい。

[19] 実験がフリーオペラント事態であるため、そもそも「試行」は定義されないが。

[20] 例えば、Hinson and Staddon (1983)

[21] 本稿より詳しいノートとして、「エントロピーを考えるドキュメント」を筆者は昔作った。必要に応じて、参照してほしい。
https://github.com/HeathRossie/memo/blob/main/%E3%82%A8%E3%83%B3%E3%83%88%E3%83%AD%E3%83%92%E3%82%9A%E3%83%BC%E3%82%92%E9%A0%91%E5%BC%B5%E3%82%8B%E3%83%88%E3%82%99%E3%82%AD%E3%83%A5%E3%83%A1%E3%83%B3%E3%83%88.pdf

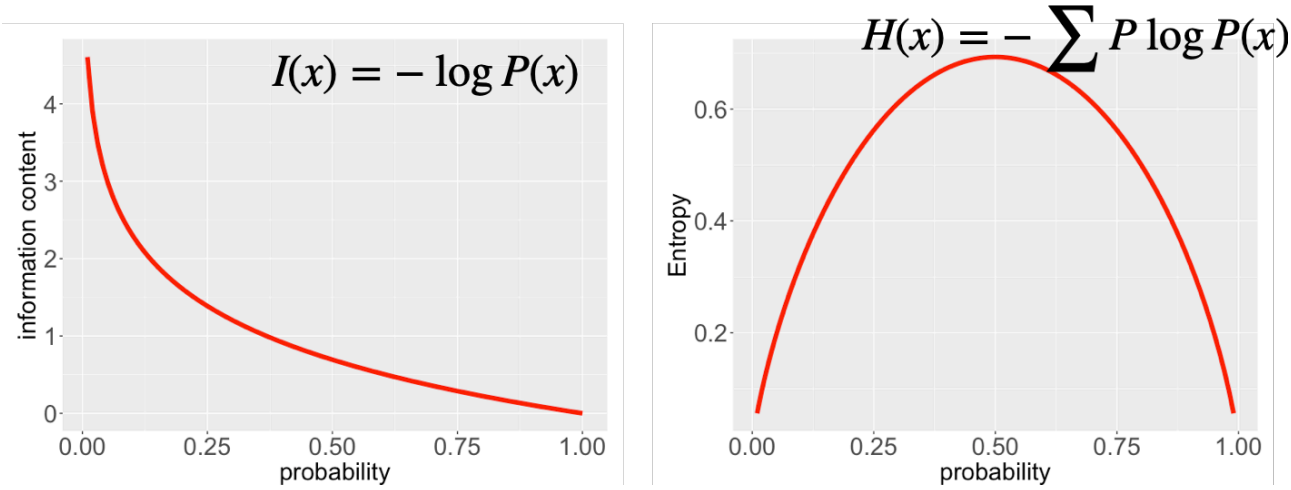

図13. 歪みがあったりなかったりするコインを投げたときの (a) 情報量 (b) エントロピー

とのこと。 前振りはこの程度にして、情報量の定義を行おう。$P(x)$ を確率 (分布) として、情報量は次のように定義される。

$$I(x) = -\log P(x) \cdots (19)$$

例えば、歪みのないコインであれば、投げたとき表が出る確率が0.5である。つまり、表が出たときの情報量は

$$-\log 0.5 = 0.69$$

である[22]。さて、次は歪みがあるコインを考えよう。「必ず表が出る」コインから「必ず裏が出るコイン」まで様々に確率を振って情報量 (19) を計算してみると、図13aのようになる。図を見れば確率が低い事象ほど、情報量が大きいことがわかる。つまり、起きづらいことが起きたときほど、情報量が高いというわけだ。そういうわけもあって、情報量は「サプライズ」を表現しているとも言える。

表が出る確率があるのだから、裏が出る確率もあるはずである。表が出る確率がすごく低いとき、表が出る情報量はとても大きくなる。一方で、そのとき裏が出る情報量は小さくなってしまうだろう (図13aを見て考えるか、あるいは自分で適当な値を代入して計算してみるとよい)。そう考え始めると、そのコインを投げて得られる平均的な情報量が知りたくなる。そこで、情報量の平均をエントロピーという名前で定義しよう。

[22] 情報理論の議論では、対数の底は通常2が使われる。そのとき単位は「ビット」(bit) と呼ばれる。ここでは自然対数を使っているが、議論の一般性には支障はない。

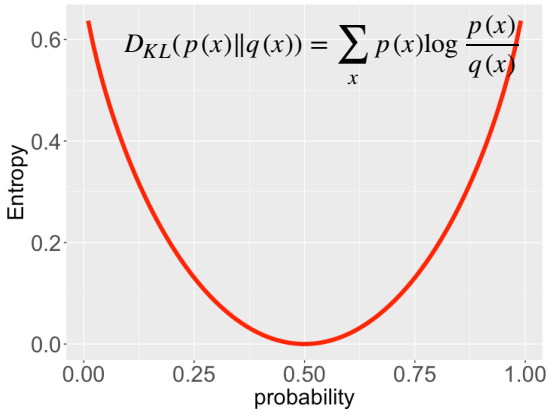

図14. 片方が歪みがなくて、もう片方はいろんな歪み方をしたコインのカルバック・ライブラー距離

$$H(x) = -\sum_x P \log P(x) \cdots (20)$$

コインの例のエントロピーを計算したのが図13bである。確率が0.5のときにエントロピーが最大になっているのがわかるだろう。すなわち、コインを投げたとき、どちらの面が出るかが一番不確定のときにエントロピーが最も高くなっている。そういうわけで、エントロピーは不確実性の尺度として使えそうなのがわかるだろう。ちなみに、連続力の場合は総和記号が積分に置き換わる。

最後の準備として、カルバック・ライブラー情報量 (Kullback-Leibler divergence) を紹介する。カルバック・ライブラー情報量は、相対エントロピー (relative entropy) とも呼ばれ、確率分布間の類似性を表す量である。$Q(x)$ と $P(x)$ という2つの確率分布があったとしよう。この2つの分布のカルバック・ライブラー情報量は次のように定義される。

$$D_{KL}(Q(x)\|P(x)) = \sum_x Q(x)\log\frac{Q(x)}{P(x)} \cdots (21)$$

見方としては、$\log\frac{Q(x)}{P(x)} = \log Q(x) - \log P(x)$ であり、両分布が完全に一致するとき0になる。この「ずれ」の部分を $Q(x)$ で期待値をとっているのがカルバック・ライブラー情報量である。すなわち両分布が一致するとき、$D_{KL}(Q(x)\|P(x)) = 0$ である。それ以外のとき、カルバック・ライブラー情報量は0以上の値をとる。いわば分布間の距離のような指

標となってる[23]。2つのコインを用意して、片方が完全無比の歪みのなさ、もう片方はいろんな歪ませ方をして表が出る確率が様々なときのカルバック・ライブラー情報量を計算したのが図14である。2つの確率分布が一致しているとき0になっていて、他のときは0以上の値をとっているのがわかるだろう。

4.2 自由エネルギー原理、超入門[24]

自由エネルギー原理とは、神経科学者のKarl Fristonが脳の機能について「統一原理」[25]として提唱しているものであり、“いかなる自己組織化されたシステムでも 環境内で平衡状態であるためには、そのシステムの (情報的) 自由エネルギーを最小化しをくてはならない” という規範的な命題であり、“適応的なシステムが無秩序へ向かう自然的な傾向に抗して持続的に存在しつづけるために必要な条件” である (Friston, 2010)。これだけではおそらく何を言っているかさっぱりであると思われるし、学習となんの関係があるのかも意味不明である。もっと「統一原理」じみた理論的主張についてはFristonの元論文などにあたってもらうとして、ここでは学習現象の取り扱いに有用な枠組みとして紹介しよう。

まず、前提を準備しよう。動物は感覚入力 o を受け取っているとしよう。感覚入力 o は環境の刺激原因 x によって生じている。例えば、目の前にりんご (x) があったとして、網膜には赤色で円形の刺激作用 (o) が生じる。我々が受け取ってるのは後者である。我々も含め動物は、原因 x を直接に知覚できるわけではない。感覚入力 o から x を推論しているのである。推論するといっても、意識的に行なっているということを含意しているわけではない。与えられたデータに相当する感覚入力 o から、それを引き起こした原因 x を特定するようなプロセスが実行されているということである。これは、Helmholtzの無意識的推論と呼ばれる知覚観でもある。

---

[23] 「ような」といったのは、期待値を $Q(x)$ でとったときと $P(x)$ でとったときで値が変わるので、正確には距離とは言えないためである。あなたから見たときと、筆者から見たときでお互いの間の距離が違うと、それは距離の定義としては変であろう。そのためカルバック・ライブラー情報量は「カルバック・ライブラー擬距離」と呼ばれることもある。

[24] 筆者が作ったもう少し詳しい資料をこちらに残してある。
https://github.com/HeathRossie/memo/blob/main/%E4%BB%8A%E5%BA%A6%E3%81%93%E3%81%9D%E3%82%8F%E3%81%8B%E3%82%8B%E8%83%BD%E5%8B%95%E7%9A%84%E6%8E%A8%E8%AB%96.pdf

[25] なぜ鉤括弧がついているかというと、統一原理であるということはまだまだ受け入れられているわけでもないし、反論もあるためだ。とはいえ、かなり幅広い現象を扱える枠組みであることには疑いようがない。

そのような推論を行うことを定式化していこう。動物は x と o の同時分布 $P(x,o)$ を内的モデルとして有しているとしよう。これを生成モデル (generative model) と呼ぶ。ようは $P(x,o)$ に対応するような脳活動があるということを仮定しているわけだが、本項では神経基盤については触れない。さらに、動物は感覚原因 x の事前分布 $P(x)$ を信念として内的に表現していることも仮定しよう。先程のりんごの例でいれば、赤色で円形の感覚入力 o を受ける前に、そもそも環境にりんごがどれくらいの確率で存在するのかについて信念を持っているということである。ここまでは、そこまで不自然な仮定でもないだろう。$P(x,o)$ と $P(x)$ を動物が有していることを認めれば、たちどころに次が成り立つ。

$$P(o\,|\,x) = \frac{P(x,o)}{P(x)}$$

上式は単に、条件付き確率の定義から成立する。この条件付き確率には尤度 (likelihood) という名前が与えられている。この場合の尤度は、りんごが目の前にあったとしたとき、赤色で円形の刺激を受け取ってる状態になるのはどのくらいの確率で起きうるのかということを指している。

しかし、動物が本当に知りたい確率は $P(o\,|\,x)$ ではない。動物は感覚入力 o からその原因 x について推論しないといけないのだから、評価したい確率は $P(x\,|\,o)$ である。ここで、ベイズの定理を用いれば、

$$\begin{aligned} P(x\,|\,o) &= \frac{P(o\,|\,x)P(x)}{P(o)} \\ &= \frac{P(o\,|\,x)P(x)}{\sum_x P(o\,|\,x)P(x)} \quad \cdots (22) \end{aligned}$$

である。しかし、ベイズ統計学の世界では、この確率は通常求めるのが困難であることが知られている。その理由は、分母の $P(o)$ の計算が難しいからである。詳しくは入門的なベイズ統計学の教科書を参照してほしい[26]が、動物の推論に引きつけて直観的に説明しておこう。$P(o) = \sum_x P(o\,|\,x)P(x)$ であるため、可能な全ての x について足し合わせなければいけない[27]。赤くて丸い感覚刺激 o の原因を推論するために、可能な全ての原因 (りん

[26] 例えば久保 (2012) 『データ解析のための統計モデリング入門』

[27] 考えている確率分布が連続分布の場合、ここは積分になる。多変量の場合、多重積分になる。

ご、マンゴー、血の染み、スイカの断面、etc...) について $P(o|x)P(x)$ を評価しているとは到底考えづらいだろう。ということは、式 (22) はそのままで動物の知覚のモデルとして利用しづらい。なんかの近似推論を利用していると考える方が、モデルとしてもっともらしそうだ。そこで自由エネルギー原理で用いるのが、変分推論 (variational inference) という考え方である (Galdo et al., 2020)[28]。

変分推論では、$P(x|o)$ を直接計算するのを諦める。代わりに、近似分布 $Q(x)$ を用意する。その近似分布 $Q(x)$ を $P(x|o)$ に近づけることを目標とする。近似分布というのだから、$P(x|o)$ より単純な形をしていそうな分布を使わないと意味がない。例えば正規分布 (Buckely et al., 2017) や一点分布 (Bogacz, 2017; 2020) が挙げられる。$Q(x)$ を $P(x|o)$ に近づけるというのだから、分布間の類似性を測らないといけない。そのような量としては、前節で紹介したカルバック・ライブラー情報量 (21) が使えるだろう。また、近似分布 $Q(x)$ は、動物の持つ信念であるという意図を込めて、Fristonの用語法では認識分布 (recognition density) と呼ばれる。

$$D_{KL}(Q(x)\|P(x|o)) = \sum_x Q(x)\log\frac{Q(x)}{P(x|o)}$$

今、動物は現在の信念 $Q(x)$ を色々と変えることで、$P(x|o)$ に近づけ、最適な認識を実現したいわけだ。しかし、このカルバック・ライブラー情報量は、そのままでは評価できない。$P(x|o)$ が計算困難であるからだ。$P(x|o)$ がわからないので、カルバック・ライブラー情報量もわからない。なので、このままでは馬鹿みたいなのだが、辛抱強く変形していってみるとしよう。条件付き確率の定義から $P(x|o) = \frac{P(x,o)}{P(o)}$ であるので、それを代入してみよう。

$$\begin{aligned} D_{KL}(Q(x)\|P(x|o)) &= \sum_x Q(x)\log\frac{Q(x)P(o)}{P(x,o)} \\ &= \sum_x Q(x)\log\frac{Q(x)}{P(x,o)} + \sum_x Q(x)\log P(o) \end{aligned}$$

---

[28] 不完全な解説だが、筆者も一応補足的な資料を公開している
https://github.com/HeathRossie/memo/blob/main/%E5%A4%89%E5%88%86%E6%8E%A8%E8%AB%96%E3%82%92%E9%A0%91%E5%BC%B5%E3%82%8B%20(%E6%9C%AA%E5%AE%8C%E6%88%90).pdf

$$= D_{KL}(Q(x)\|P(x,o)) + \log P(o)$$

2段から3段目の変形では $P(o)$ が x に依存しないこと、$\sum_x Q(x) = 1$ であることを利用している。認識分布と事後分布のカルバック・ライブラー情報量を2つの項に分解できた。ここで自由エネルギー F を

$$F = D_{KL}(Q(x)\|P(x,o)) \cdots (23)$$

と定義しておけば、

$$D_{KL}(Q(x)\|P(x\,|\,o)) = F + \log P(o) \cdots (24)$$
$$F = D_{KL}(Q(x)\|P(x\,|\,o)) - \log P(o) \cdots (25)$$

となる。まだ単に式 (23) で自由エネルギー F を名付けただけだが、(23-25)を眺めていると、F は認識分布 $Q(x)$ と生成モデル $P(x,o)$ のカルバック・ライブラー情報量である。動物は、現在の外界への信念として $Q(x)$ を持っているのに加え、これまでの経験や系統発生により $P(x,o)$ を有している (と、私たちは仮定したのだった)。従って、式 (23) は評価できる[29]。次に、式 (24) の第2項の $\log P(o)$ は感覚入力 o の情報量、つまりサプライズになっていることがわかる。

動物が、認識の中で変更できるのは $Q(x)$ である。例えば、目の前にあるのがりんごだと思っていたが、実は梨かもしれない。自由エネルギー F の定義である式 (23) を見ると、$Q(x)$ を生成モデル $P(x,o)$ に近づけるように変化させれば、F を減少させられることがわかる。一方で、$-\log P(o)$ は o に依存する量である。つまり、$Q(x)$ を変更しても変化することはない。このことを踏まえて式 (24) を見ると、F を減らすように $Q(x)$ を変化させれば、カルバック・ライブラー情報量 $D_{KL}(Q(x)\|P(x\,|\,o))$ を下げることができる。結局、F を下げればよいということに気づくことができた。また、$D_{KL}(Q(x)\|P(x\,|\,o)) \geq 0$ であるため、式 (24) を使えば

[29] 1つ誤魔化しがある。生成モデルはあくまで動物が有している内部モデルであるので、外界と感覚入力の真の関係とは限らない。この真の関係を「生成プロセス」と自由エネルギー原理用語ではいう。ここでは、生成モデルは生成プロセスには十分近いということを仮定している。しかし、より興味深い話として、生成モデルが生成プロセスを必ずしも近似していないときの方が行動が最適になるというシミュレーション結果も報告されている (Tschantz et al., 2020)。

$$D_{KL}(Q(x)\|P(x\,|\,o)) = F + \log P(o) \geq 0$$

$$F \geq -\log P(o)$$

となる。$-\log P(o)$ は感覚のサプライズなので、F はサプライズの上限になっていることもわかった。$Q(x)$ を変化させて、$P(x,o)$ に近づけることが、知覚 (認識) なのだ。

「いやいや、学習とか行動とかの話が主題じゃなかったのか？」と思われたかもしれない。F を下げるように行動すれば、外界についての不確実さを下げることができるのだ、ということをこれから紹介する。今度は、$P(x,o) = P(o\,|\,x)P(x)$ を使って式 (23) を変形すると、

$$\begin{aligned} F &= D_{KL}(Q(x)\|P(x,o)) \\ &= \sum_x Q(x)\log\frac{Q(x)}{P(x,o)} \\ &= \sum_x Q(x)\log\frac{Q(x)}{P(o\,|\,x)P(x)} \\ &= \sum_x Q(x)\log\frac{1}{P(o\,|\,x)} + \sum_x Q(x)\log\frac{Q(x)}{P(x)} \end{aligned}$$

$$F = -\sum_x Q(x)\log P(o\,|\,x) + D_{KL}(Q(x)\|P(x)) \cdots (26)$$

最後の変形 (26) で、F がやはり2つの項に分解されていることがわかる。第1項にある $\sum_x Q(x)\log P(o\,|\,x)$ は、正確性 (accuracy) と呼ばれている項である。式 (26) はその符号反転になっていて、その場合は不確定性 (uncertainty) ともいう。なんでそんな名前が付けられているかを説明しよう。$\sum_x Q(x)\log P(o\,|\,x)$ は $\log P(o\,|\,x)$ を $Q(x)$ で期待値を取ったものである。$P(o\,|\,x)$ は尤度である。なので、推測された感覚原因 $Q(x)$ の下で、所与の感覚入力を受ける確率を評価していることになる。第2項の $D_{KL}(Q(x)\|P(x))$ には、ベイジアンサプライズ (Bayesian surprise) という名前がついている。この量は、現在の推測 $Q(x)$ と事前信念 $P(x)$ の間のカルバック・ライブラー距離である。つまり、現在の推測 $Q(x)$ が事前信念から離れるほど大きくなってしまう。事前信念 $P(x)$ は個体の経験や、系

統発生から決まってくるものだが、「世界はこうなっているだろう」という信念を表している。そこから大幅に外れるような事態は、そうそう起きるものではない。例えば、目の前にりんごがあるのは不可思議なことではないが、直径10cmの巨大なルビーである確率は (あなたがどこかの国の王様か大富豪でもない限り) 低いだろう。そのように捉えれば、ベイジアンサプライズ $D_{KL}(Q(x)\|P(x))$ は、極端でありえそうなもない認識を防ぐための罰則項になっていることがわかる。

ここで、現在の推測 $Q(x)$ を固定したときにどうなるかを考えてみよう。認識はとりあえず決まって、式 (26) を再度考察してみるということだ。ベイジアンサプライズ $D_{KL}(Q(x)\|P(x))$ は変わりようがない。$Q(x)$ が決まったら、ここは定数扱いになるだろう。次に、式 (26) の第1項である正確性 $\sum_x Q(x)\log P(o|x)$ である。この項は、感覚入力 o によっても変化する。実際には式 (26) ではマイナスの符号がついているので、正確性が上がれば F は下がることに注意しよう。言い換えれば、不確実性が下がるのだ。 どういうときに不確実性が下がるのか？ $\sum_x Q(x)\log P(o|x)$ を眺めていれば、$Q(x)$ に整合するような感覚入力 o が得られればよいことがわかる。感覚入力は、単に受動的に刺激から入力されるわけではない。私たち動物は動くことができる。なんらかの行動を通じて、感覚入力を変更することができるというわけだ。例えば、りんごらしき赤くて丸い物体が視界の端に映れば、そちらを見たり触ったりすることで、それがりんごである確度を高めることが可能である。これで行動によって、F を下げられることがわかった。このように、動物が行動を通じて自由エネルギー F を低減するという枠組みを能動的推論 (active infernce) という。

行動と知覚がともに F を下げることを通じて定式化できそうなことをこれまで見てきた。最後に「学習」である。世界と感覚入力の関係を動物が内的に保持しているのが、生成モデル $P(x,o)$ であると言った。モデルというのだから、パラメータが存在する。そのパラメータを経験を通じて変容させることができると考えよう。すると、式 (25-26) を通じて F の低減させるとき、取るべき知覚・行動が変わってくる。こうして、学習が成立するというわけだ。文献によっては、θ を生成モデルのパラメータとして、$P(x,o)$ を $P(x,o|\theta)$ のように書くことがある。ここでも F を下げるように θ を変えるような学習則を考えることができる。

以上、かなりミニマルな説明[30]ではあるが、動物がどのように認識し、行動し、学習するかを F の最小化の下、理解できることが示唆された。自由エネルギー原理とは、以上を満たすような形で、知覚、行動、学習といった動物の諸相に迫っていこうという主張である[31]。各ドメインへの適用例は、個別の論文に当たるか、乾・阪口 (2021) を参照するとよい。

4.3 Bogacz (2020) のDopActモデル

自由エネルギー原理は、あくまで原理を満たす形で動物の諸側面を捉えていこうという主張である。「原理」(principle) であって、特定の理論やモデルではない。従って、具体的な検証を行うには、自由エネルギー原理を満たした特定の理論・モデルを考えなければならない。Fristonやその関係者たち自身が精力的にそのようなモデルを提案していてるが、ここではBogacz (2020) のモデル (の一部) を紹介しよう。理由としては、彼のモデルは能動的推論にインスパイアされたものであるが、難解な期待値計算などは出てこないためだ。しかし、具体的な行動現象の説明としては筋が通っているように著者には感じられるので、本節ではそのエッセンスを紹介しよう。

Bogacz (2020) は、動物がその瞬間の報酬 r と期待報酬 v の総和 R を最大化すべく、行動を a を決定していると考える。つまり、最大化すべき報酬として

$$R = r + v \cdots (27)$$

v は価値の評価システム (valuation) で計算されると仮定する。Bogacz (2020) の評価システムには、強化学習モデルでよく使われるTD学習というモデルが使われているが、本稿の主題ではないのでここでは省かせてもらう。とにかく、報酬と期待報酬が決まったのだと考えてほしい。ここでいう行動 a は、反応強度として、時間当たりの反応数である反応率を考えている。ある刺激 (あるいは刺激強度) s の下で、期待される報酬が R であるときの反応強度の確率分布 $p(a|R,s)$ を事後分布として捉えると、この分布は次のように書ける。

$$P(a|R,s) = \frac{P(R|a,s)P(a|s)}{P(R|s)} \cdots (28)$$

---

30 例えば、能動的推論では期待自由エネルギーという量が出てきて、行動選択の定式化を行う (Tschantz et al., 2020)。

31 本稿で述べたことを超えて、Fristonらは自由エネルギー原理を生命の原理であるという主張にまで拡張している (Friston, 2013)。そのような最新版自由エネルギー原理は、筆者には正直なところよくわからないし、動物の行動を考える上では、必ずしもその主張まで受け入れる必要もない。

(28) のように反応強度の分布 $P(a|R,s)$ を分解できたわけだが、分子にはそれぞれ意味合いを付与することができる。$P(R|a,s)$ は、ある刺激のもとでなんらかの反応率で反応をした際にどの程度の報酬が得られるかを表現している。Bogacz (2020) はこれを「目的志向 (goal-directed)」の分布であると仮定する。一方で、$p(a|s)$ は報酬とは無関係に、ある状態でどの程度の反応強度で行動を遂行したかを表現している。そこで $P(a|s)$ を習慣 (habit) とみなしている。

加えて、目的志向 $P(R|a,s)$ は報酬量に関する (動物が有している) 確率分布であるが、習慣 $P(a|s)$ は反応強度の確率分布であることにも注意しよう。つまり、両者は異なる量を表現した確率分布である。また、(28) はベイズ則なので、習慣 $P(a|s)$ は反応強度の事後分布 $P(a|R,s)$ の事前分布になっている。ひとまず、a は反応率という解釈ができて、R は報酬量なので、ひとまずは理解できる。また、s はそのとき提示されている刺激である。しかし、まだ $P(\cdot)$ がどんな分布なのかも、s がなんなのかも決まっていない。Bogacz (2020) では、$P(\cdot)$ に正規分布を仮定して、モデルをシンプルに表現している。

$$P(x) = f(x;\mu,\Sigma) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left(-\frac{(x-\mu)^2}{2\Sigma} \right) \cdots (29)$$

ここで、μ は平均、Σ は分散を表す正規分布のパラメータになっている。目的志向 $P(R|a,s)$ も習慣 $P(a|s)$ も正規分布であるわけだが、平均の取り方は次のような単純なモデルを用いている

$$P(R|a,s) = f(R;aqs,\Sigma_g) \cdots (30)$$
$$P(a|s) = f(a;hs,\Sigma_h) \cdots (31)$$

式 (30-31) が言っていることを行動的に翻訳してみよう。(30) は目的志向の部分の確率分布で、報酬量に関する表象を表している (いわばR–O連合)。動物にとって、報酬量に関する予期は平均値の周りを Σ_g の周りで正規分布しており、その平均値は aqs で表現される。つまり反応強度 a と刺激強度 s の関数で、パラメータ q でスケール化されている。刺激強度 s は、連続量を想定するケースも考えられるが、単純には $s = \{0,1\}$ と刺激の存在の有無を示すインジケータにしてもよい。例えば、反応レバーが提示されてるか否かを示すには $s = \{0,1\}$ としておくのが得策だろう。次に、(31) 式は習慣の部分のモデルであり、(30) 式と同様に分散 Σ_h を持つ正規分布である。反応強度 a は刺激強度 s のみに依存し、その強度はパラメータ h に依存すると仮定している。ある刺激のもとでどれほどの反応強度で行動を遂行してきたかだけを表現していて、結果事象は考慮に入れていないので、いわばS–R連合を示している。

尤度分布も事前分布も両方正規分布なので、この事後分布 $p(a|R,s)$ には解析的な解がある。結果としては、反応強度の事後分布は、目的志向と習慣の分布の中間にくる。しかし、より分散が小さい方に寄った分布になる。しかし、実際の動物が (28) を直接評価しているとは考えにくいとい

うのが、前節で紹介した議論の流れなのであった。$P(R|s)$ は取りうる行動 a で $P(R|a,s)$ を積分しないと計算ができないためだ (Bogacz, 2017)。そこで、変分推論を用いた近似計算を仮定する。そのために、自由エネルギー F を次のように仮定する。

$$F = \log[P(R|s,a)P(a|s)] \cdots (32)$$

これは、(28) 式の分子部分の対数をとったものである。対数を取ることで、正規分布の指数部分が消えるので計算が容易になる。Fristonらの本家自由エネルギー原理に登場する変分自由エネルギーとの対応としては、$P(R|s,a)p(a|s) = P(R,a|s)$ であることを押さえれば、$P(R,a|s)$ を生成モデルとして捉えると、(32) 式は生成モデルの対数であることがわかる。Friston (2005) や Bogacz (2017) で用いられている自由エネルギーは、認識分布 を $Q(\cdot)$ と置いたときに生成モデルと $Q(\cdot)$ のKL情報量である。(6) 式は近似分布に一点分布を仮定し、符号を反転した場合に相当する。つまり、Fristonらの自由エネルギー原理では、認識の不確実さ (分散) もモデル化するが、Bogacz (2020) は点推定として認識をモデル化するという単純化を図っているということだ。行動のプランニングでは、$R = v$ と期待報酬をセットし、F を最大にするように a を探索する。実際に、$F = \log[P(R|a,s)P(a|s)]$ に正規分布 (29) を仮定した式 (30)、(31) 代入すると、

$$F = \frac{1}{2}\left(-\ln \Sigma_g - \frac{(R - aqs)^2}{\Sigma_g} - \ln \Sigma_h - \frac{(a - hs)^2}{\Sigma_h} \right) + C \cdots (33)$$

(C は定数項)

となる。定数部分は、微分したら消えるのでこの後の議論では出てこない。行動強度を決めるには、式 (33) を a で微分する。

$$\dot{a} = \frac{\partial F}{\partial a} = \frac{(R - aqs)qs}{\Sigma_g} - \frac{hs - a}{\Sigma_h} \cdots (33)$$

ここで報酬予測誤差を $\delta_g = \frac{R - aqs}{\Sigma_g}$、習慣誤差を $\delta_h = \frac{a - hs}{\Sigma_h}$ としておけば

$$\frac{\partial F}{\partial a} = \delta_g qs - \delta_h$$

と簡単に書ける。こう見れば、期待報酬に近づけるために反応変えていく第1項と、これまで推敲してきた行動強度に合わせようとする習慣部分としての第2項の綱引きによって、行動強度が決定されているのがわかる。

行動し、結果事象を得たとしよう。つまり、$R = r$ を得た。そこで、式 (30)、(31) のパラメータ q、Σ_g、h、Σ_h の更新を行う。これもまた、式 (33) のF の勾配に基づいて行う。導出の詳細は省くが、式 (33) を各パラメータで微分すると、

$$\Delta q \sim \frac{\partial F}{\partial q} = \delta_g a s$$

$$\Delta h \sim \frac{\partial F}{\partial h} = \delta_h s$$

$$\Delta \Sigma_g \sim \frac{\partial F}{\partial \Sigma_g} \sim (\delta_g \Sigma_g)^2 - \Sigma_g$$

$$\Delta \Sigma_h \sim \frac{\partial F}{\partial \Sigma_h} \sim \delta_h^2 - \Sigma_h$$

を得ることができる。以上の更新則によって、動物は学習を行っていると考える。DopActモデルが能動的推論から着想を得ているのはどこかというと、行動の遂行も学習も、いずれも自由エネルギーの最適化から捉えている部分である。見てみれば、行動強度の決定は式 (33) で、F を a で微分することによって得ている。その結果として強化を受けた際には、今度は F 各パラメータで微分することで学習を更新しているのだ。

Bogacz (2020) はDopActモデルを用いて、いくつかのシミュレーションを行なっているが、ここではDickinson (1995) のシミュレーションを紹介しよう。Dickinson (1995) は、ラットに道具的条件づけを施したが、片方の群のラットには120回の強化、もう片方には360回の強化を行なった。つまり、訓練の経験数が異なる。彼らはさらにラットを2群にわけ、片方は訓練時と同様に空腹の状態にし、もう片群は実験前に餌を与え、ほとんど満腹にしてしまった。これは強化子の低価値化 (devaluation) と呼ばれる手続きである。通常、動物は低価値化を施された強化子が随伴される反応をあまり行わなくなる。実際、全ての群のラットに消去テスト[32]を実施したところ、120回の訓練を受けた群では、空腹なラットはレバーを高頻度で押したが、低価値化を受けたラットはほとんど反応を行わなかった。DopActモデルで同じような随伴性をシミュレーションしても、同様の結果になる (図15左[33])。

興味深いのは、360回の訓練を受けた群である。こちらの群では、低価値化を受けたラットも比較的高い反応率を示した。強化子の価値がもはやないにも関わらず、レバーを多く押したということである。このように、結果事象への感受性が低減した反応傾向を習慣行動 (habitual behavior) が呼ばれる。DopActモデルのシミュレーションでも、繰り返しの訓練により低価値化したレバーへの反応が再現された(図15右)。この背景には、Bogacz (2020) のモデルが習慣システム $P(a|s)$ を仮定し、その習慣の強度パラメータ h が徐々に強固になっていくことが原因にある。動物の行動も、繰り返し同じ反応を遂行することで徐々に目的指向から習慣へと制御が変化していくのだが、その移行過程をDopActモデルは $P(R|a,s)$ と $P(a|s)$ のパラメータの更新で捉えているのである。

---

[32] 餌の提示を行わず、レバーだけを提示して反応の生起頻度を記録するテスト

[33] この図は、Bogacz (2020) のMatlabコードを筆者がJuliaで書き直して再現したものである。https://github.com/HeathRossie/Bogacz2020_active_inference/blob/main/Bogacz_2020_Fig8A_devaluation.jl

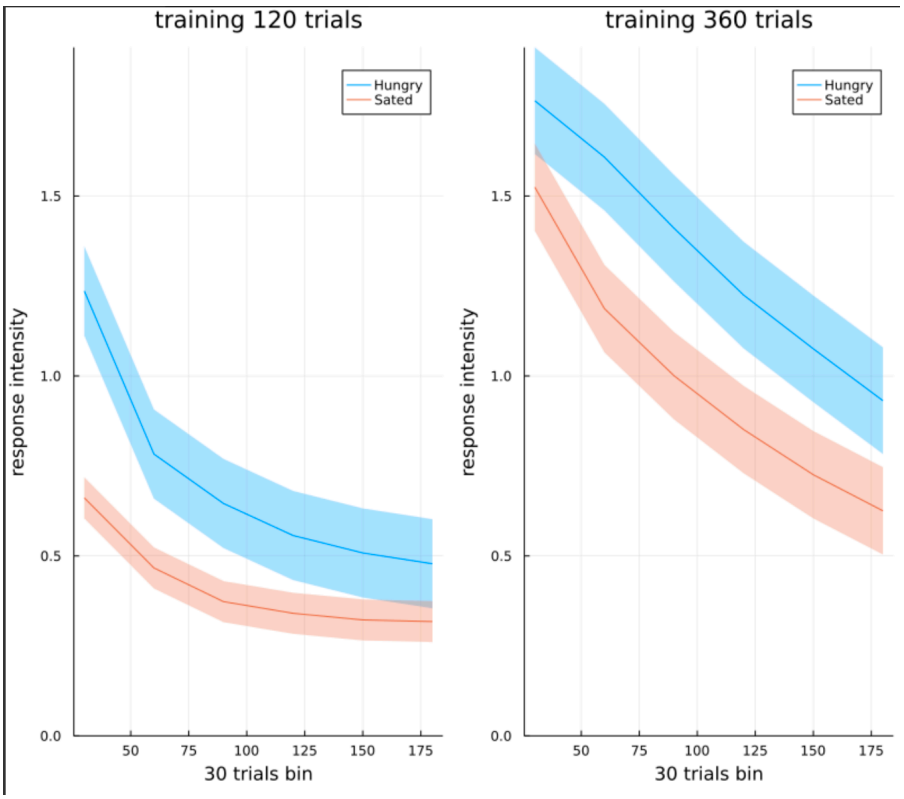

図15. Dicinson (1995) の習慣行動実験のシミュレーション

5. 参考文献

Baum, W. M. (1981). Optimization and the matching law as accounts of instrumental behavior. Journal of the experimental analysis of behavior, 36(3), 387-403.

Baum, W. M. (2021). Introduction to molar behaviorism and multiscale behavior analysis. In Contemporary Behaviorisms in Debate (pp. 43-62). Springer, Cham.

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation 1. Journal of the experimental analysis of behavior, 12(6), 861-874.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. The Psychological Record, 50(2), 251-265.

ボークス (1990). 動物心理学史―ダーウィンから行動主義まで. 宇津木保・宇津木成介 (訳), 誠信書房.

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. Journal of mathematical psychology, 76, 198-211.

Bogacz, R. (2020). Dopamine role in learning and action inference. Elife, 9, e53262.

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. Journal of Mathematical Psychology, 81, 55-79.

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement. New York, NY: Appleton-Century-Crofts.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

Friston, K. (2013). Life as we know it. Journal of the Royal Society Interface, 10(86), 20130475.

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. Biological cybernetics, 102(3), 227-260.

Galdo, M., Bahg, G., & Turner, B. M. (2020). Variational Bayesian methods for cognitive science. *Psychological methods*, *25*(5), 535.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, *122*(3160), 157-158.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS–US pairings. Journal of Experimental Psychology: Animal Behavior Processes, 5(1), 31.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. Journal of the experimental analysis of behavior, 4(3), 267.

Herrnstein, R. J. (1970). On the law of effect 1. Journal of the experimental analysis of behavior, 13(2), 243-266.

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules 1. Journal of the experimental analysis of behavior, 24(1), 107-116.

Hinson, J. M., & Staddon, J. E. R. (1983). Hill-climbing by pigeons. Journal of the Experimental Analysis of Behavior, 39(1), 25-47.

乾敏郎・阪口豊 (2021). 自由エネルギー原理入門: 知覚・行動・コミュニケーションの計算理論. 岩波書店

Killeen, P. (1972). The matching law. Journal of the experimental analysis of behavior, 17(3), 489-495.

藤巻峻, 新保彰大, 松井大, 時暁聴, & 神前裕. (2015). 条件づけにおける時間 II—オペラント計時行動, および時間学習の神経機構—. 基礎心理学研究, *34*(1), 78-90.

今田 寛 監修・中島定彦 編 (2003). 学習心理学における古典的条件づけの理論-パヴロフから連合学習研究の最先端まで-, 培風館.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. in ba campbell & rm church (eds.), Punishment and aversive behavior (pp. 279-296). New York: Appleton-Century-Crofts.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: the effect of nonreinforced pre-exposure to the conditional stimulus. Journal of comparative and physiological psychology, 52(4), 415.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological review, 82(4), 276.

ジェームズ・E. メイザー (2008). メイザーの学習と行動. 磯博行・坂上 貴之・川合 伸幸 (訳). 二瓶社.

Rescorla, R. A., & Wagner, A. R. (1972). In AH Black, & WF Prokasy (Eds.), A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement Classical conditioning II: Current research and theory (pp. 64–99). New York: Appleton Century Crofts.

坂上貴之・井上雅彦 (2018). 行動分析学. 有斐閣アルマ.

澤幸祐 (2021). 私たちは学習している行動と環境の統一的理解に向けて. ちとせプレス.

Shannon, C. E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological review, 87(6), 532.

Rachlin, H., & Baum, W. M. (1969). Response rate as a function of amount of reinforcement for a signalled concurrent response 1. Journal of the Experimental Analysis of Behavior, 12(1), 11-16.

Shull, R. L., Gaynor, S. T., & Grimes, J. A. (2001). Response rate viewed as engagement bouts: Effects of relative reinforcement and schedule type. Journal of the experimental analysis of behavior, 75(3), 247-274.

Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. The Journal of General Psychology, 12(1), 40-65.

Skinner, B. F. (1956). A case history in scientific method. American psychologist, 11(5), 221.

Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. PLoS computational biology, 16(4), e1007805.