

速さと正確さのトレードオフにおける適度な思考時間の存在: 資源合理的分析

Optimal time in speed-accuracy trade-off: Resource-rational analyses

白砂 大[†], 香川 璃奈[‡], 本田 秀仁[†]

Masaru Shirasuna, Rina Kagawa, Hidehito Honda

[†] 追手門学院大学, [‡] 産業技術総合研究所

Otemon Gakuin University, National Institute of Advanced Industrial Science and Technology

m.shirasuna1392@gmail.com

概要

速さと正確さのトレードオフの文脈では、時間をかけることでよい結果(正確な判断)を得られると考えられている。しかし従来、人の限られた認知資源、特に考えること自体に対するコストが考慮されていなかった。本研究では、資源合理性の枠組みに基づいて、思考コストを抑えつつ高い正確さを維持できるような適度な思考時間が存在するという仮説を立てた。そのうえで、理論(計算機シミュレーション)と実験(行動実験)の双方から、この仮説を検証した。

キーワード: 速さと正確さのトレードオフ (speed-accuracy trade-off), 資源合理性 (resource rationality); 計算機シミュレーション (computer simulation); 行動実験 (behavioral experiment)

1. はじめに

判断の正確性と反応時間との関係について、心理学・認知科学では「速さと正確さのトレードオフ」(speed-accuracy trade-off, 以下「SAT」; e.g., Heitz, 2014; Wickelgren, 1977)が古くから知られている。これは、素早い判断は正確さが低くなりやすく、正確な判断を行うには長い時間をかける必要があるという、思考時間と正答率とのトレードオフ関係のことである。

しかし、従来の SAT の文脈においては、認知資源(e.g., 計算能力など)という観点が考慮されていなかった。人の認知資源には限りがあるため、考えれば考えるほどよい結果をもたらすとは限らず、むしろ考えることそれ自体に認知的な負荷(以下「思考コスト」)がかかると考えられる。例えば、考えすぎることによって、かえって判断に迷って認知コンフリクトが生じたり、メンタルワークロードが増大したりするかもしれない。近年の認知科学では、人は思考コストを最小限に抑える形で合理的な判断を行っているとする「資源合理性」という枠組みが提唱されている(e.g., Griffiths, et al., 2015)。実際に、時間に伴う思考コストを仮定した計算論モデルの方が、それを仮定しないモデルよりも人のふるまいをよく説明することが、理論(計算機シミュレーション)と

実験(行動実験)の双方から数多く示されている(e.g., Callaway et al., 2023; Ho et al., 2023)。人の認知資源を考慮しない場合、長く考えることがよい結果をもたらすと予測される(i.e., SAT)。一方、認知資源を考慮する場合、長く考えすぎると思考コストが増加するため、必ずしもよい結果にはならないだろう。すなわち、正確性を高めつつ思考コストが最小限になるような、適度な思考時間が存在すると予測される(i.e., 資源合理性)。

SAT の文脈において、従来議論されてこなかった認知資源という側面を考慮することは、正確さと思考コストとのトレードオフという新たな視点から人の認知・行動をより深く理解することにつながるだろう。本研究では、計算機シミュレーションと行動実験を通して上記の予測を検証することを目的とした。

2. 実験 1: 計算機シミュレーション

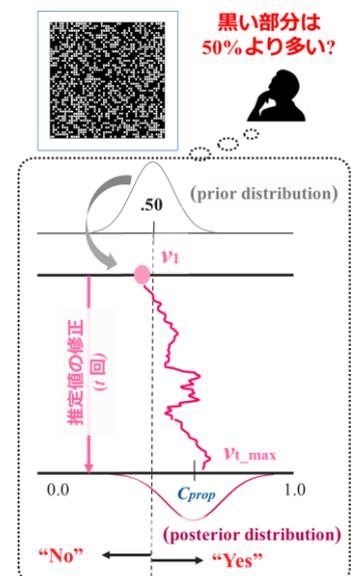
SAT の文脈において思考コストを考慮すると、適度な(短すぎず、長すぎない)思考時間が存在するだろう。実験 1 ではまず、この点を計算機シミュレーションによって検証した。

図 1 グリッド課題の計算機シミュレーション(概略図)

2.1. 方法

題材: SAT は主に知覚判断の場面で議論されることから(e.g., Heitz, 2014), 本研究でも単純な知覚判断課題を題材とした。具体的には、白黒の格子画像が呈示され、黒の面積が全体の 50%より多いか否かを判断する課題を想定した(図 1)。

手続き: 本シミュレ



シオンは、数値推定課題を資源合理性の枠組みで検証した Lieder et al. (2018)に基づいて行われた。大筋として、人が黒グリッドの割合を推定するにあたり、 t 時点における推定値 v_t がメトロポリス・ヘイスティングス法(Hastings, 1970)に基づいて修正されていくものとした。まず、本題材は「50%より多いか否か」を判断する課題であるため、 $N(0.50, \sigma^2)$ に従って初期推定値 v_1 を生成した。そして、正答値 C_{prop} および推定者の信念 b のもとで、最終的な推定値(事後分布)が $N(C_{prop}, b^2)$ に従うことを仮定した。具体的には、 $Prob(X = v_{t-1} + \delta) > Prob(X = v_{t-1})$ であれば推定値は必ず修正され(i.e., $v_t = v_{t-1} + \delta$)、 $Prob(X = v_{t-1} + \delta) \leq Prob(X = v_{t-1})$ であっても $Prob(X = v_{t-1} + \delta) / Prob(X = v_{t-1})$ の確率で修正が行われ、それ以外の場合は修正が行われないものとした(i.e., $v_t = v_{t-1}$)。なお δ は、推定値がどの程度修正されるかを表現するパラメータであり、 $N(0, 0.05^2)$ に従うと仮定した。そして、最終的な推定値 $v_{t_{max}}$ が.50を上回っていれば正答、下回っていれば誤答と見なした。

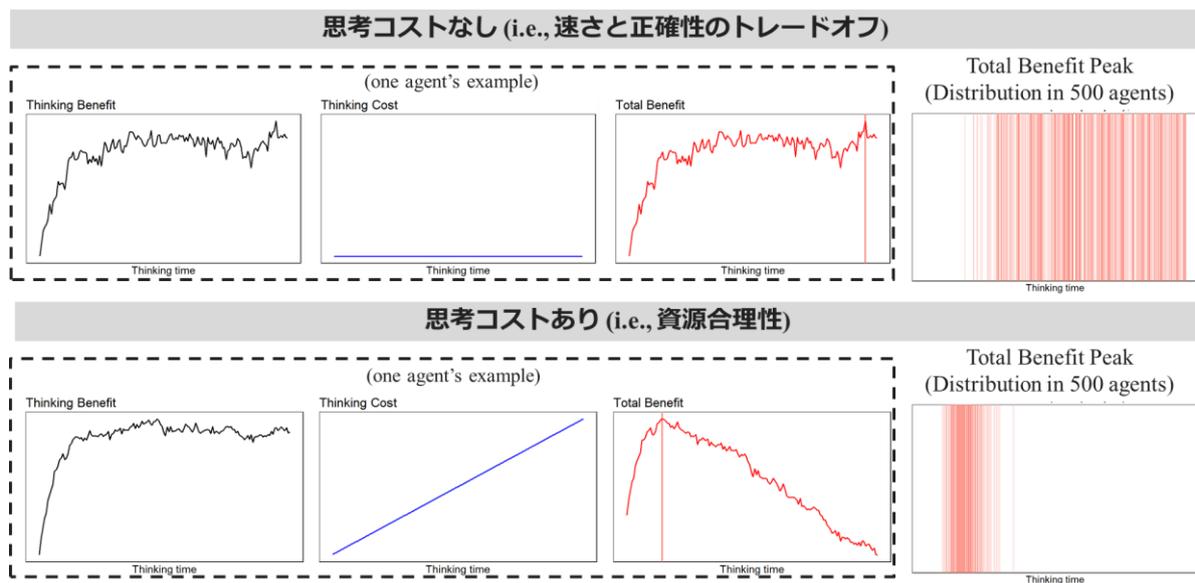
パラメータ設定: シミュレーションでは、正答値 C_{prop} を「.55」(i.e., 黒い部分の面積が55%)とし、初期値の標準偏差 σ を「0.1」、推定者の信念 b を「0.1」に設定した。また、 t の最大値 t_{max} は「150」(i.e., v_1 から v_{150} まで150回の修正が行われる)とした。

分析指標: 一般に、人の推定値は一定にはならないことが考えられる(e.g., 黒の割合が同じでも、グリッドの配置パターンによって推定値は異なりうる)。そこで本シミュレーションでは、 v_1 から v_{150} までの推定手続きを1,000回繰り返した。そのうえで、各時点 t における1,000回の推定値の正答率を「 $ThinkBenefit_t$ 」と定義し、これを判断の正確性の指標とした¹。また、資源合理性の枠組みを踏まえて、各時点 t における思考コスト「 $ThinkCost_t$ 」も導入し、これを思考コストの指標とした。 $ThinkCost_t$ は、コストを考慮しない場合(i.e., SAT)は常に0とした。一方、コストを考慮する場合(i.e., 資源合理性)には t の増加に伴って線形に増加すると仮定し(Lieder et al., 2018)、その際のコストの最大値 $ThinkCost_{150}$ は「0.50」に設定した。最後に、「 $TotalBenefit_t$ 」を

$$TotalBenefit_t = ThinkBenefit_t - ThinkCost_t$$

と定義し、これを正確さと思考コストとのトレードオフ下における判断の良さの指標とした。この値が大きいくほど、両者のトレードオフ下で良い判断を行うことができたと解釈された。もし適度に短い思考時間が望ましいのであれば、 $TotalBenefit_t$ は t の値が小さい時点でピークを迎えることが予想された。本シミュレーションでは、500人の推定者がいることを想定し、SAT

図2 計算機シミュレーションの結果



注) 枠内は「1人」の推定者による結果の例(左から $ThinkBenefit$ 、 $TotalBenefit$ 、 $ThinkCost$ の結果)を、枠外は「500人」の推定者における $TotalBenefit$ のピーク(枠内右「 $TotalBenefit$ 」で示された赤線の分布)を、それぞれ示す。上: 思考コストを仮定しなかった場合(i.e., $ThinkCost_{150} = 0$)。下: 思考コストを仮定した場合(i.e., $ThinkCost_{150} = 0.50$)。いずれのグラフも、横軸は時間 t を表す。

¹ $ThinkBenefit_t$ の値は、0から1の範囲に収まるように正規化された。なお、本シミュレーションでは、

正規分布に従うパラメータはすべて最小値0、最大値1の範囲で生成された。

を仮定した場合(i.e., $ThinkCost_{150} = 0$), および資源合理性を仮定した場合(i.e., $ThinkCost_{150} = 0.50$)で, 500人における $TotalBenefit_t$ のピーク時点を検証した。

2.2. 結果と考察

シミュレーションの結果を図2に示す。まず, 1人の推定者の結果例について述べる(点線枠内)。SATに基づけば(図2上), 思考コストが考慮されないため, $TotalBenefit$ は $ThinkBenefit$ と等価であった。 $TotalBenefit$ は, 時間とともに向上し, ある時点で頭打ちになるものの, 大きく低下することはなかった。そしてそのピークは, ある程度の時間が経過した頃に見られた。このグラフの形状は, SATの研究で典型的に見られるものである(e.g., Heitz, 2014)。一方, 資源合理性に基づけば(図2下), 思考コストが考慮されるため, $TotalBenefit$ はある時点を境に急激に低下した。このことは, 時間をかけて考えれば正確性($ThinkBenefit$)を高められるが, 同時に思考コスト($ThinkCost$)も増加するため, $TotalBenefit$ は比較的早い時点でピークを迎えることを意味する。

続いて, 500人の推定者における $TotalBenefit$ のピーク時点について述べる(点線枠外)。思考コストを仮定しない場合はピークが比較的遅い時点で幅広く見られた(mean = 95.2, min = 23, max = 150)のに対し, 仮定した場合はピークが比較的早い段階に集中して見られた(mean = 23.2, min = 10, max = 51)。以上より, 人の限られた認知資源を考慮すると, 思考コストを抑えつつ高い正答率を維持するための適度な思考時間が存在することが, 理論的に示された。

3. 実験2: 行動実験

実験2では, 実験1で見出された理論的知見が実際の人の行動と整合的かどうかについて, 行動実験を通して検証した。

3.1. 方法

実験参加者: 日本人83名が実験に参加した。このうち, 42名が後述の「通常群」に($M_{age} = 39.7, SD_{age} = 10.9$), 41名が後述の「長考群」に($M_{age} = 35.0, SD_{age} = 10.5$), それぞれランダムで割り当てられた。なお, 通常群の参加者のうち3名は, データを適切に測定・取得できなかったため, 分析から除外された。

題材・刺激: 実験1で想定したグリッド課題を用いた。

黒の割合は「35%」「45%」「55%」「65%」の4通りとし, それぞれについて40枚の異なるグリッドパターン, 計360枚を刺激として用意した。ただし, 黒35%と黒65%ではほぼ全ての試行が正答だったこと(天井効果), また実験1(正答値が「黒55%」)との整合性を取ることを考慮し, 分析には黒45%または黒55%が呈示された試行のみを用いた。

条件・手続き: 実験はPCおよびマウスを用いて行われた。各試行では, まず0.5秒間の固視点が画面中央に呈示され, 続いてグリッド刺激が1枚呈示された。参加者は, 黒の割合が50%より多いかどうかについて, 「Yes」か「No」のボタンをクリックすることで回答した(図3)。いずれかのボタンが押されると, 再び固視点が出現し, 以上の手続きが繰り返された。いずれの試行も, 開始時にはマウスカーソルが画面中央に配置され, グリッド刺激が呈示されている間, 参加者のマウスカーソルの位置(画面上のxy座標)がフレームごとに記録された。1ブロックは40試行からなり, 1人当たり全9ブロック, 計360試行を行った。本実験では, 「通常群」と「長考群」の2群を設けることで, 思考時間を操作した。通常群では, 参加者は通常通りにグリッド課題に回答した。一方, 長考群では, 刺激呈示から2.5秒間はボタンが呈示されず, 参加者は各試行冒頭の2.5秒間は回答できない(i.e., 強制的に「考える」時間が与えられた)形となっていた。なお, この2.5秒の間でも, マウスカーソルを動かすことは可能であった。

分析指標: 各個人における正答率の平均値(以下「個人正答率」)を, 全体的な正確性の指標とした。また, 時間経過に伴う変化については, 反応時間が「1秒未満」「1~2.5秒」「2.5~4秒」「4秒以上」の4カテゴリに分けて分析した。思考コスト

の指標として, マウストラッキング分析におけるArea Under Curve (AUC; 図4)を用いた。AUCは, 「カーソルの開始位置と選択肢位置とを結

図3 行動実験の手続き(概略図)

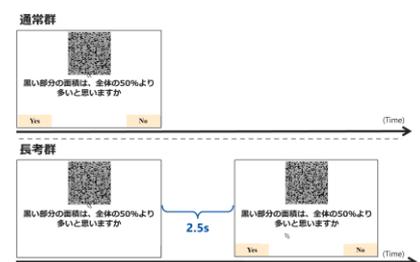


図4 マウストラッキングにおけるAUC (Area Under Curve)



ぶ直線」と「実際の軌跡」とで囲まれた面積と定義される。一般に、AUC が大きいほど、選択時の認知コンフリクトが大きかったと解釈される(e.g., Stillman et al., 2018)。本研究ではこの AUC の大きさを、一種の思考コストの程度と見なした。

3.2. 結果と考察

行動実験の結果を図 5 に示す。まず個人正答率については、通常群よりも長考群の方が有意に高かった(図 5 左; $M_{\text{Normal}} = .671, M_{\text{Long}} = .775, t(78) = 3.11, p = .003, 95\%$ 信頼区間[0.037, 0.171])。よって、考える時間を与えられ、全体としては正答率が向上したといえる。他方で、認知コンフリクトの指標である AUC を見ると、両群とも時間経過に伴って増大する傾向にあった(図 5 中央)。さらに、時間経過による正答率の変化を調べたところ、両群とも、長く考えすぎても(e.g., 2.5 秒以降)正答率は向上しない傾向が見られた(図 5 右)。特に通常群では、1~2.5 秒に正答率のピークが見られ、それ以降では正答率が低下していた。これらの結果は、考えること自体に対して認知的なコストがかかっていること、また長く時間をかけることが必ずしも正答率の向上に繋がるのではなく、むしろ少ない思考コストで正答率を最大化させる適度な思考時間が存在することを示唆している。実験 2 で観察されたデータは、実験 1 で得られた理論的知見を行動レベルで裏付けているといえる。

4. 総合考察

本研究では、従来議論されてきた SAT の枠組みを、資源合理性の観点から検証した。人の認知資源には限りがあることから、考えること自体に認知的なコスト(思考コスト)がかかることが想定される。よって、思考

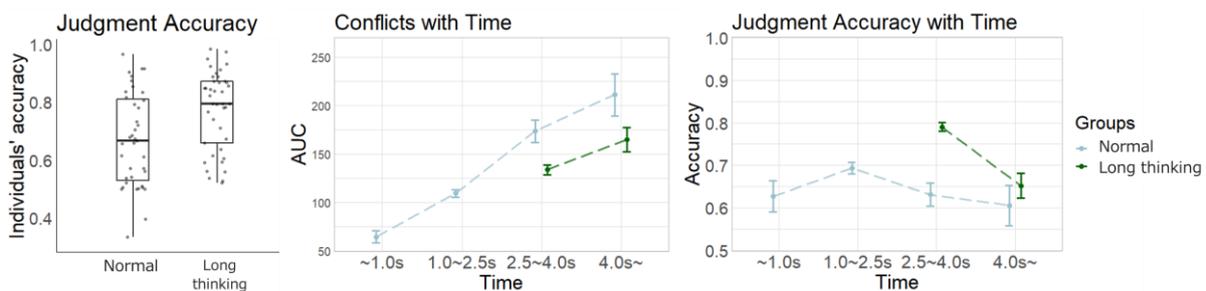
時間は長いほどよいわけではなく、思考コストを抑えつつ正確性を最大化するような適度な長さ存在すると考えられる。本研究ではこの予測を、理論(計算機シミュレーション)と実験(行動実験)の双方から示した。

本研究では、グリッド課題という単純な知覚判断課題のみを題材としていた。そのため、結果の一般性・応用可能性については、今後検討の余地があるだろう。また、実際に「適度な思考時間」を与えた際に、(思考時間を与えない、または長すぎる思考時間を与えた場合と比べて)正答率が向上し、かつ判断者の認知負荷も抑えられるかどうかを検討することで、よい判断を促す介入策の検討にもつながることが期待される。

文献

- Callaway, F., Griffiths, T. L., Norman, K. A., & Zhang, Q. (2023). Optimal metacognitive control of memory recall. *Psychological Review*, Advance online publication.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *TopiCS*, 7(2), 217–229.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Front in Neuros*, 8, 150.
- Ho, M. K., Cohen, J. D., & Griffiths, T. L. (2023). Rational simplification and rigidity in human planning. *Psych Sci*, 34(11), 1281–1292.
- Lieder, F., Griffiths, T. L., Quentin, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin & Review*, 25(1), 322–349.
- Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends CS*, 22(6), 531–543.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41(1), 67–85.

図 5 行動実験の結果



注) 時間をかければ個人正答率は向上する(左)。しかし時間経過に伴って認知コンフリクトは増加し(中央)、また正答率が上昇し続けるわけではない(右)。「Normal」は通常群を、「Long thinking」は長考群を、また「AUC」はマウストラッキングにおける Area Under Curve を、それぞれ示す。中央および右のグラフで、各点は当該カテゴリにおける平均値を、エラーバーは 95%信頼区間を、それぞれ表す。