# 逐次的意思決定における探索と知識利用のジレンマへの対処方法
# A Behavior of Human for the Exploration-Exploitation Dilemma in Decision Making

並木 尚也[†]，大用 庫智[†]，高橋 達二[‡]
Naoya Namiki, Kuratomo Oyo, Tatsuji Takahashi

[†]東京電機大学大学院，[‡]東京電機大学理工学部
School of Science and Engineering, Tokyo Denki University
14rmd17@ms.dendai.ac.jp

**Abstract**

In an uncertain environment, decision-making is entailed two opposing demands. One of these demands is gathering new information, another is exploiting already known information. These opposing demands are called the exploration-exploitation dilemma. In brain science, it's known that human's brain estimates options relatively and correlates to SoftMax that is policy premised on choosing options randomly. The other side, in cognitive science, it is indicated that human is not cognizant of random sequence correctly. There is one contradiction. Although it's difficult for human to be cognizant of random sequence correctly, does human really choose options randomly? Both research results pose a conflict. In this study, we analyzed how human behave for the exploration-exploitation dilemma through experiments that is N-armed bandit problem and comparing with some policies commonly used in reinforcement learning, from a viewpoint of whether human really choose options randomly.

**Keywords — Bandit problem, SoftMax**

## 1. はじめに

不確実な環境下における意思決定は，多数の選択肢から良い選択肢を探し出す「探索」と，既知の情報・経験を活用し最良の選択肢を選択し続ける「知識利用」という 2 つの相反する行動が要求される. これを探索と知識利用のジレンマと呼ぶ. 得られる利益を最大にするという目的を達成するためには，このジレンマは無視できない厄介な要素である．収益を最大化するためには，最良の選択肢を見きわめ，選択し続ける必要がある（知識利用）．しかしながら，不確実な環境下では，どの選択肢が有益なのかが未知である．そのため，選択肢を一つ一つ試しながら検証して，その価値を見きわめる必要がある（探索）．このジレンマを表現した強化学習の基本的な課題である N 本腕バンディット問題[1]があり，この問題に対するさまざまなモデルが提案されている．

脳科学の分野では人間の脳が選択肢を相対的に評価していることが知られている．また SoftMax 法というコンピュータのようにランダム性を仮定している方策と相関があることが明らかになっている[2]．一方で，認知科学の分野では，人間のランダム系列に対する認識の困難さが指摘されている[3]．たとえば，コイン投げにおいて，各試行は独立にもかかわらず，数回連続で表が出た際に次は裏が出るであろうと予想してしまうギャンブラーの誤謬などが挙げられる．このように人間はランダム系列をコンピュータのように正しく認知することが困難であり，ランダム系列に対して何らかの規則性（e.g. 少数の法則）を見出してしまう傾向がある．

ここで一つの矛盾が生じる．人間はランダム系列の認識が困難であるにも関わらず，果たして本当に確率的，つまりランダムに選択を行うことができるのだろうか. 両者の結果は一貫していない. また, 前述した SoftMax 法はあくまで複数の人間の平均データと相関があったのであり，個々の振る舞いに対して一致しているかどうかは定かではない.

本研究では，探索と知識利用のジレンマに対して人間がどのような振る舞いをするのか，人間が本当に確率的に選択を行うかどうかという視点から，強化学習のタスクであるバンディット問題を通して実験を行い，さまざまな方策と比較しながら分析した.

## 2. 探索と知識利用のジレンマ

ここでは，探索と知識利用のバランスが重要であることと，このジレンマが人間にとってどのようなものなのかを述べる．

不確実な環境下での逐次的な意思決定課題において，探索と知識利用の行動のバランスをとることは重要であり，どちらかの行動に偏ることは目的の達成から遠ざかることになる．知識利用を重視すると，最良の選択肢を見誤る可能性があり，結果的に目的の達成から遠ざかってしまう．探索を重視すると，利益の回収が遅れてしまい，制限のある環境下（たとえば時間，資金など），あるいはその制限が不透明な環境では利益の回収が不十分になり，こちらもまた目的の達成から遠ざかってしまう．現実では無制限に試行できる環境はめったになく，さまざまな要素によって制限されるだろう．そのため，目的を達成するためには，探索と知識利用のバランスをうまく保つ必要がある．

探索と知識利用のジレンマは人間の経験的学習・意思決定の性質と深く関わっている．このジレンマに対して人間がどのようにうまく対処しているのかを解明することは，人間の経験的学習・意思決定の性質を理解することにつながると考えられる．また，その性質を応用することによって，人工知能やロボットなどが未知の環境において自律的に学習する事を可能にするかもしれない．そのような意味で，探索と知識利用のジレンマに対する人間の振る舞いを研究することは意義のある事であると考える．

## 3. N本腕バンディット問題

N 本腕バンディット問題とは，強化学習のもっとも基本的な課題の一つであり，前述した探索と知識利用のジレンマを最も単純に表現する課題である．具体例として，スロットマシンを挙げて説明する．任意の N 台のスロットマシンが存在し，それぞれに異なる当たり確率が設定されており，その当たり確率に従って報酬を返す．プレイヤーは得られる報酬を最大化する事を目的とする．このときプレイヤーは各腕の当たり確率を知らず，1度に1つの腕を選択する．目的を達成するために，プレイヤーは各腕の中で最良の腕を探す事（探索）と，最良と思われる腕を引き続ける事(知識利用)を要求される．このように，バンディット問題は探索と知識利用の 2 つの要素を含んでおり，単純に表現している．バンディット問題とは，このように N 個の選択肢の中から逐次的に選択し，報酬を最大化するという目的のある形態をとる問題の事である．本研究では，探索と知識利用のジレンマに対する人間の振る舞いを観測する事に都合が良いため，実験のタスクとして用いた．また，より単純な枠組みで行うために 2 つの選択肢，即ち，2 本腕バンディット問題と呼ばれる形式で行った．

## 4. 人間の探索と知識利用のジレンマの扱い方

探索と知識利用のジレンマは，強化学習の中で中心的なトピックとして研究されてきた．近年，強化学習のタスクを通して，探索と知識利用のジレンマは脳科学でも研究され初めて来た[2]．その中でも，fMRI を用いたバンディット問題をプレイ中の参加者の脳の観測により，探索と知識利用のジレンマや学習等の人間の脳内での扱われ方が，だんだんと解明されつつある．ここで，我々は探索と知識利用のジレンマと脳科学，そして，バンディット問題と関係が深い論文を二つ紹介する．

Daw らは 4 本腕バンディット問題をプレイ中の人間の参加者の脳活動の観測によって，探索に関連する神経基質の関係と探索と収穫の切り替えの形式的な問題を調査した．その結果，彼らは前頭前野腹内側部（ventral medial prefrontal cortex：vmPFC）が相対的な報酬の大きさをコード化する事と探索時に前頭極が活性化する事を示した．Daw らは初めて，探索と神経基質の関係を明らかにし，探索と知識利用のモードの間の行動戦略のスイッチングを容易にするための管理機構を映す事を可能にした．Boorman らは，2 本腕バンディット問題をプレイ中の人間の参加者の脳活動の観

測によって，主に二つの脳領域の活性化と探索と知識利用のジレンマの関係を調査した. その結果，彼らは前頭前野腹内側部が選択された腕の相対的な価値をコード化することを示した．また，前頭極が選択されていない腕の相対的な報酬確率をコード化することを示した．彼らは，不確実な環境に対処可能な人間の行動の柔軟性に関して，前頭葉における計算の重要性を示した．ただし，これらの 2 つのバンディット問題のタスクは非定常であった.

以上から，不確実な環境で発生する探索と知識利用のジレンマに対処するために，人間は絶対的評価よりも相対的な評価を行っていることが分かる．その証拠に，バンディット問題をプレイ中の人間の振る舞いが相対評価を行なう SoftMax 法で最も特徴づけられている[2]しかし, SoftMax 法の様な評価・選択は人間には難しいと考えられる（ランダム系列を正しく認知出来ない[3]）．また，実際に行動としてどのように表れるかは具体的に明らかになっていない.

## 5. ランダム系列の誤認知

ランダム系列の誤認知とは，ランダムな系列を人間が正しく認知することが困難であり，その系列に対してなんらかの規則性を見いだしたりしてしまうようなことである．具体例としてコイン投げを挙げる. 連続するコイン投げにおいて,「表表表」と出ると「次もきっと表が出るであろう」と考えてしまう．このような思考の傾向はホットハンドと呼ばれている．また，逆に「表表表」と出ると､「そろそろ次には裏が出るだろう」と考えてしまうこともある．このような思考の傾向はギャンブラーの誤謬と呼ばれている．紹介した 2 つの思考の傾向のどちらも，考え方としては正しくない．なぜならば、コイン投げにおいて表が出る確率も裏が出る確率も等確率であり，独立であるため，各試行間に関わりはなく，結果の確率に影響しないのである.

他には,「少数の法則」,「代表性ヒューリスティック」などがある.「少数の法則」とは，ある母集団から抽出された標本にも，その母集団の本質的な特徴が現れていると人間が認識する傾向である. コイン投げで例えれば，無限に続く長い系列ではなく，ごく短い系列においても表と裏が等しく出ていると考えることである（極端に言えば、系列の長さが 2 であるとき,「表裏」もしくは「裏表」となるであろうと考える）.「代表性ヒューリスティック」とは, ある母集団から抽出された標本が，どの程度母集団を近似（代表的）しているかどうかの判断におけるバイアスである．コイン投げでいえば,「表表表表表裏表」「裏表裏表裏表裏」という 2 つの系列が存在するとき，人間は前者の事象よりも後者の方の数が多い，つまり代表性が高いと考える傾向があるということである.

## 6. 実験 1

### 6.1 実験設定

本研究では, 39 名の参加者が 2 本腕バンディット問題にコンピュータ上で取り組んだ．取り組むタスクを 2 つのスロットの当たり確率差が大きい問題（以下，簡単な問題）と小さい問題（以下，難しい問題）を用意した．簡単な問題では，2 つのスロットの当たり確率をそれぞれ（0.8, 0.2）とし，難しい問題では 2 つの腕の当たり確率をそれぞれ（0.6 0.2）とした．参加者の可能な試行回数は，簡単な問題は 20 回，難しい問題は 40 回にそれぞれ設定した．またスロットが出力する報酬は「当たり」か「はずれ」の 2 通りのみにした．さらに，参加者を次のような 2 群に分けた．一つの群は，最初に簡単な問題を行った後に難しい問題を行う ED 群（Easy→Difficult），もう一つはその逆の順序で行った DE 群（Difficult→Easy）である．参加者の人数はそれぞれ ED 群は 17 人，DE 群は 22 人となっている.

また，人間の直観性をより重視するために，どれだけ試行できるか，どの腕が今までどれだけ当たったか，あるいは外れたかなどの情報はすべて参加者には分からないようにした. 先行研究では，これらの情報が可視化されている場合がほとんど

である（e. g. Zhang and Yu 2013）また，人間のデータと強化学習でよく用いられている方策（Greedy 法，ε-greedy 系，SoftMax 法）を比較した．

## 6.2 人間と比較する方策

人間のデータと比較する方策をここで紹介する．また，スロットの評価値は客観的な条件付確率によって算出される．

（1） Greedy 法

この方策は，選択肢それぞれの評価値に基づいて，常に一番評価値が高い選択肢を選択する方策である．Greedy というのは貪欲という意味である．

（2）ε-greedy 法

このモデルは，探索と知識利用の行動を明確に分離する方策である．具体的にはパラメータ ε（0.0～1.0 の間をとる）の確率でランダムに選択肢の選択をし，1-ε の確率で greedy に選択を行う．ε-greedy 法にはいくつか種類があり，今回はその中の 3 つの方策を比較対象として使用した．

- 序盤探索法（Epsilon First）

序盤探索法は，定められた挑戦可能な試行回数の ε の割合だけ完全にランダムに選択を行う方策である．

- ε－一定法（Epsilon Constant）

ε－一定法は，最初の試行から最後の試行まで ε の確率が変化しない方策である．

- ε－減衰法（Epsilon Decreasing）

ε－減衰法は，試行回数を重ねるごとに徐々に ε の確率が減衰してゆく方策である．本研究で用いた減衰式を以下に示す． $\tau$ は減衰のスピードのパラメータ，$t$ はその時点までの試行回数である．

$$\varepsilon = \frac{1.0}{1.0 + \tau * t} \tag{1}$$

（3）SoftMax 法

SoftMax 法は，条件付き確率によって算出されたスロットマシンそれぞれの評価値を選択確率に重みづけし，選択を確率的に行うモデルである．探索と知識利用の行動をバランシングする方策である．本研究では，SoftMax 法を拡張した Modefied SoftMax Algorithm を使用した[5]．以下に式を示す．$P(X)$ はある選択肢の選択確率，$M(1|X)$はある選択肢 $X$ に対する評価， $\tau$ は減衰率，$t$ は現在までの試行回数である．

$$P(X) = \frac{exp\,(M(1|X)\times\tau t)}{\Sigma_{x'\in\{A,B\}} exp\,(M(1|X')\times\tau t} \tag{2}$$

## 6.3 実験結果

人間が確率的に選択を行っているかどうかを調べるために，「Win-Shift」という指標を用いる．Win-Shift とは，ある腕を選択し，当たったにも関わらず次の試行では違う腕を選択する確率である．単純に違う腕を選択する確率（Shift）や，外れて違う腕を選択した確率（Lose-Shift）ではランダムに選択したことかどうかは判断できない．特に後者は違う腕を選択することの理由に「外れたから」という事が考えられる．しかしながら Win-Shift が起こる理由はランダムに選択を行う事以外にはないと考えられる（「当たった」から違う選択肢を選択する事は目的を考慮すると，理由として考えづらい）．従って，Win-Shift は確率的に選択しているかどうかのみを確認できる指標であると考える．そして各個人の Win-Shift のデータ傾向をもとに，分類を行った．図 1 に前述した方策の Win-shift の図を示す．また，図 2，3 に ED 群における Win-Shift の分類, 表 1, 2 に図 2, 3 に対応したそれぞれのタイプとモデルごとの正解率とそのタイプの割合を示す．図 1～3 は縦軸が Win-shift が起きる確率，横軸が試行回数を表している．Win-Shift を各個人のデータで分類した理由は，平均化する事によりデータがつぶれ性質が見えなくなるためである．また，Win-Shift が発生したステップの期間によって分類を行っている．正解率とは 1 回目の試行から最後の試行までの, 当たり確率の高い腕を選択した割合である．図 1 は前述した方策のコンピュータ上のシミュレ

ーションの結果である．greedy 法は Win-Shift が起こらない事がわかる． ε 一定法， ε 減衰法，序盤探索法は ε -greedy 系の方策である． ε 一定法は一定の確率で， ε 減衰法は減衰しながら，序盤探索法はある試行回数まで 0.5 の確率でそれぞれ Win-Shift が起こる事がわかる．SoftMax 法は ε 減衰法のような傾向で Win-Shift が起こる事がわかる．即ち，greedy 法以外の確率的に選択を行う方策は Win-Shift が起こる．また図 1 からわかるように，方策によって Win-Shift の傾向が異なっている事がわかる． 特に顕著に異なるのが，Win-Shift が起こるステップである．人間もまた同様に，個々で異なる方策を持っている可能性も考えられるため，方策の特徴を決定づける「Win-Shift がどのステップで起きたか」ということを基準に人間のデータを分類した今回は単純に Win-Shift が起きたステップを前期，中期，後期の３つに分割し，その組み合わせによって分類した．Win-Shift が出現したことを「S」とし，出現しないことを「N」とし，前期・中期・後期（前期／中期／後期）のそれぞれどこに Win-Shift が出現したかで分類し，「NNN 型」，「SNN 型」，「NSN 型」，「NNS 型」，「SSN 型」，「SNS 型」，「NSS 型」，「SSS 型」の８通りに分類する．

まず，ED 群について見る．表１より，簡単な問題における ED 群では, 最も多いタイプは NNS 型と NNN 型であった．表２より，難しい問題における ED 群では，最も多いタイプは NNN 型であった．

次に，DE 群について見る．表３より，簡単な問題における DE 群では, 最も多いタイプは NNN 型であった．表 4 より，難しい問題における DE 群では，最も多いタイプは SSS 型であった．

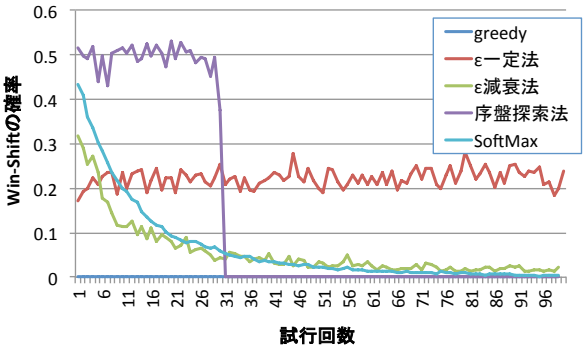


図 1.　各方策の Win-Shift

表１．簡単な問題における ED 群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 89 | 35 |
| SNN | 93 | 12 |
| NNS | 80 | 6 |
| SSN | 73 | 35 |
| SNS | 60 | 12 |
| Greedy 法 | 93 | |
| 序盤探索法 | 83 | |
| ε 一定法 | 93 | |
| ε 減衰法 | 84 | |
| SoftMax 法 | 77 | |

表２．難しい問題における ED 群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 63 | 35 |
| SNN | 79 | 23 |
| NSN | 74 | 12 |
| SSN | 34 | 12 |
| SNS | 55 | 6 |
| NSS | 63 | 6 |
| SSS | 48 | 6 |
| Greedy 法 | 72 | |
| 序盤探索法 | 72 | |
| ε 一定法 | 69 | |
| ε 減衰法 | 73 | |
| SoftMax 法 | 69 | |

表3．簡単な問題におけるDE群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 89 | 45 |
| SNN | 80 | 14 |
| NSN | 58 | 9 |
| SSN | 72 | 14 |
| SNS | 58 | 9 |
| NSS | 70 | 5 |
| SSS | 50 | 5 |

表4．難しい問題におけるDE群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 64 | 27 |
| SNN | 63 | 14 |
| NNS | 86 | 9 |
| SNS | 60 | 14 |
| NSS | 88 | 5 |
| SSS | 56 | 32 |

## 7. 実験２

### 7.1 実験設定

実験１とは異なる条件で，25名の参加者が2本腕バンディット問題にコンピュータ上で取り組んだ．取り組むタスクを2つのスロットの当たり確率が双方とも高い（以下，高確率環境）と双方とも低い（以下，低確率環境）を用意した．またスロットが出力する報酬は実験１と同様に「当たり」か「はずれ」の2通りのみにした．さらに，参加者を次のような2群に分けた．一つの群は，最初に高確率環境を行った後に低確率環境を行うHL群（High→Low），もう一つはその逆の順序で行ったLH群（Low→High）である．参加者の人数はそれぞれHL群が11人，LH群は14人となっている．高確率環境では，２つのスロットの当たり確率をそれぞれ（0.7， 0.8）とし，低確率環境では，２つのスロットの当たり確率をそれぞれ（0.3, 0.2）と設定した．参加者の試行可能な回数はどちらの環境とも 50 回とした．実験１と異なる点はスロットマシンの確率設定と参加者の試行可能な回数のみであり，その他の点は実験１と同様である．

### 7.2 実験結果

実験１と同様に，「Win-Shiftがどのステップで起きたか」ということを基準に人間のデータを分類した．

まず，HL群について見る．表5より，高確率環境におけるHL群では，最も多いタイプはSSS型であった．表6より，低確率環境における HL群では，最も多いタイプはSSS型であった．

次に，LH群について見る．表7より，高確率環境におけるLH群では，最も多いタイプはNNN型であった．表8より，低確率環境における LH群では，最も多いタイプはNNN型であった．

表5．高確率環境におけるHL群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 73 | 36 |
| SNN | 58 | 9 |
| SSS | 43 | 55 |

表6．低確率環境におけるHL群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 79 | 18 |
| NSN | 46 | 18 |
| NNS | 36 | 9 |
| SNS | 58 | 9 |
| SSS | 51 | 45 |

表７．高確率環境における LH 群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 91 | 50 |
| SNN | 78 | 7 |
| SSN | 74 | 7 |
| SSS | 51 | 36 |

表８．低確率環境における LH 群の正解率とタイプの割合

| タイプ/モデル | 正解率(%) | タイプの割合(%) |
|---|---|---|
| NNN | 56 | 36 |
| SNN | 45 | 14 |
| NSN | 56 | 7 |
| NNS | 10 | 7 |
| SSN | 42 | 7 |
| SNS | 35 | 21 |
| SSS | 60 | 7 |

## 8. 総合議論

実験１において，難しい問題における DE 群以外の状況では，人間において最も多いタイプは Win-Shift が見られないタイプであった．難しい問題における DE 群に関しても２番目に多いタイプが Win-Shift が見られないタイプであった．特に，簡単な問題における DE 群では約半数近い人間に Win-Shift が出現しなかった．従って，一般的に人間が確率的に選択を行っていないということが考えられる．

実験２においては，HL 群では Win-Shift が出現しないタイプよりも，すべての期間で Win-Shift が出現するタイプが高確率環境と低確率環境の双方で最も多く割合を占めていた．Win-Shift が出現しないタイプは，２番目に多かった．しかしながら LH 群では，高確率環境と低確率環境の双方で最も多く割合を占めていたのが，Win-Shift が出現しないタイプであった．特に高確率環境において顕著であり，半数がそのタイプであった．このことから，すべての環境ではないが，確率的に選択を行っていない人間は多数存在することが確認できる．また，２つの群両方ともが高確率環境よりも低確率環境における Win-Shift がないタイプの割合が低いことがわかる．これは低確率環境であると，必然的に当たる回数よりも負ける回数が増えることが予想され，負けた回数の多さから，当たったことに対する確信が揺らぎ，たとえ当たったとしてもその結果に疑念を持ち，Win-Shift を引き起こすことが考えられるかもしれない（この場合の Win-Shift は確率的に選択をするという意味とは異なるかもしれない）．

２つの実験から，多くの環境で，一般的に人間が確率的に選択をしないことが考えられる．また良い情報（Win）が与えられたときには選択肢を切り替えないという事も考えられる．したがって，悪い情報（Lose）が与えられたときのみに選択肢を切り替える可能性もある．

どの実験環境でも後期に Win-Shift が見られるタイプが存在した．このように試行の後期に Win-Shift が出現することは，前述した強化学習で一般的に用いられる方策ではあまり無い．そのため，この後期に Win-Shift が出現する性質は人間特有のものであると捉えられるかもしれない．理由として，飽きや疲れなどの感情的な要因が考えられる（特に本実験では，参加者には試行可能な回数が未知であり，負担が大きく，疲れを誘発した可能性も考えられる）．このような性質は，非定常な環境，つまり，試行の途中でスロットマシンの確率が変動するような状況で有効にはたらくと考えられる．さらに，後期に Win-Shift が出現する理由が感情的な要因であるとすれば，定常・非定常かどうかを疑うという意識を持たず，無意識に非定常環境に適応できるということになる．これは，人間が環境に素早く適応可能である特徴を有するということが言える可能性がある．そうであれば，その特徴の意味をより詳細に解明し，モデル化し応用することは有用であると考えられる．

## 7. 結論

本研究では，探索と知識利用のジレンマに対する人間の振る舞いの性質・傾向を調査した．その結果，探索と知識利用の行動を明確に分ける方策や，人間と相関があるといわれている SoftMax 法などの方策とは違う傾向があることが確認できた．確率的に選択が行われないのである．また，探索行動，選択肢の切り替えは一般的には負の情報がもたらされた時のみ起こることも確認できた．さらに，後期に Win-Shift が起きるという人間特有の傾向があることが確認できた．これらの結果は，現在研究されている人間の認知的な特性を利用するモデルに対して，より詳細な形式化を可能にすると考えられる．

## 参考文献

[1] Sutton, R． S., Barto, A. G., 1998. Reinforcement Learning: An Introduction. *MIT Press*, Cambridge, MA.Sidman, M. (1994). Equivalence relations and behavior: A research story. Boston, M.A.: Authors Cooperative.

[2] Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., Dolan, R. J., 2006. Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876–879, 2006.

[3] Tversky, A., Kahneman, D., Judgement under Uncertainty: Heuristics and Biases, Science, 185(4157), 124-1131, 1974.

[4] Zhang, S., Yu, A.J. (2013). Cheap but Clever: Human Active Learning in a Bandit Setting. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

[5] Oyo,K., Takahashi,T. A cognitively inspired heuristic for two-armed bandit problems: The loosely symmetric (LS) model. *Procedia Computer Science* 24 (2013) 194-204, 2013.