

認知的柔軟性の基盤としての潜在構造の学習： 行動指標と計算モデリングによる分析

Learning a latent structure as a basis of cognitive flexibility: Analysis with behavioral indicator and computational modeling

真隅 暁, 佐藤 尚
Akira Masumi, Takashi Sato

沖縄工業高等専門学校
National Institute of Technology, Okinawa College
masumi.a@okinawa-ct.ac.jp

1. はじめに

我々は日常生活において、しばしば、とるべき行動の柔軟な切り替えを必要とするような状況に遭遇する。身近な例として自動車の運転が挙げられる。自動車を意図した通りに操作するためには、直接的には、ハンドルやアクセルの適切な操作が必要である。しかし、例えば「どれくらいの速度が適切であるか」は混雑具合や歩行者の存在といった周囲の状況に依存する。そしてこの状況自体が時々刻々と変化していくため、常に周囲の状況に注意を払って、「適切な速度」という基準自体を更新し続けなければならない。あるいは別の例として、友人グループでの会話が挙げられる。会話においては、次々と切り替わっていく話題にうまくのって円滑にコミュニケーションをとるためには、過去の話題に固執しすぎず、適度に頭を切り替えて、いま話されていることに焦点を合わせる必要がある。

このように、周囲の状況が刻々と変化していく環境では、我々は自身のとるべき行動を柔軟に変化させる必要がある。このような柔軟性は認知的柔軟性と呼ばれ[Scott, 1962], 現在までに、実験課題による参加者の認知的柔軟性の特徴付けやその神経基盤、さらには薬物依存症や強迫性障害などの精神疾患との関連などが研究されてきた[Izquierdo et al, 2017].

認知主体の認知的柔軟性を特徴づける実験課題として、確率的逆転学習課題が知られる[Izquierdo et al, 2017]. この課題では、はじめに刺激と報酬の対応関係を学習した後、この対応関係を課題中に変化させ¹, これに対する応答を調べることで、被験者の認知的柔軟性を特徴づけることができる(図1に課題の模式図を示した)。

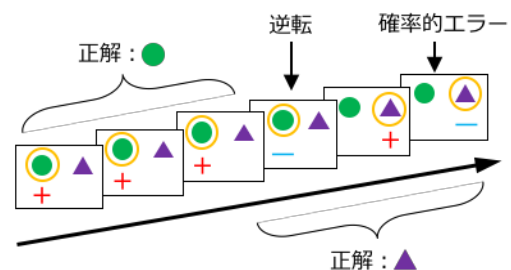


図1：確率的逆転学習課題の模式図。○は参加者による選択を表す。また、+は正解フィードバック、-は不正解フィードバックを表す。

従来この課題は、逆転が生じる前に学習した内容を適切に抑制できるかどうかを調べるために用いられる事が多かった。しかし近年になって、刺激と報酬の対応関係のパターンを課題の「状態」ととらえ、参加者が、状態やその切り替わりといった、事前に教示を受けていない、課題の潜在的な構造を学習することが柔軟な意思決定の基盤になっていることを示唆する知見が示されている[Wilson et al, 2014]. しかし、この知見はまだ十分に検証されてはおらず、さらに、被験者が課題の潜在構造を学習し、これに基づいた意思決定を行うようになっていく過程などについては未解明な点が多い。

また、薬物依存やアルコール依存などの物質依存症、あるいはギャンブル依存などの行動嗜癖の罹患者を対象とした研究を通じて、課題の成績と、衝動性傾向などの心理的・性格的傾向が相関を示すことが知られている[Izquierdo et al, 2017]. 従って、個々の被験者の心理的・性格的傾向を測り、これを他の測定量と合わせて被験者の振る舞いを分析するために

¹ 刺激が2種類の場合は逆転(反転)になる。

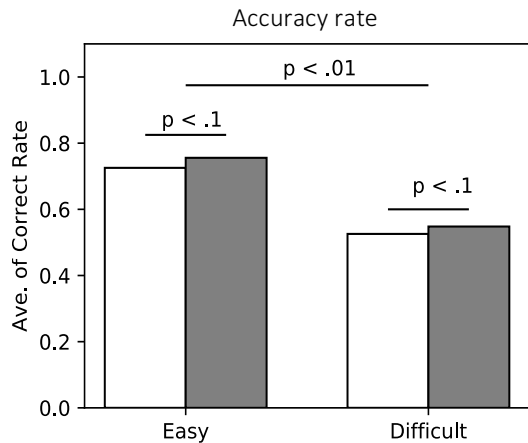


図2：正答率。縦軸：正答率の被験者平均，横軸：課題の難易度（易条件（Easy），難条件（Difficult））。白は課題前半，灰色は課題後半を示す。

利用することは，認柔軟な意思決定の認知基盤を明らかにするうえで有益であると考えられる。

そこで本研究では，確率的逆転学習を用いた実験を実施し，被験者が課題の潜在構造を学習しているかどうかを検証する。そのために，正答率や反応時間などの行動指標，および，強化学習モデルを用いた行動データの計算モデリングによる分析を行う。さらに，質問紙調査による心理尺度の測定を実施し，これら異なる分析手法を組み合わせ，参加者の意思決定や学習，およびそれらが各参加者の個別的特徴といかなる関連を持っているかを分析・検討する。

2. 方法

本研究で用いた実験課題について説明する。この課題では，被験者は提示された画像刺激の弁別を行う。被験者は，パソコン画面に表示された2種類の画像のどちらか一方を，キーボード入力によって選択する。各画像は一方が「正解」，他方が「不正解」に対応付けられていて，被験者による選択の後，自身が選んだ画像が「正解」であったか否かのフィードバックが画面に表示される。各画像刺激と「正解」・「不正解」の対応付けは20トライアル（=1ブロック）ごとに逆転するものとした。また，この課題には「確率的なエラー」があり，本来「正解」に対応付けられているはずの画像が「不正解」となることがある。このエラーの発生確率を p として， $p=0.3$ および $p=0.1$ のもとで実験を実施した（前者を難条件，後者を易条件と呼ぶ）。試行数は条件ごとに200トライアルとした（100トライア

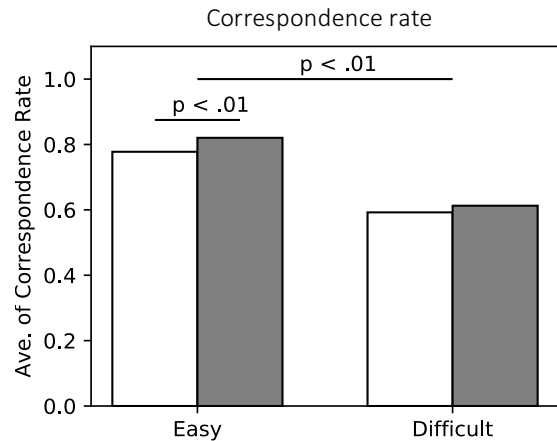


図3：一致率。縦軸：一致率の被験者平均，横軸：課題の難易度（易条件（Easy），難条件（Difficult））。白は課題前半，灰色は課題後半を示す。

ルを基準にして全トライアルを2つに分け，それぞれ前半・後半と呼ぶ。参加者には「正解は不変ではないこと」のみを伝え，なるべく多く「正解」を選択するように教示した。ブロックの存在やその切り替わり，エラーの発生については教示しなかった。被験者は健常な青年男女30名（男性27名，女性3名）で，平均年齢は19.43歳であった（18~22歳，SD=1.20）。

3. 結果

3.1. 行動指標

まずはじめに，易・難条件間，および各条件下での課題の前半・後半の間で，正答率を分析した（図2）。正答率とは「全トライアル数に対する，“正解”フィードバックを得た割合」である。分析の結果，正答率は，難条件に比べて易条件の方が有意に高いことがわかった（ $p < .01$ ）。また，易・難条件の双方において，前半よりも後半の方が，正答率が高いという有意傾向が見られた（ $p < .1$ ）。また，一致率についても同様の分析を行った（図3）。一致率とは，“全トライアル数に対する，“各ブロック内で正解と定義された”画像を選んだ割合”である。その結果，一致率は，難条件に比べて易条件の方が有意に高く（ $p < .01$ ），かつ，易条件では前半よりも後半の方が，一致率が有意に高いということがわかった（ $p < .01$ ）。

次に，被験者の意思決定（画像の選択）が，どの程度，被験者による潜在構造の推定に依拠しているかを調べるため，意思決定における一貫性を分析した。具体的には「 t 回目のトライアルで“不正解”となったにも関わらず，直後（ $t+1$ 回目）のトライア

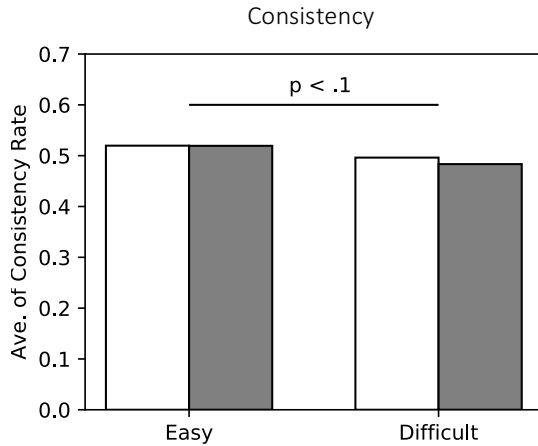


図4：一貫性．縦軸：一貫性の被験者平均，横軸：課題の難易度（易条件（Easy），難条件（Difficult））．白は課題前半，灰色は課題後半を示す．

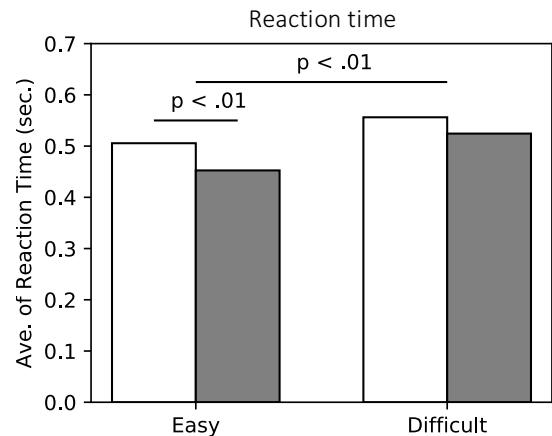


図5：反応時間．縦軸：反応時間の被験者平均，横軸：課題の難易度（易条件（Easy），難条件（Difficult））．白は課題前半，灰色は課題後半を示す．

ルにおいて同じ画像を選択した割合」を分析した．その結果，易条件の方が難条件よりも一貫性が高いという有意傾向（ $p < .1$ ）が見られた．

さらに，被験者の意思決定の過程をより詳細に分析するため，画像表示からキー押下までの反応時間を分析した．反応時間の長さは，被験者の確信度（迷い）と相関があると予測され，反応時間が短いほど課題の学習が進んでいることを示唆する．分析の結果，難条件に比べて易条件の方が，反応時間が有意に短いことがわかった（ $p < .01$ ）．また易条件では，前半よりも後半の方が，反応時間が有意に短かった（ $p < .01$ ）．

3.2. 強化学習モデルを用いたモデルベース解析

3.2.1. 強化学習モデル

上記の行動指標に基づく分析に加え，本研究では，得られた行動データに対して強化学習モデルを用いたモデルベース解析を行った．強化学習とは，自身がとった行動に対する報酬の重み付き累積値によって行動を価値付けし，その価値の条件付き確率に従って行動を選択するモデルである．この重み付き累積値のことを価値関数という．

本研究では，被験者が課題の潜在構造を学習しているかどうかを検証するために，課題の潜在構造を「未学習」および「学習済み」と仮定したモデルを用いて，実験で得られた行動データに対する尤度を計算し，モデル選択を行った．

本研究で用いた強化学習モデルは，Rescorla-Wagner モデル（RW モデル），Q-learning モデル，状態推定付き Q-learning モデルの3つである．RW モデルは以下の式(1)で定義される．

$$Q(a_t) = Q(a_t) + \alpha(r(a_t) - Q(a_t)) \quad (1)$$

ここで $Q(a_t)$ は，時刻 t で選んだ行動 a_t の価値関数を表す． r_t は時刻 t において得た報酬を表す． α は学習率と呼ばれ，価値関数の更新速度を規定する．式(1)の第二項，すなわち，得られた報酬と価値観数の差分は報酬予測誤差と呼ばれ，もしこの値が正であれば価値関数は増加し，負であれば減少する．

Q-learning モデルは以下の式(2)で定義される．

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r(s_t, a_t) - Q(s_t, a_t)) \quad (2)$$

式(2)における s_t は課題の「状態」を表す．RW モデルとの違いは，この状態変数を有する点である．ここで「状態」とは，本研究の場合，逆転学習課題における「刺激と正解・不正解の対応付けのパタン」のことを指す．本研究では2種類の刺激を用いたので，対応付けのパタン数は2である．すなわち，課題の状態は2つあり，これを $s_t = \{0, 1\}$ と表す．

他方，被験者が「状態」の情報を活用して意思決定を行うためには，彼らは毎回のトライアルにおいて，現在の状態をある程度正確に推定することが必要になる．この点，本研究で用いた Q-learning モデルでは，各トライアルで被験者は，現在の状態を

正確かつ確実に（確率1で）識別できるものと仮定した。

しかし、この仮定は明らかに強すぎるため、新たに3つのモデルとして、状態推定付き Q-learning モデルを導入する。このモデルは、式(2)で表される仕方で価値関数を更新することに加え、被験者が現在の状態を確率的に推定することを仮定したモデルである。このモデルでは、被験者は毎トライアル、現在の状態が $s_t = 0$ なのか $s_t = 1$ なのか推定し、この推定に基づいて意思決定を行う。この推定は確率 q で正解するものとし、 q の値が1に近いほど正確に推定できているとする。言い換えれば、状態推定付きモデルは、Q-learning モデルに対して q というパラメータが1つ加わったモデルだと考えることができる。あるいは、Q-learning モデルは、状態推定付き Q-learning モデルにおいて q の値を $q = 1$ とした場合に相当する。

上記3つのモデルのうち、RW モデルが、潜在構造を「未学習」だと仮定したモデルであり、Q-learning モデルと状態推定付き Q-learning モデルが「学習済み」と仮定したモデルである。

これらのモデルでは、行動選択確率（画面の左右に表示された二つの画像刺激のうち的一方を選択する確率）は以下で定義される。

$$p(a_t|Q(a_t)) = \frac{\exp(\beta Q(a_t))}{\sum_{L,R} \exp(\beta Q(a_t))}$$

ここで β は鋭敏性と呼ばれ、選択のランダムネスを特徴づけるパラメータであり、0の場合に一様ランダム、無限大のときには一方の選択肢を決定論的に選ぶようになる。

3.2.2. モデル選択

本研究では、前節で述べた3種類の強化学習モデルを用い、実験で得られた行動データについて、AIC（赤池情報量基準）に基づくモデル選択を実施した。AIC は以下で定義される量である。

$$AIC = -2 \log L + 2k$$

ここで L は尤度、 k はモデルの自由パラメータの数を表す。AIC の値が小さいほど、データをよく説明するモデルとされる。

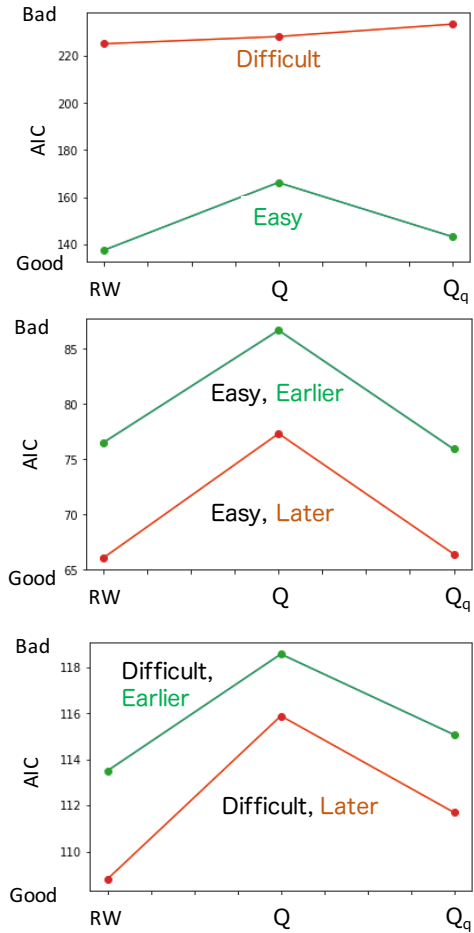


図6：各条件における AIC。縦軸：AIC の値，横軸：モデル。状態推定付き Q-learning モデルの状態推定確率の値は $q = 0.9$ とした。Q_q は状態推定付き Q-learning モデルを表す。

前節で述べた3種類の強化学習モデルについて、それぞれの自由パラメータの数 k は、RW モデルが $k = 3$ 、Q-learning モデルが $k = 4$ 、状態推定付き Q-learning モデルが $k = 5$ である。

各モデルについて AIC を計算した結果を図6に示す。AIC によるモデル選択の結果、被験者平均としては、難条件の前半と後半のそれぞれにおいて、RW モデルが選択されるか傾向があることがわかった。また他の条件では、RW モデルか状態推定付き Q-learning モデルのいずれが選択されるか傾向がある、という結果が得られた。

さらに、各被験者の個別の振る舞いを把握するために、被験者ごとに状態推定付き Q-learning モデルの AIC と RW モデルの AIC の差分を計算した結果を図7に示す。この値が負の場合、その被験者については状態推定付き Q-learning モデルが選択されることを意味する。これを見ると、被験者平均としては RW モデルが選択される傾向がある一方で、

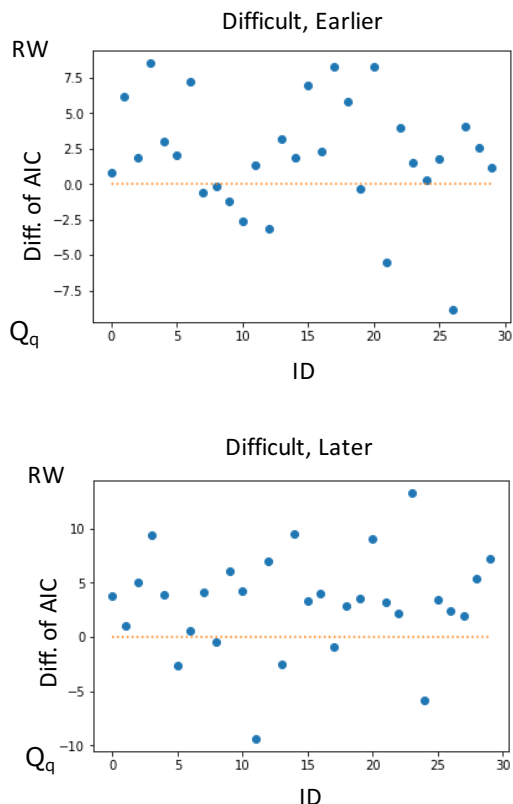


図7：各被験者におけるAICの差。縦軸： $AIC(Q_q) - AIC(RW)$ の値，横軸：被験者ID。状態推定確率の値は $q = 0.9$ 。上図は難条件・前半，下図は難条件・後半

なかには状態推定付き Q-learning モデルが選択される被験者がいることがわかった。特に難条件の前半・後半においては，被験者全体の 25%程度で状態推定付き Q-learning モデルが選択されることがわかった。

上記の結果を踏まえて，状態推定付き Q-learning モデルによって行動がよく説明される被験者が，どの程度正確に，課題の潜在構造（状態）を推定できているのかを把握するための分析を行った。具体的には，状態推定付き Q-learning モデルについて，各条件で q の値を $[0.0, 1.0]$ の範囲で 0.1 刻みで変化させて AIC を計算した結果を図 8 に示す。これを見ると，いずれの条件においても，状態推定確率 q の値の増加とともに AIC が減少し， $q = 0.9$ 付近で最小値をとることがわかった。これはすなわち，状態推定付き Q-learning モデルにおいては，現在の状態を高い確率で正確に推定できることを仮定したモデルの方が，行動データをよく説明するということを意味している。一方で， $q = 1.0$ においては

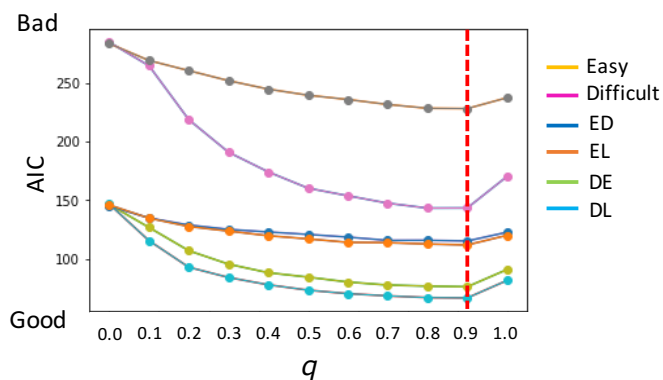


図8：状態推定付き Q-learning モデルの AIC。縦軸：AIC の被験者平均，横軸：状態識別確率 q 。各線は，それぞれ各条件を表す。図中の赤破線は，AIC が最小となる q の値を示す。

AIC の値が上昇しているため，現在状態を確実に推定できるモデルは行動データをうまく説明しないモデルであることがわかる。

3.3. 心理尺度

本研究では，逆転学習課題を用いた行動実験とは別に，質問紙調査による心理尺度の測定を実施した。具体的には，以下の 4 つの尺度を測定した。

- エフォートフル・コントロール (ECSA) [山形他, 2005]
 - 抑制制御，集中力，注意の切替えのうまさ
- 行動制御系・行動賦活系尺度 (BISBAS) [高橋雄介・繁樹, 2008]
 - 罰・報酬感受性，罰の回避，報酬への接近傾向の強さ
- 衝動性 (BIS11) [小橋・井田, 2013]
 - 衝動的行動傾向の強さ
- 首尾一貫感覚 (SOC, Sense of Coherence) [Antnovsky, 1987]
 - 先を見通せることによる，ストレス耐性の強さ

これらの尺度では，質問紙に対する解答からスコアが得られ，その値が高いほど，被験者は，各質問紙が対象とする心理的・認知的傾向を強く持つものとされる。

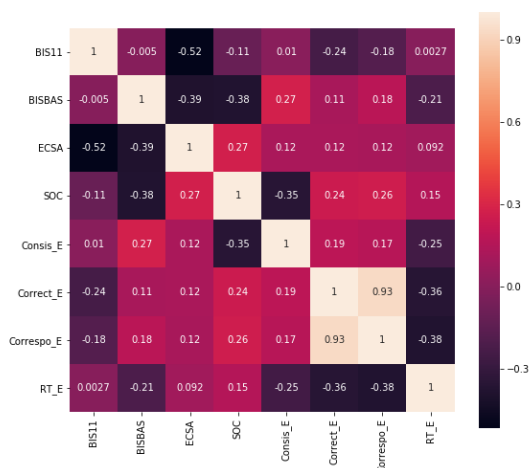


図9：行動指標と心理尺度との間の相関行列。図中の値は Spearman の順位相関係数を表す。例として、易条件での計算結果を示す。

上記の心理尺度を測定したうえで、それらと実験で測定した行動指標との間の相関を分析した。

図9に、行動指標と心理尺度との間の相関行列を示す。これは、4つの行動指標（正答率、一致率、一貫性、反応時間）と4つの心理尺度（ECSA, BISBAS, BIS11, SOC）との間で、Spearman の順位相関係数を計算したものである。図9には例として易条件での計算結果を示してあるが、分析は全ての条件に対して実施した（また図9には、行動指標同士、および心理尺度同士との間の相関を計算した結果も含まれている）。分析の結果、易条件においては、「衝動性と正答率の間の負の相関」および「首尾一貫感覚と正答率の間の正の相関」があることがわかった。また、難条件では「罰・報酬感受性と正答率の間の正の相関」があることがわかった。衝動性、および首尾一貫感覚はともに、刺激に対する即時的応答をどの程度抑制できるかを反映すると考えられる。また罰・報酬感受性は、自身が受けたフィードバックに対する行動の可塑性の程度を反映すると考えられる。これを踏まえると、得られた結果は、課題において高い成績を得るためには、直近のトライアルで得られた報酬（フィードバック）に対して即時的な応答をするのではなく、得られた即時的報酬と長期的報酬（期待報酬）との間の関係性のある程度よく“吟味”する必要があることを示唆している。

4. まとめ

本研究では、逆転学習課題を用いた行動実験を実施し、被験者の振る舞いを、行動指標、強化学習モデルを用いたモデルベース解析、および、心理尺度という異なる手法によって分析した。

行動指標、およびモデルベース解析の結果は、易条件で参加者は、トライアルを重ねるにつれて課題の潜在構造を学習し、それによって成績が向上していることを示唆している。また心理尺度と行動指標との相関の分析結果は、課題で高い成績を得るためには、報酬に対する即時的応答を抑え、長期的報酬を志向することの重要性が示唆された。今後は、認知的柔軟性の基盤となる認知機能の計算論的理解へ向けて、モデルベース解析で得られた結果と行動指標、および心理尺度との間の関連をより詳細に分析する必要があるだろう。

本研究で行ったモデルベース解析では、上述のように、被験者の行動を説明するモデルとして、被験者平均では、RW モデルが選択される傾向が見られた。これは「課題を解くために潜在構造を学習・活用している」という。先行研究で得られている知見と一見食い違っているように見える。本論の最後に、この点について議論する。

先行研究では、脳計測実験および脳損傷個体の行動データから、課題の潜在構造は眼窩前頭皮質（Orbitofrontal Cortex, OFC）において表象されていることを示唆する知見を示している [Wilson et al, 2014]。一方、ヒトの発達過程において、OFC を含む前頭前野は、報酬系などの進化的に古い脳領域に比べて発達が遅いことが知られ、特に20歳前後の思春期においては、この発達のアンバランスが、リスクテイキング行動や新奇性探索傾向といった、思春期に特徴的な振る舞いを引き起こすことが知られている [Hartley and Somerville, 2015]。さらに、[Hauser et al, 2015]では、本研究で用いたRWモデルは、成人期よりも思春期の行動の方をよく説明することが示されている。

一方、本研究では20歳前後の若年者が被験者となっている（平均19.43歳（18~22歳, SD=1.20））。これらの知見および事実を踏まえると、本研究で得られた結果が先行研究で得られた知見と異なるものとなったのは、対象とした被験者の年齢が若く、成人に比べて前頭前野が未成熟であることを反映したものである可能性がある。しかし、図7で示したように、なかにはRWモデルではなく、状態推定付きQ-learningモデルによってその

行動がよく説明される被験者もいるため、一概には言えない。言い換えると、本研究で得られた結果は、各被験者の脳（特に前頭前野）の発達における個体差が反映されたものである可能性がある。従って今後は、成人期の被験者を対象とした実験を実施し、その結果を今回得られた結果と合わせ、世代間比較を行い、上述の議論を検証する必要があるだろう。これにより、認知的柔軟性の基盤となる脳領域や認知機構を発達段階ごとに特徴づけるための端緒とできると考えられる。

謝辞

本研究はJSPS科研費JP16H06397の支援を受けた。

参考文献

- [1] W.A. Scott, “Cognitive complexity and cognitive flexibility,” *Sociometry*, pp. 405–414, 1962.
- [2] A. Izquierdo, J. Brigman, A. Radke, P. Rudebeck, and A. Holmes, “The neural basis of reversal learning: an updated perspective,” *Neuroscience*, vol. 345, no. 2, pp. 12–26, 2017.
- [3] R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, and Y. Niv, “Orbitofrontal cortex as a cognitive map of task space,” *Neuron*, vol. 81, no. 2, pp. 267–279, 2014.
- [4] 山形伸二, 高橋雄介, 繁榎算男, 大野裕, & 木島伸彦. (2005). 成人用エフォートフル・コントロール尺度日本語版の作成とその信頼性・妥当性の検討. *パーソナリティ研究*, 14(1), 30-41.
- [5] 高橋雄介・繁榎算男. (2008). 罰の回避と報酬への接近の感受性を測定する 3 尺度の比較. *パーソナリティ研究*, 17, 72–81.
- [6] 小橋真理子, & 井田政則. (2013). 改訂日本語版 BIS-11 の作成: 信頼性と妥当性の検討.
- [7] Antonovsky, A. *Unraveling The Mystery of Health - How People Manage Stress and Stay Well*, San Francisco: Jossey-Bass Publishers, 1987
- [8] Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision-making. *Current opinion in behavioral sciences*, 5, 108-115.
- [9] Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., & Brem, S. (2015). Cognitive flexibility in adolescence: neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage*, 104, 347-354.