Tsukuba Psychological Research
2023, **61**, 53-61

Q学習モデルにおけるモデルリカバリ率の確認と改善[1)]

名古屋大学　藤田　和也[2)]

筑波大学人間系　山口　一大[3)]

Confirming and improving the model recovery rates of Q-learning models

Kazuya Fujita (*Graduate School of Informatics, Nagoya University, 464-8601, Japan*)
Kazuhiro Yamaguchi (*Faculty of Human Sciences, University of Tsukuba, 305-8572, Japan*)

Several studies that discuss the issues of determining an appropriate model highlight the importance of model selection. The reliability of a model selection result depends on its model recovery rate, referring to the probability of selecting an appropriate model that can be enhanced by improving its model recovery rate. In this paper, we confirm the model recovery rate of a Q-learning model and further improve its model recovery rate by manipulating the stimuli, particularly reward probabilities. First, we confirm that the model recovery rates of four basic Q-learning models are insufficient at approximately 60 percent. Second, we confirm that the model recovery rates increased by between 10～20 percent due to inversing the reward probability. It is important to verify the model recovery rate, which is fundamental quantity. Moreover, the model recovery rate can be increased by appropriately manipulating the stimuli, which can be controlled by an experimenter.
Key words: Model recovery rate, Q-learning model, adaptive stimulus selection, KL information

1．序

心理学実験の中には，モデル選択が重要である研究がある。認知実験の中では，認知プロセスを表現する認知モデルを複数用意して，そのどれが尤もらしいかをデータから判断する研究が典型例である。また重回帰分析における変数選択もモデル選択の一種と考えることができる。とりわけベイズ統計学やベイジアン認知モデリング（Gelman et al., 2014; Lee & Wagenmakers, 2013; 豊田編, 2017）が広まりつつある現在においては，この認知モデルに関するモデル選択の問題は益々重要になると予想される。ベイジアン認知モデリングにより，複雑な認知プロセスを表現したモデルを利用した分析の可能性が広がるからである。

モデル選択を利用した研究においては，p(*fit model* | *simulated model*) と p(*simulated model* | *fit model*) の２つの確率が重要な情報となる（Wilson & Collins, 2019）。前者は，シミュレーション上の真のモデルである simulated model でデータ生成した時に，fit model が選ばれる確率である。例えば，モデル1, 2, 3があった時に，simulated model がモデル２だとする。この時，実際にモデル２を元にデータ生成してモデル選択した場合に，モデル１が選ばれる確率，モデル２が選ばれる確率，モデル３が選ばれる確率を考えることができる。当然，モデル２が選ばれる確率が高いほど良い。各モデルが選ばれる確率を行列の形でまとめたものを confusion matrix と呼ぶ（表１を参照）。また，p(*simulated model* | *fit model*)は，モデル選択の結果，あるモデル（fit model）

1）データ分析に使用したコードは Open Science Framework（https://osf.io/w7n5r/）で公開している。
2）〒464-8601 愛知県名古屋市千種区不老町
E-mail: fujita.kazuya@k.mbox.nagoya-u.ac.jp
ORCID: https://orcid.org/0000-0002-1062-3215
3）ORCID: https://orcid.org/0000-0001-8011-8575

が選ばれた時に，そのモデルがデータ生成モデル（simulated model）である確率である。実験によりデータを取得しモデル選択をした場合でも，選ばれたモデルが真であるとは断定できないため，この確率が重要となる。各モデルがデータ生成モデルである確率を行列の形でまとめたものをinversion matrixと呼ぶ。inversion matrixはベイズの定理を利用すればconfusion matrixから計算できる。ただし考察で触れるが，inversion matrixも，候補モデルの中に厳密に真のモデルが含まれている場合と含まれていない場合で，意味が異なる。本研究ではconfusion matrixをシミュレーションにより求めることで，モデルリカバリ率の確認を行う。モデルリカバリ率とは，各モデルが選ばれる確率である。特に，真のモデルが実際に選ばれる確率が重要である。confusion matrixは対角要素が1で，非対角要素が0である単位行列に近いほど良い。つまり真のモデルが100%選ばれる状況に近いほど望ましい。

近年，臨床心理学領域との関連で，認知モデルの中でもQ学習モデル（Daw, 2009; 片平, 2018; Watkins & Dayan, 1992）が計算論的精神医学（国里他, 2019）の中で注目されている。例えば，抑うつなどの臨床的な性質と，Q学習モデルのパラメータの相関が研究されている（Chase et al., 2010）。Q学習モデルは多腕バンディット課題に対して適用されることが多い。多腕バンディット課題では，複数のスロットマシーンがあり，参加者はその内1つのマシンを選択してギャンブルを行う（Steyvers et al., 2009）。当たりが出れば報酬が与えられ，はずれであれば報酬は与えられない。これを数十から数百試行行う。参加者は報酬最大化を目指す。Q学習モデルにおいては，神経科学における知見も踏まえ（Schultz, Dayan, & Montague, 1997），Q値と実際にフィードバックされる報酬額との差（報酬予測誤差）に基づいてQ値を更新していくと仮定する（詳しくは2.1節参照）。Q値とは各選択肢の価値（魅力度）のようなものである。

Q学習モデルにはいくつかの基本的なモデルが存在する。本研究では最も基礎的なモデルを標準的なQ学習モデルと呼ぶことにする（詳しくは2.1節参照）。本研究では以下の3つのモデルも含め，計4つのモデルを考える。標準的なQ学習モデルでは選択されなかったスロットのQ値は更新されないと仮定するが，選択されなかったスロットのQ値も更新（減衰）されるとする忘却モデル（Forgetting model, F model; Ito & Doya, 2009; Toyama, Katahira, & Ohira, 2019）が対象モデルの1つである。また，報酬予測誤差が正（獲得）と負（損失）で異なる学習率を仮定する非対称モデル（Asymetric model, A model; Chase et al., 2010; Kunisato et al., 2012）もある。さらに，同じ選択を繰り返す固執性という性質を入れたモデル（Katahira, 2018）も対象とする。先行研究によっては，異なるモデルを仮定してパラメータ推定を行ったり，モデル選択結果が異なったりする。著者が知る限り，これらの基本的なQ学習モデルを考えた時の，モデルリカバリ率ですら不明である。

またモデルリカバリ率に影響する要因として，認知実験における実験刺激も重要である。実験刺激は実験者側がコントロールできるものであり，パラメータの推定精度や検出力に影響するだけでなく（Heck & Erdfelder, 2019; Murphy & Brincke, 2018），モデル選択結果に影響する。そのため刺激を適切に選ぶことによって，推定精度やモデル選択などのパフォーマンスを向上させることも可能である。本研究では特に，当たりが出る確率（報酬確率）に着目する。Q学習モデルにおける典型的な選択肢数2のバンディット課題では，スロットAおよびBの報酬額が1で，スロットAの報酬確率とスロットBの報酬確率の和が1になるように設定されることが多い（e.g., Gershman, 2016）。またスロットAとスロットBの報酬確率を逆転させる操作（報酬確率逆転）を行う場合（e.g., Dezfouli et al., 2019; Katahira et al., 2011）と行わない場合（e.g., Beevers et al., 2013）もある。本研究では選択肢数が2のバンディット課題を対象とする。

モデル選択の観点からの刺激選択法としてはAdaptive design optimization（ADO; e.g., Myung et al., 2013; Myung & Pitt, 2009; Cavagnaro et al., 2010; Cavagnaro et al., 2013a, b）が提案されている。またバンディット課題を使用して，モデル選択の観点から刺激選択法を調べた研究もある（Zhang & Lee, 2010）。このような客観的かつ最適な刺激選択法においては，様々な刺激候補の各値における“効用”（刺激の良さ）を定義して，効用が最大になる刺激を選択することになる。モデル選択の観点からの一つの効用の定義としてKullback-Leibler（KL）情報量があり（Cavagnaro et al., 2010），本研究でも付録においてKL情報量を利用して刺激の良さを検討する。KL情報量とは，モデルAとモデルBの違いや離れ具合を表すような量である。

本研究で行うことは以下の3点にまとめられる。1点目は，まず基本的なモデルリカバリ率を確認することである。特に先述した基本的な4モデルを考えた状況において，confusion matrixを算出する。2点目は，KL情報量を見ることで，モデル選択の観点

において刺激の良さを検討することができることを説明する。特に報酬確率逆転の効果を検討する。3点目は，実際に刺激デザインを変更することでモデルリカバリ率を向上できることを示す。特に報酬確率逆転によりモデル選択のパフォーマンスが向上することを示す。

本論文の構成は以下の通りである。2節ではQ学習モデル，モデル選択，KL情報量の基本的な部分を説明する。3節ではシミュレーション実験を行う。4節において総合考察を行う。

2．方法

2.1　Q学習モデル

選択肢が2つのバンディット課題では，参加者はスロットA, Bの内どちらかを選択してギャンブルを行う。スロットを表す文字をkとし，$k=1$はスロットAを$k=2$はスロットBを表すとする。また，$RP_{t,k}$と$Re_{t,k}$をそれぞれ，試行tにおける報酬確率と報酬額（の候補値）とする。$Re_{t,k}$は当たりが出た場合に得られる報酬額である。外れた場合は報酬額は0とする。実際に参加者にフィードバックされる報酬結果は$r_{t,k}$と表し，$Re_{t,k}$とは区別する。先行研究における刺激デザインは，$Re_{t,k}=1$, $RP_{t,2}=1-RP_{t,1}$（e.g., Gershman, 2016）と設定することが多い。本研究でもこの設定を使用する。

まず標準的なQ学習モデル（Standard Q-learning model; Sモデル）を説明する。Q学習モデルではQ値の更新において，Q値とフィードバックされる報酬額の差分（報酬予測誤差），$r_{t,k}-Q_{t,k}(\theta)$，が更新量の規定要因となる。ただし，$Q_{t,k}(\theta)$は試行tにおけるスロットkのQ値であり，θはパラメータベクトルでSモデルの場合は学習率パラメータα $(0\leq\alpha\leq1)$と逆温度パラメータβ（$0<\beta$）である。y_tを意思決定データとして，$y_t=1$はスロットAを，$y_t=0$は試行tでスロットBを選んだことを意味するものとする。Q値は

$$Q_{t,k}(\theta)=Q_{t-1,k}(\theta)+I_{\mathrm{A}}\,(k-1+y_{t-1})\alpha\Delta_{t-1,k,} \qquad (1)$$

$$\Delta_{t,k}=r_{t,k}-Q_{t,k}(\theta) \qquad (2)$$

という式で更新されると仮定する。ただし，$I_A(x)=1$ *if* $x=1$, $I_A\,(x)=0$ *if* $x\neq1$ である。$I_A(k-1+y_{t-1})$は要するに，スロットkをt−1試行目で選んでいたら1を，選んでいなかったら0を返す関数である。今後は省略して$I_A(k,y_{t-1})$と書く場合がある。

試行tにおける，スロットAの選択確率$p(y_t=1\,|\theta)$については，

$$p(y_t=1\,|\theta)=\frac{1}{1+\exp(-\beta(Q_{t,1}(\theta)-Q_{t,2}(\theta))} \qquad (3)$$

と仮定する。そしてデータ生成は

$$y_t\sim Bernoulli(p(y_t=1\,|\theta)) \qquad (4)$$

と仮定する。

Q値の更新式を修正した拡張モデルとして，忘却モデル（Fモデル）と学習率の非対称モデル（Aモデル）がある。本研究においてはFモデルでは，t−1試行目で選択されなかった選択肢のQ値が0に近づく（忘却）と仮定する。式で書くと

$$Q_{t,k}(\theta)=Q_{t-1,k}(\theta)+I_A(k,\,y_{t-1})\alpha\Delta_{t-1,k}-(1-I_A\,(k,\,y_{t-1}))\alpha_FQ_{t-1,k}(\theta) \qquad (5)$$

となる。ただしα_F（$0\leq\alpha_F\leq1$）は忘却率パラメータである。

Aモデルでは，報酬予測誤差$\Delta_{t,k}$の符号によって異なる学習率α_L^+，α_L^-を仮定する。つまり，Q値の更新式は

$$Q_{t,k}(\theta)=Q_{t-1,k}(\theta)+I_A(k,\,y_{t-1})\{I(\Delta_{t-1,k}\geq0)\alpha_L^+\Delta_{t-1,k}+I(\Delta_{t-1,k}<0)\alpha_L^-\Delta_{t-1,k}\} \qquad (6)$$

とかける。ただし，$I(\cdot)$は中身の条件が満たされたら1，満たされなかったら0を返す関数である。

また選択確率のモデルの方を修正した固執性モデル（perseverance model; Pモデル）も存在する。Pモデルでは，基本的に選択されたスロットは次試行も選ばれやすくなるという仮定を置く。これを表現するために，選択確率のモデルを

$$p(y_t=1|\theta)=\frac{1}{1+\exp(-\beta(Q_{t,1}(\theta)-Q_{t,2}(\theta))-\phi C_t(\tau))} \qquad (7)$$

と修正する。$C_t(\tau)$がt試行目の固執性の程度（choice trace; Katahira, 2018）を表しており，選択行動への影響の強さをパラメータϕが決めている。Pモデルでは，この$C_t(\tau)$は

$$C_t(\tau)=(1-\tau)C_{t-1}(\tau)+(-1)^{1-y_{t-1}}\tau \qquad (8)$$

という更新式を仮定する。なお，τ（$0\leq\tau\leq1$）はchoice traceの減衰率パラメータである。

2.2　モデル選択に使用する情報量規準

本研究において以下のAkaike Information Criterion (AIC; Akaike 1973; Symonds & Mousslli, 2011)，Bayesian Information Criterion (BIC;

Claeskens, & Hjort, 2008)，およびWidely applicable (Watanabe) Information Criterion (WAIC; 渡辺，2012) を利用する。m 個目のモデルにおける AIC は

$$AIC_m = -2 \log p_m(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_m) + 2l_m \quad (9)$$

とかける。ただし，p_m はモデル m の尤度，$\boldsymbol{y} = (y_1, \cdots, y_T)$，$\boldsymbol{\theta}_m$ はモデル m のパラメータ，l_m はモデル m のパラメータ数，T は試行数を表す。また m 個目のモデルにおける BIC は

$$BIC_m = -2\log p_m(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_m) + \log(T)\ l_m \quad (10)$$

と書ける。WAIC (渡辺, 2012) に関しては

$$WAIC_m = -2\sum_t^T \log E_{post}\ [p_m(y_t|\boldsymbol{\theta}_m)] + 2\sum_t^T var_{post}(\log p_m(y_t|\boldsymbol{\theta}_m)) \quad (11)$$

で計算できる。ただし，E_{post},var_{post} は事後分布における期待値および分散である。AIC, BIC, WAIC いずれにおいても，モデル選択においては，その値が最小のモデルを最良のモデルとして選択する。

注意点として，本研究においては MCMC 推定で出した事後分布のモードを点推定値 $\hat{\boldsymbol{\theta}}_m$ として利用する。そのため，AIC においては厳密にはベイズ推定などにも利用できる一般化情報量規準など（小西・北川，2004）を利用すべきである可能性がある。学習率 α などに関しては無情報事前分布を利用しているため，そのまま AIC を利用している。また BIC の計算におけるサンプル数の部分はそのまま試行数 T を利用している。

2.3 KL 情報量

KL 情報量は，真の分布を q, 仮定した分布を p とした時，

$$KL(q;\ p) = \int q(x)\ \log \frac{q(x)}{p(x)}\, dx \quad (12)$$

と定義される（小西・北川，2004）。KL 情報量は，真のモデル q と仮定したモデル p の近さを測る測度と捉えられている。要するに KL 情報量は，対数尤度比の平均挙動を表している。今回のモデルの場合は，KL 情報量は

$$E_q\left[\log \frac{p(\boldsymbol{y}|\theta_q^*)}{p(\boldsymbol{y}|\theta_p^*)}\right] = \sum_t^T \left\{\eta_t^* \log \frac{\eta_t^*}{\eta_t} + (1-\eta_t^*)\log \frac{1-\eta_t^*}{1-\eta_t}\right\} \quad (13)$$

となる。ただし，θ_q^*,θ_p^*は真の分布および仮定した分布でのパラメータ真値，η_t^*,η_t は真の分布，仮定した分布での選択肢 A の選択確率である。つまり，S, F, A モデル間では更新された Q 値にズレが生じ，それが選択確率のズレに繋がり，結果として KL 情報量は正の値となり，モデル弁別が可能となる。刺激選択の観点から言うと，どれだけモデル間で選択確率の違いを生じさせることができるか，そのような実験刺激デザインは何か，という点が重要となる。

3．シミュレーション実験

3.1 方法

本節では，シミュレーションによって，4 つの Q 学習モデルを考えた時の，モデルリカバリ率を実際に求めることを目的とする。さらに，実験刺激デザインによってモデル選択のパフォーマンスが変わることを示す。どのような実験刺激デザインが良いか，特に報酬確率逆転の効果を KL 情報量によって検討する部分については付録で示す。

シミュレーションの流れは以下の通りである。1．まず刺激デザイン等を決める。報酬額は 1 で固定し，報酬確率は (0.8, 0.2) などの設定を予め決めておく。確率 (0.8, 0.2) とは，選択肢 A の報酬確率が0.8，選択肢 B の報酬確率が0.2であることを表す。2．真のデータ生成モデル m を固定する。3．データ生成モデル m から，反応データを生成する。2 節で説明したモデルおよび 1．で定めた真のパラメータ値を利用する。4．4 つの Q 学習モデル（S, F, A, P モデル）で MCMC 推定を行い，モデル選択を AIC, BIC, WAIC を規準として行う。このシミュレーションを50回繰り返す。2 ～ 4 のステップを，全てのモデル（m = 1,2,3,4）で繰り返す。

シミュレーションの設定は以下の通りである。試行数は200, パラメータの真値は $\alpha = 0.5$, $\alpha_L^+ = 0.7$, $\alpha_L^- = 0.3$, $\alpha_F = 0.3$, $\tau = 0.2$, $\phi = 1$, $\beta = 2$ とした。報酬確率は (0.5, 0.5)，(0.8, 0.2) 条件および，(0.8, 0.2) で報酬確率逆転あり条件の 3 条件を用意した。ただし報酬確率逆転は10試行ごとに行った。AIC の計算などには，MCMC 推定を行い事後分布のモード (maximum a posteriori probability; MAP) を利用した。事前分布は，α_L, α_L^+, α_L^-, α_F, $\tau \sim U(0,1)$, $\phi \sim$ *student* $t(4,0,3)$ $(-5 \le \phi \le 5)$, $\beta \sim$ *student* $t(4,0,3)$ $(0 \le \beta \le 10)$ を利用した。ただし，*student t*(*df*,*mu*,*scale*) は自由度 df，位置パラメータ mu，スケールパラメータ scale の t 分布を表す。MCMC 推定は，5000個のサンプルを取り出し，最初の2000個をウォームアップ期間として捨て，間引きは 1 として，3 チェイン

を回した。

3.2　結果と考察

モデルリカバリ率の結果（confusion matrix）を表したのが表1a~1c である。まず情報量規準により違いを見ておくと，BIC はシンプルな S モデルを選びがちである。特にPモデルが真の場合は，パフォーマンスが極端に悪い。AIC と WAIC は似た傾向である。AIC, BIC は参考程度に，理論的な整合性も考え今後は WAIC に特に着目して結果を見ていく。

全体的にモデルリカバリ率は 6 割程度とあまり良いパフォーマンスとは言えない（表1a）。このような状況では，実際の実験で例えば S モデルが選ばれても，（S，F，A，P モデルのどれかが厳密に真のモデルだったとしても）S モデルが真にデータ生成モデルであるとは断定できないことになる。F，A，P モデルが真の状況でも S モデルは選ばれることがあるためである。また，刺激デザイン（報酬確率）が異なると，モデル選択のパフォーマンスは変わっている（表1a，1b）。

さらに，報酬確率逆転の効果も表1b，1c を比べると見える。付録のシミュレーションから報酬確率逆転によってモデル弁別力が高まることが予想されている。実際に，報酬確率逆転によってモデルリカバリ率は，A モデルを除いて，10〜20％程度上昇している。このように，モデル選択のパフォーマンスは，実験者がコントロール可能な実験刺激デザインにより，ある程度改善できることが分かる。

4．総合考察

4.1　考察

本研究では，基本的な 4 つの Q 学習モデルにおける，モデルリカバリ率を確認した。また，KL 情報量によってどのような実験刺激でモデルリカバリ率が改善するか検討し，実際に MCMC ベースのシミュレーションによって，実験刺激によりモデルリカバリ率が変化することを確認した。

まず基本的なモデルリカバリ率は 6 割程度と高いものとは言えなかった。このことから，少なくとも

表 1a.
モデルリカバリ結果 : 報酬確率（0.5, 0.5）

(A) WAIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.56	0.16	0.28	0.24
	F	0.10	0.68	0.00	0.18
	A	0.22	0.12	0.62	0.10
	P	0.12	0.04	0.10	0.48

(B) AIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.78	0.22	0.36	0.32
	F	0.12	0.72	0.00	0.34
	A	0.06	0.02	0.54	0.02
	P	0.04	0.04	0.10	0.32

(C) BIC		真のモデル			
		S	F	A	P
仮定したモデル	S	1.00	0.34	0.82	0.64
	F	0.00	0.64	0.00	0.32
	A	0.00	0.02	0.18	0.02
	P	0.00	0.00	0.00	0.02

表 1b.
モデルリカバリ結果 : 報酬確率（0.8, 0.2）

(D) WAIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.68	0.20	0.34	0.46
	F	0.00	0.50	0.00	0.10
	A	0.14	0.14	0.62	0.18
	P	0.18	0.16	0.04	0.26

(E) AIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.82	0.28	0.44	0.58
	F	0.00	0.54	0.00	0.14
	A	0.08	0.10	0.56	0.14
	P	0.10	0.08	0.00	0.14

(F) BIC		真のモデル			
		S	F	A	P
仮定したモデル	S	1.00	0.56	0.78	0.88
	F	0.00	0.40	0.00	0.08
	A	0.00	0.02	0.22	0.04
	P	0.00	0.02	0.00	0.00

表 1c.
モデルリカバリ結果 : 報酬確率（0.8, 0.2）報酬確率逆転あり

(G) WAIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.78	0.14	0.28	0.18
	F	0.00	0.70	0.00	0.16
	A	0.16	0.06	0.62	0.16
	P	0.06	0.10	0.10	0.50

(H) AIC		真のモデル			
		S	F	A	P
仮定したモデル	S	0.90	0.16	0.50	0.28
	F	0.02	0.74	0.02	0.24
	A	0.06	0.02	0.46	0.10
	P	0.02	0.08	0.02	0.38

(I) BIC		真のモデル			
		S	F	A	P
仮定したモデル	S	1.00	0.48	0.86	0.70
	F	0.00	0.52	0.00	0.24
	A	0.00	0.00	0.14	0.02
	P	0.00	0.00	0.00	0.04

Note. S モデル＝標準的な Q 学習モデル , F モデル＝忘却モデル，A モデル＝学習率非対称モデル，P モデル＝固執性ありモデルを表す。確率（0.8, 0.2）はスロット A の報酬確率0.8, スロット B の報酬確率0.2を表す。

本研究で扱った4つのモデルを考える場合には，実際の実験で得られたモデル選択の結果あるいは選ばれたモデルをベースにした解釈については注意が必要であることが分かる。特に個人ごとにモデル選択結果を出している場合には，そのモデル選択結果を完全に信頼できるかは分からない。このことを踏まえると，モデルリカバリ率を何等かの方法で改善した方が良いことも分かる。そこで本研究では特に，報酬確率逆転の効果に着目し，報酬確率逆転を行うことによって，モデルリカバリ率が改善することを確認した。

ADOに関する先行研究においては，実験刺激によってモデル選択のパフォーマンスが改善できることが示されている（Cavagnaro et al., 2013; Myung & Pitt, 2009）。先行研究と異なる点は，本研究ではQ学習モデルという時系列依存があるモデルを扱った点と，adaptiveなデザインではなくfixed designを使用した点である。adaptiveデザインとは参加者の反応データを見てから実験中に刺激を適応的に変化させる手法で，fixed designとは実験実施前に提示する刺激を，参加者の特性に関係なく，決める手法である。Q学習は1試行ごとに報酬確率を変化させることはできないため，fixed designの方が便利な部分がある。Q学習モデルにおいてfixed designを想定した状況でも，実験刺激によってモデルリカバリ率は改善できた。このように，実験者がコントロールできる実験刺激という要因によって，モデル選択のパフォーマンスおよびモデル選択の結果がどれくらい信じられるかの程度が変わる。そのため，実験刺激デザインを客観的かつ適切に設計することは重要である。

4.2 今後の展望

本研究の限界を3点指摘しておく。第一に，confusion matrixは単位行列（対角部分が1で，それ以外は0）が理想的であるが，本研究の結果ではそうはなっていない。報酬確率逆転によりモデルリカバリ率は改善し，単位行列に近づいたものの，まだ改善の余地は残されている。モデルリカバリ率の改善が完璧ではなかった原因としては，fixed designであったこと，2腕バンディット課題という課題構造自体は固定していたこと，報酬確率しか動かしてないので刺激のパターン数が少ないなどの点が考えられる。場合によっては課題構造自体を変えたりといった，更なる工夫を行うことで，モデルリカバリ率を理想的な状況により近づけることが重要である。

第二に，本研究は個人ごとのデータに対して，個人ごとにモデル選択を行う状況を想定した分析となっている点である。実際には，階層モデルを組んでWAICなどで集団全体のデータを利用してモデル選択を行ったり，どの認知モデルに従うかを潜在変数として，混合モデルのように集団レベルの分布からその潜在変数が生成されたと仮定するモデルを使用して分析することもできる。この場合，参加者が複数になる分，全体のデータ数は増えるため，モデルリカバリ率は今回の結果より良くなると予想される。このような状況で，モデルリカバリ率は十分かを確認するのも重要である。

第三に，概念的にモデルリカバリ率は，考えるモデルセットによって値が変わる。モデルセットは考えている全てのモデルのことであり，今回の場合はS, F, A, Pモデルの4つのモデルのことである。例えば，新たなモデル5を入れた場合，本研究の結果とはconfusion matrixの値は変わると想定される。モデルセットが変わるたびに，再度計算し直す必要がある。またそもそも，厳密に真のモデルがモデルセットに含まれていない場合は，confusion matrixの結果をどう解釈するかなど理論的な部分を含めて議論する研究が今後必要になってくるだろう。

引用文献

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. & Csaki, F. (Ed). *Proceedings of the 2nd International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado. https://doi.org/10.1007/978-1-4612-1694-0_15

Cavagnaro, D. R., Gonzalez, R., Myung, J. I., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science*, *59*(2), 358–375. https://doi.org/10.1287/mnsc.1120.1558

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, *22*(4), 887–905. https://doi.org/10.1162/neco.2009.02-09-959

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. I. (2013). Discriminating among probability weighting functions using adaptive design optimization. *Journal of Risk and Uncertainty*, *47*(3), 255–289. https://doi.org/10.

1007/s11166-013-9179-3

Chase, H. W., Frank, M. J., Michael, A., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2010). Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychological Medicine*, *40*, 433-440. https://doi.org/10.1017/S0033291709990468

Claeskens, G. & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511790485

Daw, N. D. (2009). Trial-by-trial data analysis using computational models. *Attention and Performance*, 1-26. https://doi.org/10.1093/acprof:oso/9780199600434.003.0001

Dezfouli, A., Griffiths, K., Ramos, F., Dayan, P., & Balleine, W. (2019). Models that learn how humans learn: The case of decision-making and its disorders. *PLosComput Biol*, *15*(6), e1006903. https://doi.org/10.1371/journal.pcbi.1006903

Gelman, A., Carlin, J.B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis (3 ed)*. Boca Raton: CRC Press. https://doi.org/10.1201/b16018

Gershman, S. J. (2016). Empirical priors for reinforcement learning models. *Journal of Mathematical Psychology*, *71*, 1-6. https://doi.org/10.1016/j.jmp.2016.01.006

Heck, D. W., & Erdfelder, E. (2019). Maximizing the Expected Information Gain of Cognitive Modeling via Design Optimization. *Computational Brain and Behavior*, 2(3-4), 202-209. https://doi.org/10.1007/s42113-019-00035-0

Ito, M., & Doya, K. (2009). Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia. *The Journal of Neuroscience*, *29*(31), 9861-9874. https://doi.org/10.1523/JNEUROSCI.6157-08.2009

片平健太郎 (2018). 行動データの計算論モデリング 強化学習を例として オーム社

Katahira, K. (2018). The statistical structures of reinforcement learning with asymmetric value updates. *Journal of Mathematical Psychology*, *87*, 31-45. https://doi.org/10.1016/j.jmp.2018.09.002

Katahira, K., Fujimura, T., Okanoya, K., & Okada, M. (2011). Decision-making based on emotional images. *Frontiers in Psychology*, *2*, 311. https://doi.org/10.3389/fpsyg.2011.00311

小西貞則・北川源四郎 (2004). 情報量規準 シリーズ＜予測と発見の科学＞2. 朝倉書店

国里愛彦・片平健太郎・沖村 宰・山下祐一 (2019). 計算論的精神医学：情報処理過程から読み解く精神障害 勁草書房

Kunisato, Y., Okamoto, Y., Ueda, K., Onoda, K., Okada, G., Yoshimura, S., Suzuki, S., Samejima, K., & Yamawaki, S. (2012). Effects of depression on reward-based decision making and variability of action in probabilistic learning. *Journal of Behavior Therapy and Experimental Psychiatry*, *43*, 1088-1094. https://doi.org/10.1016/j.jbtep.2012.05.007

Lee, M. D. & Wagenmakers, E. J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge: Cambridge University Press.（井関龍太（訳）(2017). ベイズ統計で実践モデリング――認知モデルのトレーニング―― 北大路書房）https://doi.org/10.1017/CBO9781139087759

Murphy, R. O., & Brincke, R. H. W. (2018). Hierarchical Maximum Likelihood Parameter Estimation for Cumulative Prospect Theory: Improving the Reliability of Individual Risk Parameter Estimates. *Management Science*, *64*(1), 308-326. https://doi.org/10.1287/mnsc.2016.2591

Myung, J. I., Cavagnaro, D. A., & Pitt, M. A. (2013). A tutorial on adaptive design optimization. *Journal of Mathematical Psychology*, *57*, 53-67. https://doi.org/10.1016/j.jmp.2013.05.005

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*(3), 499-518. https://doi.org/10.1037/a0016104

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*, 1593-1599. https://doi.org/10.1126/science.275.5306.1593

Steyvers, M., Lee, M. D., & Wagenmakers, E. (2009). A Bayesian analysis of human decision-making on bandit problems. *Journal of Mathematical Psychology*, *53*(3), 168-179. https://doi.org/10.1016/j.jmp.2008.11.002

Symonds, M. R. E., & Moussalli, A. (2011). A brief

guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*(1), 13-21. https://doi.org/10.1007/s00265-010-1037-6

Toyama, A., Katahira, K., & Ohira, H. (2019). Biases in estimating the balance between model-free and model-based learning systems due to model misspecification. *Journal of Mathematical Psychology*, *91*, 88-102. https://doi.org/10.1016/j.jmp.2019.03.007

豊田秀樹（編）(2017). 実践ベイズモデリング――解析技法と認知モデル―― 朝倉書店

渡辺澄夫 (2012). ベイズ統計の理論と方法. コロナ社

Watkins, C. J. C. H., & Dayan, P. (1992). Q-Learning. *Machine Learning*, *8*, 279-292. https://doi.org/https://doi.org/10.1007/BF00992698

Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *ELife*, *8*, 1-33. https://doi.org/10.7554/elife.49547

Zhang, S., & Lee, M. D. (2010). Optimal experimental design for a class of bandit problems.*Journal of Mathematical Psychology*, *54*(6), 499-508. https://doi.org/10.1016/j.jmp.2010.08.002

付　録

方法

本シミュレーションでは，各刺激条件においてKL情報量を算出することで，どのような刺激デザインがモデル選択の観点から望ましいかを当たりを付けることを目的とする。一部のパラメータ推定を省略し，かつKL情報量を計算することで，計算負荷を下げることができるため，複数の刺激候補を検討する時に向いている。とりわけ，刺激のパターン数が数百など大きい時には，この計算負荷の低さは顕著になるだろう。

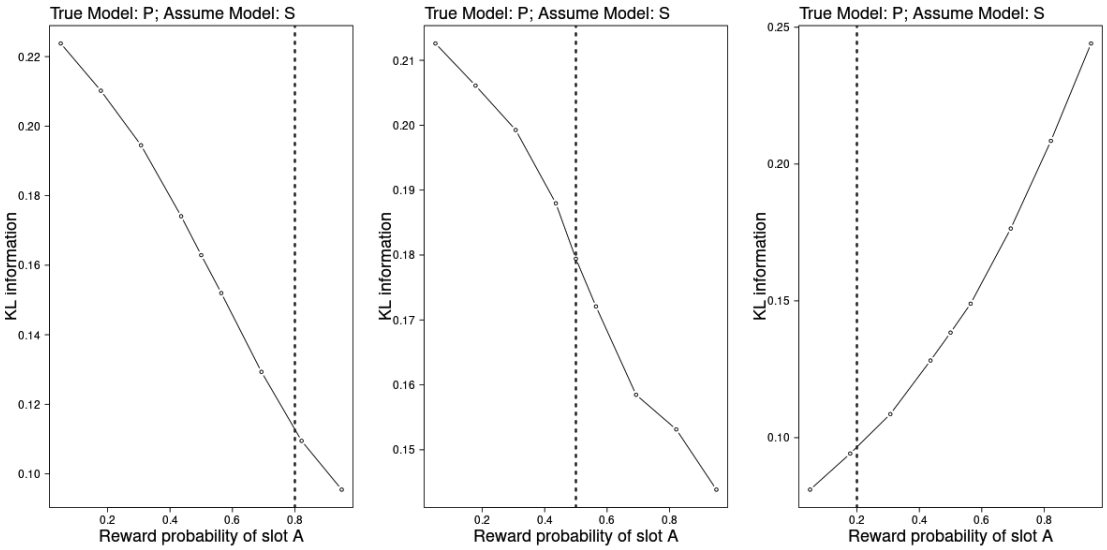
シミュレーションの流れは以下の通りである。1．真のモデルおよびパラメータ真値を固定する。2．t試行目までの刺激を固定する。3．t試行目までのデータ生成を行う。4．仮定するモデルを固定して，パラメータ推定値を求める。5．t+10試行目までの刺激を固定してデータ生成を行う。6．t+10試行目までのKL情報量を計算する。シミュレーションの設定は以下の通りである。真のモデルや仮定するモデル，パラメータ真値の設定は，本文のシミュレーションと同様である。試行数tは100である。100試行目までの実験刺激（報酬確率）は (0.2, 0.8), (0.5, 0.5), (0.8, 0.2) の3種類用意した。t+1試行目からt+10試行目までの実験刺激（報酬確率）は0.05から0.95を7等分する各点の値および0.5の計9種類用意した。

結果

結果例として，真のモデルがPモデルで，仮定したモデルがSモデルにおけるKL情報量の結果を図S1に示す。青の縦線は，t試行目までのスロットAの報酬確率で，横軸の値はt+1からt+10試行目までのスロットAの報酬確率である。特に (0.2, 0.8), (0.8, 0.2) 条件において，報酬確率は現在の状態から離した方がKL情報量は高くなることが分かる。つまり，報酬確率逆転はKL情報量を高め，モデル選択の観点においては望ましい実験デザインと予想される。

（受稿10月12日：受理10月25日）

図 S1. 各刺激条件における KL 情報量



Note. 各報酬確率の値（横軸）における KL 情報量の値（縦軸）を表す。縦の青のダッシュ線は t (100) 試行目までの報酬確率を表す。つまり，左から，t (100) 試行目までの報酬確率は (0.8, 0.2)，(0.5, 0.5)，(0.2, 0.8) である。真のモデルは P モデルで，仮定したモデルは S モデルの結果例である。