

合理性と目標多重性：限定合理性と二重合理性を超えて

Rationality and purpose-multiplicity: Beyond bounded rationality and dual rationality

服部 雅史[†]
Masasi Hattori

[†]立命館大学総合心理学部
College of Comprehensive Psychology, Ritsumeikan University
hat@lt.ritsumei.ac.jp

Abstract

According to the heuristic and bias approach, Type 2 processes correct errors caused by Type 1 processes. Some recent studies, however, emphasize the advantage of unconscious Type 1 processes, which are supposed to be rational in both normative and adaptive senses. I propose rationality should be located in view that tasks we face are concerned with compound purposes. I also propose Type 1 process fulfills the habit-bounded rationality rather than normative rationality, and Type 2 process involves intentional diversion and decontextualization from a view of different (meta-)levels. I speculate on a possible extension of the idea toward a unified theory of cognition

Keywords — normativity, adaptiveness, automaticity

1. 二重過程と合理性

思考の認知科学において、合理性の概念は重要な役割を果たしてきた。それは、意思決定や推論の認知過程が、規範と記述の溝 (Stanovich, 1999), すなわち、論理や確率などの規範理論と実際の人間の行動の乖離を糸口として明らかにされてきたことと無関係ではない。規範的合理性は、定められた目標に対するシステムの出力として、どういったものが望ましいかを定義する。しかし、規範的合理性は、その過程を規定するものではないので、認知過程の特性は、規範と実際の出力の差、すなわちエラーから推定されるしかない。

この点において大きな成功を収めたのが「ヒューリスティックとバイアス」アプローチ (see, e.g., Evans, Newstead, & Byrne, 1993; Kahneman, 2011) である。このアプローチの成功によって、自動的に素早く直感的な「タイプ1」過程 (システム1) がしばしばエラーを引き起こし、反対に、熟慮的で時間がかかりコントロールされた「タイプ2」過程 (システム2) が、バイアスを修正して錯誤の回避を可能にするという二重過程 (dual process) の考えが広まった。すなわち、タイプ1過程は、最小限の認知資源しか使わずに、大抵の場合にうまくいく処理を行う。一方、タイプ2過程は、多

くの認知資源を消費して、統合的で合理的な解を導き出すための処理を行う。この見方によれば、タイプ1は適応的合理性 (または進化的合理性) を満たし、タイプ2は規範的合理性を満たすことになる。

なお、ここでいう規範的合理性 (normative rationality) と適応的合理性 (adaptive rationality) は、Anderson (1990) による区別で、それぞれ、数学的規範から逸脱しないこと、環境の中で適応的に生存することを意味する。この二重合理性の概念は、Harman (1995) による理論的合理性 (theoretical rationality 信念の合理性) と実践的合理性 (pragmatic rationality 行為あるいは計画や意図の合理性) の区別と緩やかに対応しており、その源流は、Simon (1955, 1956, 1978) にさかのぼる。それは、環境の構造と認知資源の有限性を考慮に入れた限定合理性 (bounded rationality) の概念である。

しかし、二重過程と二重合理性の対応については、対照的な主張もある。たとえば、Gigerenzer らの研究グループは、現実的な意思決定において、直感的で単純なヒューリスティックが、むしろ正解に導くことが多いと主張してきた (e.g., Gigerenzer, Todd, & The ABC Research Group, 1999)。洞察問題解決研究では、無意識的なタイプ1過程が、分析的なタイプ2過程では到達できない新しい解に導くとされている (e.g., Gilhooly, Ball, & Macchi, 2015)。さらに進んで、判断課題や推論課題については、系統発生的に古いとされるタイプ1過程が、合理的選択理論といった規範に合致する行動を実現し、むしろ、言語と結びついたタイプ2過程こそがエラーを引き起こすという主張もある (Oaksford & Hall, 2016)。この最後の見方によれば、タイプ1は、適応的合理性も規範的合理性も満たすことになる。もしそうなのであれば、進化的に新しいはずのタイプ2過程が非合理的であるのはなぜだろうか。また、タイプ1過程が合理的というときの合理性とは、正確にどのような意味なのであろうか。

2. 目標多重性

二重過程と合理性の関係を明らかにするために、エラーと合理性の関係について再検討しておきたい。Sloman & Fernbach (2008) は、因果推論のシステムチェックなエラーの例として、自己欺瞞による行動変容や、四人のジレンマにおける「協力」を挙げている。たとえば、Quattrone & Tversky (1984) は、腕を冷水につける実験で、心臓が健康な人は長く耐えられると聞いた被験者は、これ以上耐えられないと感じるまでの時間が長くなることを明らかにした。これは、明らかに自己欺瞞によるもので、規範に照らせば「因果推論のエラー」(Sloman & Fernbach, 2008, p. 489) である。しかし、このエラーは非合理的であろうか。自己欺瞞が満足感や自己効力感をもたらす、その結果、幸福感が得られる (Taylor & Brown, 1988) とすれば、この行動は、「気持ちよく生きる」という生物としての基本的な目標に合致するという意味で、合理的なのではないだろうか。

合理性の定義は目標に依存する。たとえば、冷水実験の課題では、正しい因果推論がなされることが暗黙のうちに目標に設定されているが、被験者の目標は、必ずしも実験者の目標に一致するとは限らない。さらに重要なのは、おそらく被験者は、同時に複数の目標を保持しているという点である。つまり、被験者にとっては、正しい因果推論をすることも幸福感を得ることも、さらには別の（もしかしたら個人的な）何かも、すべてが同時に目標であり得る。そうした目標が相互排他的でない限り、どの目標もある程度の水準で満足がいくように達成しようとするのが、実践的な意味で合理的な行動と見ることができる。

もちろん、課題としてあることが要求されており、それを被験者は十分に理解・納得しているにもかかわらず、それに応えられないとすれば、それを以って非合理的とすることには一理ある。しかし、その課題が、本来自然に引き起こす認知過程を抑制して、その場で与えられた課題要求だけで上書きすることを要求しているとすれば、それが本当に適応的であるのかどうかについては議論の余地がある。ただ、この点については、本稿では議論しない。

3. 習慣限定合理性とメタ合理性

タイプ 1 過程は、おそらくこうした多重化された目標を並行的に考慮することができるように、自動的で並列的な処理過程として実現されているのではないだ

ろうか。こうしたタイプ 1 過程は、Oaksford & Hall (2016) によれば、規範に合致するとされている。確かに、ベイズ推定という規範に合致するという意味で、タイプ 1 過程の規範性の主張には一定の妥当性はある。しかし、問題は、この処理が固定配線であることである。たとえば、ムクドリの意味決定が、合理的選択理論の公理を満たしている (Monteiro, Vasconcelos, & Kacelnik, 2013) という事実は驚きではあるが、それが一定の環境における固定化された行動であることを思えば納得できる。つまり、タイプ 1 の推論や判断は進化的に古いもので、その出力の有効性は領域依存性である。したがって、これは規範的合理性というよりは、むしろ**習慣限定合理性 (habit-bounded rationality)** と呼ぶべきものであろう。

タイプ 2 処理は、こうした習慣限定合理性を満たすタイプ 1 処理を上書きすることができる。ただし、それが結果としてエラーを引き起こすことがあるのも事実である (Oaksford & Hall, 2016)。しかし、タイプ 1 過程のモジュール（もしくは、学習された硬直化したヒューリスティック）は、そのモジュールがチューニングされた文脈以外の文脈では、望ましい出力をしないことがある。そのため、望ましくない出力を慎重に防ぐよう行動することを可能にするのは、やはり、タイプ 2 過程であろう。つまり、多重化された目標の中で、特定の目標を焦点化して合理的にそれを追求するといった規範的合理性が実現できるのは、タイプ 2 処理に限られる。

しかし、こうした合理性よりも、おそらくずっと重要な合理性は、大局的観点から意図的に（あるいは非意図的に）方向転換するといったダイナミックな思考や行動ではないだろうか。それは、たとえば、問題解決が行き詰まったらまったく別のことをする、雑談の最中に気まずい雰囲気になりかけたらしき話題を切り替える、締切り間際になって緊急性のない別の何かをやり始める、ある日突然思い立って旅に出るといったことがある。これらの行動は、文脈の切り替えや脱文脈化、目標レベルの変更といった概念でとらえることができる。この種のメタレベルの思考や行動を起こすのは、見かけほど簡単なことではない。近年、ディープラーニングで再び脚光を浴びている人工知能にとって、おそらく、こうした目標直結型ではない振る舞いは、最も実現が難しいタイプのものではないだろうか。こうした知性のもつ性質としての合理性を、**メタ合理性 (meta rationality)** と呼ぶことにする。

4. 意識の役割

本研究では、メタ合理性を実現するためには、高水準の意識が必要であると提案したい。Humphrey (2011) が主張するように、いま、意識の機能を、それがなければできない何かをさせることではなく、それがなければやる気が起こらないことにやる気を出させることであると仮定しよう。あるいは、自分の一生を生き甲斐のあるものにする、自分自身をかけがえのないものとみなして尊ぶといったことと仮定してみる。また、自己効力感や責任感を介して生き甲斐や幸福感を得ることが、進化的な適応度を上昇させるとする。そうすると、自己欺瞞、自己中心性、自我関与性、自己受容、自我防衛といった要素は、与えられた実験的・無機的な課題の中で、多重化する目標の一つに入り込む余地が常にあると考えてよいであろう。つまり、意識を持つことによって発生する目標があり、その目標の達成が、生物としての環境への適合度を高めている可能性があるということである。

タイプ1処理は、そうした多重化目標を並列的に達成することを目指すものと捉えることができる。一方、タイプ2処理は、第1に、そうした多重化目標を分離して、単一化して合理性を追求することを志向するものと位置付けることができる。第2に、多重化目標の達成が容易でない場合に、メタレベルから介入することによって、ダイナミックなタスク切り替えが促され、それを通して創造的解決を志向するものと見ることができるのではないだろうか。

以上の考えは、まだ単なる推測にすぎないが、こうした観点は、これからの認知研究の中に、自己の概念やその影響、協力行動や援助行動、責任感、生きがい、幸福感といった概念、延いては、意識のはたらきによる志向性や死の脅威といった概念までも自然に取り込んで、統合的な認知理論を構築するための重要なステップになるのではないだろうか。

5. 引用文献

- Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Evans, J. St. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hove, UK: Lawrence Erlbaum Associates.
- Gigerenzer, G., Todd, P. M., & The ABC Research Group (1999). *Simple heuristic that make us smart*. Oxford, UK: Oxford University Press.
- Gilhooly, K. J., Ball, L. J., & Macchi, L. (2015). Insight and

- creative thinking processes: Routine and special. *Thinking & Reasoning*, 21, 1–4.
doi:10.1080/13546783.2014.966758
- Harman, G. (1995). Rationality. In E. E. Smith & D. N. Osherson (Eds.), *Thinking* (2nd ed., pp. 175–211). Cambridge, MA: MIT Press.
- Humphrey, N. (2011). *Soul dust: The magic of consciousness*. Princeton, NJ: Princeton University Press.
- Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.
- Monteiro, T., Vasconcelos, M., & Kacelnik, A. (2013). Starlings uphold principles of economic rationality for delay and probability of reward. *Proceedings of the Royal Society B: Biological Sciences*, 280.
doi:10.1098/rspb.2012.2386
- Oaksford, M., & Hall, S. (2016). On the source of human irrationality. *Trends in Cognitive Sciences*, 20, 336–344.
doi:10.1016/j.tics.2016.03.002
- Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46, 237–248.
doi:10.1037/0022-3514.46.2.237
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69, 99–118.
doi:10.2307/1884852
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63, 129–138.
doi:10.1037/h0042769
- Simon, H. A. (1978). Rationality as process and as product of thought. *The American Economic Review*, 68, 1–16.
- Sloman, S. A., & Fernbach, P. M. (2008). The value of rational analysis: An assessment of causal reasoning and learning. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 485–500). New York, NY: Oxford University Press.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
doi:10.1037/0033-2909.103.2.193