

# 遅延を伴う確率的報酬獲得課題におけるヒトの意思 決定に関する研究

○吉田岳彦<sup>(1,2)</sup>, 伊藤 真<sup>(2)</sup>, 森村哲郎<sup>(1,2)</sup>, 鮫島和行<sup>(3)</sup>, 奥田次郎<sup>(4)</sup>, 吉本潤一郎<sup>(2,1)</sup>,  
銅谷賢治<sup>(2,1,5)</sup>

1,奈良先端科学技術大学院大学・2,OIST・  
3,玉川大学脳科学研究所・4,京都産業大学・5,ATR 脳情報研究所

あらまし:

確率に関わる価値の割引と遅延に関わる価値の割引モデルはそれぞれ別々に研究され、同時に扱う実験はされていない。そこで本研究では、異なる遅延と異なる確率が割り当てられた二つの選択肢から一つを選ぶ課題をヒトに対して行った。報酬には金銭などの間接的な報酬ではなくより直接的と考えられる一次報酬(ジュース)を用いた。得られたデータに対して複数の割引モデルをフィッティングした結果、指数関数型の遅延、確率割引モデルが最もよく被験者の行動を表していることが分かった。

## A study of decision-making for delayed and probabilistic reward

Takehiko Yoshida<sup>(1,2)</sup>, Makoto Ito<sup>(2)</sup>, Morimura Tetsuro<sup>(1,2)</sup>, Kazuyuki Samejima<sup>(3)</sup>, Jiro  
Okuda<sup>(4)</sup>, Junichiro Yoshimoto<sup>(2,1)</sup>, Kenji Doya<sup>(2,1,5)</sup>  
1,NAIST・2, OIST・3,Tamagawa university Brain science institute・  
4,Kyoto Sangyo university ・5, ATR

Abstract:

Effects of delay discounting and probabilistic discounting on decision making have been separately studied. In this study, we examined human's discounting function in a choice task where different delays and different probabilities primary reward (juice) were assigned to the options. As a result of behavioral analysis, exponential discounting model of delay or probability shows human's behavior better.

### 1, はじめに

ヒトや動物が生きていく上で、いくつかの選択肢の中から適切な選択肢を選ぶ能力は重要である。選択の結果得られる利得は遅延を伴ったり確率的であったりすることから、選

択行動はそれらの要素に影響を受ける。例えば、遅延のある場合の例として、時間(D1)がかかるとしても欲しい物(R1)を注文して購入するのか、それとも、近くの店でそこそこの物(R2:  $R2 < R1$ )をすぐに(D2:  $D2 < D1$ )手にいれ

るかという状況を考えることが出来る。ここでR1とR2は手に入れたときの報酬量を意味して、報酬量は遅延D1やD2があることによって選択時にはその価値が減少していると解釈することが出来る。この減少まで考慮した価値を報酬価値(V)と呼ぶ。注文して購入する方を選ぶ人は遅延による減少効果が少ないと解釈できるだろう。同様に、報酬が確率的にしか得られない場合にもこのような価値の減衰がおこるとして選択行動を捉えることが出来る[1]。これまでの研究においては、その減衰の仕方がどのような関数で表すことができるかということに焦点が当てられてきたが、遅延と確率による効果は別々に実験されてきた。しかしながら、実際の状況においては遅延と確率を同時に考慮しなくてはならない場合も少なくはない。

そこで本研究では、遅延と確率を同時に考慮しなければならない状況下におけるヒトの選択行動を調べ、どのような関数がよくその行動を表すことができるかを検討した。実験では、金銭などの二次報酬ではなく、より直接的な一次報酬(ジュース)を用いた。

## 2, 遅延もしくは確率的報酬獲得に関する研究

### 2.1, 遅延時間のある報酬獲得

これまでの研究ではラット[5]、鳩[6]に2つのレバーを選択させる実験が行われてきた。1つレバーは早くもらえるが少量の報酬、もう一方は得るために時間がかかるかわりに報酬量が大きいという設定で行われた。その遅れ時間と報酬量との関係を調べた結果、報酬価値は、指数関数型

$$V = R \times \exp(-K \times D) \quad (1)$$

(V:報酬価値, R:報酬量, K:割引率パラメータ, D:遅延時間)

の関数で表されると考えられてきた[6]が、双曲線関数型

$$V = R \times \frac{1}{1 + K \times D} \quad (2)$$

の方がよりデータに合うことが後に示された

[5,6]。ここで、動物は二つの選択肢に対してそれぞれ報酬価値を計算し、より大きい報酬価値を持つ選択肢を選ぶと仮定している。指数関数型モデルは価値割引率はどの遅延時間でも一定であるが、双曲線関数型モデルは、小さい時間遅れの場合には価値割引が大きく、遅延時間が長くなるにつれて次第に割引率が小さくなる特徴がある。

### 2.2, 確率的に報酬が得られる選択実験

同様の結果が報酬獲得に不確実性を伴った研究でも得られている[8]。実験では遅延割引課題同様に二つのレバーが提示され、一つは確実にもらえるが小さな報酬、もう一方には確率は低い大きな報酬が割り当てられた。どちらがどのくらいの確率で報酬が得られるかは動物に十分に学習をさせてから行われた。この場合、直感的には報酬量Rに直接、得られる確率Pを掛けた期待値で報酬価値Vを表せると予想できるが、先行研究ではその他の指数関数型

$$V = R \times \exp(-H \times \theta) \quad (3)$$

(V:報酬価値, R:報酬量, H:割引率パラメータ,  $\theta$ :獲得確率がPの時、の odds against( $\theta = (1-P)/P$ ))

と双曲線型

$$V = R \times \frac{1}{1 + H \times \theta} \quad (4)$$

の方がよりよく報酬価値を表せることが示されてきた[8]。

## 3, 研究の動機

本研究では、報酬獲得の際に遅延と確率的な要素が同時に含まれる場合に、報酬価値がどのような関数で表せるかを検討した。具体的には、先行研究で提案されてきた遅延と報酬モデルを積の形で組み合わせ、どのパターンがもっとも良く報酬価値を表すことができるかを調べた。

## 4, 実験概要

### 4.1, 被験者

本実験では右利きで健康な14人の男女(男性12人、女性2人)左利きの男性3人に対して実験を行った。実験を行うにあたっては玉川大学の倫理、安全委員会の承認のもとに行われた。

#### 4.2, 刺激の設定

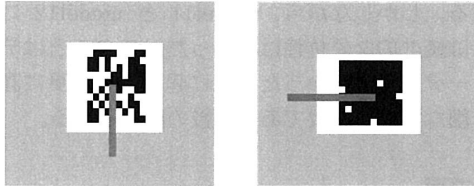


Fig1. 刺激のペア

実験の刺激には Fig1.が代表する組を用意し、被験者に選択してもらった。刺激は背景が灰色で中央にモザイク柄の正方形が出る。その真ん中から濃い灰色の棒が真上を0度に45度角度ずつずれたいずれの向きのもが一本出ている。モザイク柄は報酬が出る確率を表しており、白い部分が多いほど確率が高く、逆に黒い部分が多いほど確率は低い。確率は0%、20%、40%、60%、80%、100%を用意した。そして、棒の向きは待ち時間を表しており、真上の向きが待ち時間無しとし、反時計回りに角度がずれるほど待ち時間は長くなる。45度ずれるごとに2秒の待ち時間が設けてあり、待ち時間は0秒、2秒、4秒、6秒、8秒、10秒、12秒、最大14秒の待ち時間がある。このような情報を持った刺激を1試行当たり2つ順番に出し、被験者の好ましい方を選択してもらおう。刺激は図の左のように確率は高いが、待ち時間が長いものと、右の図が表す確率は低い待ち時間が短いものを提示している。

刺激提示の流れは Fig2.のようになっており、

- 1、青い十字[ I T I (inter trial interval) ] を表示 (3000-10000 ミリ秒)
- 2、一枚目の刺激 (500 ミリ秒)
- 3、灰色の十字 (3000-10000 秒)
- 4、2枚目の刺激 (500 ミリ秒)
- 5、灰色の十字 (3000-10000 ミリ秒)

- 6、選択期間 (2000 ミリ秒)
- 7、選択した方を被験者にフィードバック (500 ミリ秒)
- 8、選んだ図柄を表示 (被験者が選んだ図柄分だけ表示し続ける)
- 9、報酬獲得表示。白ければ獲得、黒ければ逃したことになる (500 ミリ秒)
- 10、報酬 (ジュース) が出てくる (1000 ミリ秒)
- 11、赤い十字。ジュースを飲む (2000 ミリ秒)

をそれぞれ表している。この11個の流れが1試行になり、これを1セッションにつき34試行行い、全2セッション行った。

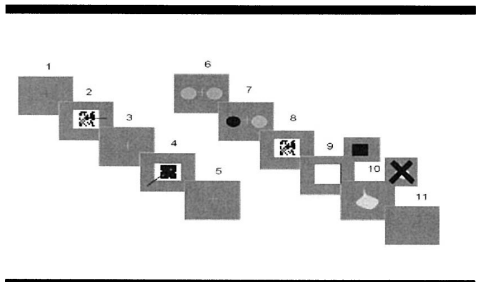


Fig2. 1試行の流れ

## 5, 実験結果

### 5.1 バイアスのある選択に関する検定

被験者の選択に何らかの偏りがいないかを二項検定を用いて調べた。その結果17人中1人のみ左の選択に重きがおかれていた ( $p < 0.05$ )、以後の解析にはそれを除いた16人の被験者データを用いた。

### 5.2 モデルフィッティング

モデルには遅延割引に関しては指数関数型、双曲線型を用いた。そして確率割引に関しては単純な期待値、指数関数型、双曲線型を用いた。選択行動は価値の比重から確率的に行っているものとし、ソフトマックスを用いた。

$$P_1 = \frac{\exp(\beta V_1)}{\exp(\beta V_1) + \exp(\beta V_2)} \quad (5)$$

( $V_1$ : 選択された選択肢の価値,  $V_2$ : 選択されなかった選択肢の価値,  $\beta$ : パラメータ  $P_1$ :

$V_1$  を選ぶ確率)

ここで  $P_1$  は被験者が選択肢 1 を選ぶ確率を表す。モデルが選択データをどれだけ良く表しているかの指標としては、赤池情報量基準 (AIC) を用いた。

$$2 \times (-\sum \log(P_i)) + 2 \times m \quad (6)$$

( $m$ : モデルのフリーパラメータの数)

その結果は Fig3. に表した。

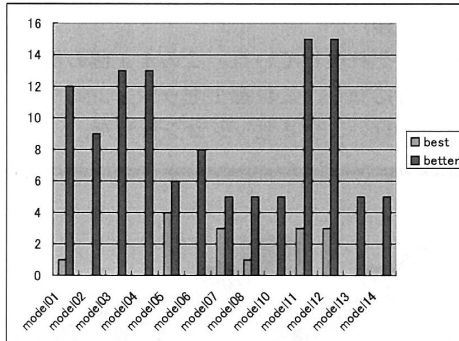


Fig3. モデルとその行動データフィット数  
灰色の棒グラフはそのモデルのAICの値が良かった行動被験者数、黒い棒グラフはランダム選択よりAICの値が良かった被験者数を表している。

本実験から行動データをより良く説明するモデルとしては

model11

$$V = R \times \exp(-H \times \theta) \times \exp(-K \times D) \quad (7)$$

model12

$$V = R \times \exp(-H \times \theta) \times \frac{1}{1 + K \times D} \quad (8)$$

(7)(8)の両モデルが被験者毎の最もAICの値の良いモデルとして、又ランダム選択よりもAICの値で勝っているという数で他のモデルより勝っていた。

## 6, まとめ

遅延の割引率が確率に依存したモデルに関しても試したが、(7)(8)モデルに対してデータを説明できるものではなく、Hoの予測していたモデル

$$V = R \times \frac{1}{1 + H \times \theta} \times \frac{1}{1 + K \times D} \quad (9)$$

よりも確率、遅延混合環境では指数関数型の減衰モデルが組み合わさったモデルの方が指示された。これは確率だけの場合などはリスク回避的な行動をとるのであるが、より利益に結びつきやすい選択をしたからだと思われる。しかしながら、model11とmodel12には統計的な優位性はなかった。このことはデータ点が少なかったことに起因する。更に詳細な実験が必要であり今後の課題とする。

## 文献

- [1] Sozou, P.D. (1998), 'On hyperbolic discounting and uncertain hazard rates', Proceedings of the Royal Society of London. Series B: Biological Sciences 265, 2015-2020.
- [2] 松島, 俊. & 青木, 直.(2005)ヒヨコの経済学: 採餌効率に基づく選択 (特集 動物の認知) 認知科学, 日本認知科学会, 12, 177-187
- [3] Tanaka, S.; Doya, K.; Okada, G.; Ueda, K.; Okamoto, Y. & Yamawaki, S. (2004), 'Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops', Nature Neuroscience 7(8), 887-893.
- [4] Ho, M.; Mobini, S.; Chiang, T.; Bradshaw, C. & Szabadi, E. (1999), 'Theory and method in the quantitative analysis of impulsive choice' behaviour: implications for psychopharmacology', Psychopharmacology 146(4), 362-372.
- [5] Roesch, M.; Taylor, A. & Schoenbaum, G. (2006), 'Encoding of Time-Discounted Rewards in Orbitofrontal Cortex Is Independent of Value Representation', Neuron 51(4), 509-520.
- [6] Mazur, J. (1985), 'Probability and delay of reinforcement as factors in discrete-trial choice.', J Exp Anal Behav 43(3), 341-51.
- [7] Samuelson, P. (1937), 'A Note on Measurement of Utility', The Review of Economic Studies 4(2), 155-161.
- [8] Rachlin, H.; Raineri, A. & Cross, D. (1991), 'Subjective probability and delay.', Journal of the Experimental Analysis of Behavior 55(2), 233.