

認識的パターンナリズムへの道は開かれたか

——現代知性改善論序説のための断片——

植原亮

1. はじめに——自然化された認識論と知性改善論

自然化された認識論では、認識論を科学と連続した営みと見なし、心理学や認知科学などが提供する経験的知見や概念の導入により認識論上の問題を再定式することで、その解決が目指される。論文「自然主義者は帰還する」の中でキッチャーは、認識論がこのような自然化を経た後でもなお認識論と呼べるのは、ひとつにはそれが知性改善論の構想という認識論の伝統的課題を引き継いでいるからだと主張している。知性改善論は、人間の認識の特性を把握したうえで、それを改善するための「方略 strategy」を特定しようとする。ベイコンの『新機関』やデカルトの『精神指導の規則』などの古典的著作は、まさしくそうした知性改善論の試みにほかならない。「信頼性主義 reliabilism」の表現を借りれば、人間の信念形成プロセス⁽¹⁾の信頼性、すなわちそのプロセスが真なる信念を生み出す割合を向上させる方略を見出すことが知性改善論の課題なのである。何が実際に信頼性の高い信念形成プロセスなのかは、人間において現に生じる認識のありようをつぶさに見ることで明らかになるが、そのための最良の知見は心理学などが提供してくれる。したがって知性改善論は、そうした経験科学の知見を積極的に参照することで構想されねばならない。このように考えれば、知性改善論は科学と連続的なものとして、自然化された認識論の中に位置づけられるのである⁽²⁾。

本稿ではこのような見方を踏まえたうえで、次のように問うことにしたい。それでは、具体的には現代の知性改善論は、とりわけキッチャーの論文以降においてはどのように論じられているのか、と。この問いに対する部分的解答として、知性改善論の現状の一端を私なりに描き出すことで、最終的には「認識的パターンナリズム epistemic paternalism」への道筋を明らかにし、それが真剣な検討に値する立場であることを示すのが、本稿の目的である。だが以下ではまず出発点として、いわゆる「内在主義 internalism」に依拠する知性改善論の方向が困難に突き当たるということを手短かに確認しておきたい。そうすることにより、現代的な知性改善論が自然主義に即して構想されざるを得ないということがいつそう鮮明となるだろう。

知識の古典的定義によれば、知識とは真なる信念であるとともに正当化されたものでな

なければならない。内在主義ではこの正当化の中身を、当該の信念が真であるという理由を認識主体が自らの心の中に信念という形で抱き、しかもそれに対して「内的な」アクセスが可能であること、つまり意識の中においていわば透明な仕方では把握しうるものとして規定する。これは、何かを知っているといえるには、少なくともその理由を主体が自分で述べることができねばならないという直観を表明したものといえる。こうした内在主義的な知識観に依拠するなら、知識を獲得するには信念の正当化のための理由を自らの意識の中で透明な仕方では把握せよ、という助言を与えることが、知性改善論の役割となる⁽³⁾。なるほどこの助言に従えば、信念の形成ないしは保持のプロセスの信頼性が向上させられるように思われる。

だがこの方向はすぐに大きな困難に突き当たる。それはひとえに、内在主義が正当化に求める基準を満たすことが現実の人間にはできないと考えられるからだ。ある信念を正当化するために他の信念に訴える、という状況を考えてみよう。そのためには、主体はふたつの信念だけでなく、一方が他方の理由になっているという関係をも内的に把握していなければならない。このとき、その理由関係の把握もまた信念形成という仕方ではなされないと考えられる。しかし、そこで出てきた、信念間の理由関係についての信念が正しいと思う理由は何か。この二階の信念にも正当化を与えない限り、最初の信念も十分に正当化されたとはいえない。ところがその正当化には、また別の信念を理由として挙げねばならないため、その理由関係についての内的な把握も、信念という形で必要となるだろう。そうすると今度は、その三階の信念にも何らかの正当化が求められるのではないか。

一見したところこの状況は、内在主義的な正当化は無限遡行に陥るがゆえに実行できないということを示唆しているように見える。しかしここはもう少し検討が必要である⁽⁴⁾。まず、無限遡行の懸念に対しては、人間の信念が有限個であることを挙げ、有限個の信念からなる信念体系の内部では理由関係の把握も無限遡行に陥ることなくいずれ止まる、との指摘がなされるかもしれない。だがそこからは別の問題が出てくる。信念体系内の信念がたとえ有限だとしても、それら全体の論理的整合性を確かめる作業でさえ、人間の頭では現実には実行不可能なほどの莫大な時間を要するからである(cf. Cherniak, 1986, p.93, 邦訳, 145-146 頁)。ただし、実際にどのくらいの時間を要するのかを正確に述べるには、人間の認識の限界に関する経験的知見の参照が欠かせないといえる。

そこで次に、人間の認識に関するそうした知見を具体的に参照して検討してみよう。すると今度は、内在主義の助言に従うことがそもそも本当に信念形成プロセスの信頼性を高めるのか、という疑念が新たに生じてくる。たとえば「位置効果 position effect」(Nisbett & Ross, 1980)や「選択盲 choice blindness」(Johansson, Hall, Sickstrom & Olson, 2005)などの研究

は、人間が自分の選好についての判断ないし信念の理由を問われたときに、その本当の理由・原因とは無関係なものを挙げ、しかも本人はそのことに気づいてすらいない場合も多いということを明らかにしている。こうした知見は他にもあるが、いずれも内在主義から提出される助言の実行可能性を疑わせるのに十分である。ある信念を正当化する本来の理由が主体の内的なアクセスの届く範囲にあるとは限らないなら、内在主義の課す正当化の基準は満たしえない過大な要求なのではないか、と考えられるのである。

ここまでの検討から、内在主義に依拠した知性改善論では、現実の人間には実行困難な方向に進んでしまうように思われる。規範（「すべし」）が可能性（「できる」）を含意するなら、そうした方向は進むべき根拠がないことになる。こうして目指すべき方向が見えてくる。内在主義的な知識観から脱し、人間の認識の限界に関する経験的知見を参照しながら（つまり自然主義的な観点から）、実行可能な知性改善論を構想せねばならないのである。ではそうした知見として、たとえば上に見た位置効果や選択盲の研究は役立つだろうか。本人にも見通せない不透明な心理的領域が人間には存在するという知見から、信念形成プロセスの信頼性を向上させるための方略について何らかの示唆が得られるのではないだろうか。次節ではまさにこのような観点から、内的アクセスの及ばない領域に重要な位置を与える「二重過程理論 dual process theory」という認知科学上の学説に着目したいと思う。

2. 二重過程理論とその応用としての知性改善

二重過程理論によれば、人間の認知は「システム1」および「システム2」がそれぞれ律するふたつの過程からなる⁽⁵⁾。システム1は広くいって直観的な働きを司り、素早く自動的で、意識的な努力は不要か、あるいはわずかしかな必要としない。ここに含まれているのは、突然の物音の方向を直ちに感知するような生得的な能力や、8×9の答えを瞬時に導き出すといった反復的訓練により培われる技能などである。他方のシステム2は、熟慮的な思考過程を実行するものだが、素早く自動的に働くシステム1に比べるといわば怠け者であり、そのスピードは遅く、働かせるには意識的な努力を要する。15×13を暗算で計算したり、新しい冷蔵庫の購入について性能や価格などの点から総合的に検討したりするときに、その出番である。前節での内容に即せば、システム2には主体の内的アクセスが及びうるがシステム1にはそれが及びえない、と対比することができるだろう。

二重過程理論と知性改善論の接点は、内在主義的な方向の限界に関わるところだけではない。二重過程理論の妥当性は多くの研究によって強く支持されているが、そのかなりのものが知性改善のための方略を具体的に示唆してくれる。システム1はバイアスの系統的な発生源であることが明らかにされており、そうしたバイアスの実態を解明し、システム

1の働きに適切に対処できれば、我々の信念形成プロセスの信頼性を高めることができると考えられるのである。

さてそれではバイアスとはどのようなものだろうか⁽⁶⁾。その代表例を見ておくことで二重過程理論の内実がもう少し明確になる。次の問題を考えてほしい——日本で溺死の数は、屋内と屋外のどちらが多いだろうか？ 直観的には屋外での溺死が圧倒的に多いと感じる。我々はメディアの報道でよく目にする海や川での水死・遭難事故を直ちにイメージし、屋外での溺死が多いという結論に飛びついてしまいがちだ。だがこれは不正解である。実際には、少数の人間がたまに行くだけの海や川での溺死よりも、ほとんどの人間が毎日入る風呂での溺死の方がずっと多いのである。これは落ち着いて考えればわかることだが、システム1が働いて、どうしても「屋外」という誤答を素早く自動的に導き出してしまう。このように、頭に浮かびやすい事例についてはその発生件数・確率を大きく見積もってしまう傾向のことを「利用可能性バイアス」と呼ぶ。こうしたバイアスはいくつも存在が知られているが、内的アクセスの及ばないシステム1の領域で生じるだけに、人間はその存在に気づきにくく、結果的に誤信念を抱きがちになる。

というわけで、バイアスの脅威を深く自覚し、その影響から免れる「脱バイアス debiasing」が、二重過程理論に依拠した知性改善論の主要な目標となるように思われる。システム1は、日常生活における多様な問題を迅速かつ少ない労力で解決してくれるという点で、非常に有益であることも確かである。だが真なる信念の形成が決定的に重要であるような場面では、システム2を働かせて、システム1で生じるバイアスのせいで自分が誤信念を抱いていないかに注意せねばならない。そうしてあとはバイアスに影響されないように信念を形成するための適切なやり方で問題解決を図ればよい。具体的には、論理学の規則を念頭に置きながら推論してもよいし、論証図を紙に書きつけながら考えてもよい。あるいはあなたが医師なら、病気を診断するときに、ミールが考案した予測モデルを使えば、数値を式に代入するだけで人間がするよりも正確な診断結果を簡単に得ることができる(Kahneman, 2011, chap.21, cf. Ahlstrom-Vij, 2013, p.24)。こうして誤りの可能性を避け、種々の思考ツールを用いながら、真なる信念を形成する見込みの増大を目指すというわけだ。

このような脱バイアスを目標に掲げる知性改善論の方向は、ひとつには批判的思考力の育成を目指す教育実践として具体化できると考えられるだろう。直観的には正しく思われる主張ないし信念に懐疑の目を向けて批判的に吟味する能力を培うことは、システム1において生じるバイアスの影響を減じるべくシステム2を適切に働かせ、適宜必要な思考ツールを運用する能力の習得として理解できるように思われる⁽⁷⁾。実際、批判的思考に関する近年の研究でも、二重過程理論から教育実践上の含意を引き出すべく、この理論に依拠

した議論が現れ始めている⁽⁸⁾。

3. 脱バイアスはどこまで可能か

二重過程理論を援用した批判的思考力育成の試みは、バイアス研究を通じて突き止められてきた人間共通の弱点を克服しようとする点で、知性改善論として実に理にかなった有望な方向であるかに見える。しかしそう判断する前に、内在主義的な方向が突き当たったのと同種の問題が生じないか、つまりその実行可能性について検討しておく必要があるだろう。はたして、システム2を働かせて脱バイアスを図れという助言は、現実においてはどのくらい有効なのだろうか。

あいにくバイアス研究そのものから少し悲観的な示唆が得られている。最も深刻な影を落とすのは、「バイアス盲点」として知られる現象である(Pronin, Lin & Ross, 2002)。バイアス盲点もまた人間に共通して見られるバイアスの一種といえるが、簡単にいえばそれは、自分への自信過剰などにより、バイアスに自分が影響されていることにも気づかないという現象である。そしてこれが脱バイアス一般への動機づけを甚だしく困難にしてしまう。たとえば利用可能性バイアスについて教示し、その脅威に注意すべくシステム2を働かせるよう助言したとする。だが助言された側は、バイアス盲点が生じている場合にはこう考える。「なるほど人には利用可能性バイアスなるものが生じるのか。だがこの俺は別だ。そんなバイアスなんかには引っ掛からない自信がある」。このように、バイアス盲点が生じていると、脱バイアスなど自分には不要と考えて、批判的思考力を身につける動機が著しく低下してしまうことになる。

ここでもしかすると、それならバイアス盲点についても教示し、その脅威に注意を払うべくシステム2を働かせるよう助言してはどうか、といいたいくなるかもしれない。しかしこれもうまくいかない。なぜならバイアス盲点もまたバイアスの一種である限り、それに関するバイアス盲点が生じかねないからだ。「なるほど人にはバイアス盲点なるものが生じるのか。だがこの俺は別だ。そんなバイアス盲点なんかには引っ掛からない自信がある」という具合である。こうした場合、いったいどうすれば脱バイアスを図れという助言に従う動機が与えられるのだろうか。

バイアス盲点はきわめて厄介だが、それでも脱バイアスがバイアス盲点のせいで必ず失敗するというわけではない。バイアス盲点の存在を深く自覚させ、その影響の排除を動機づけるための有効な手段として、信念形成プロセスの信頼性についての次のようなフィードバックを与える方法が知られている(Sieck & Arkes, 2005)。すなわち、信念形成のあとすぐに、①自分の信念にどの程度の修正が必要だと思うかを明確にする、②それが実際に必

要な修正の度合いとどのくらい合致しているかをグラフで示してもらい、③自分がバイアス盲点に陥っている度合いに関してそのグラフが何を示唆しているか問われる、というものである。なるほどこれなら、バイアス盲点のせいで自分の信念形成プロセスの信頼性を過信していた場合でも、自分のバイアスに否応なく向き合うことになり、脱バイアスが強く動機づけられるだろう。そうしてバイアス盲点が克服できれば、もっと信頼のおける信念形成プロセスを実現するための代替的な方法の習得への意欲も引き出される、というわけである。

ところがこれで片付くほど事態は容易ではない。というのも、それに伴うコストを考慮するなら、現実における実行可能性への疑念が依然として払拭できないからである。脱バイアス一般を動機づけるために、たいていの場合にはバイアス盲点の克服も必要だとしよう。しかしそのためには、一定以上の数の人間を対象にする場合にはとりわけ多くの資源を投入せねばならない。上で見たような迅速できめ細かいフィードバックを個々の人間に与えようとすれば、膨大な時間や人的労力などのコストが生じると見てよい。はたしてそうしたコストを負担する用意が、誰にどれほどあるのだろうか⁽⁹⁾。

さらに、この方略の実行可能性の問題がコスト面ではかりに解決できたとしても、実のところ、その有効性の範囲については疑念を差し挟む余地が残っている。きわめてありそうなことだが、せっかく資源を投じて教育を施し、いったんは脱バイアスを達成させて批判的思考力を身につけさせても、教室の外ではそれが発揮されないかもしれないのである。そうした能力は、教室の中で与えられるような課題にのみ発揮される領域特異的なもので、しかも教育後は時間の経過とともに低下していくことを示唆する研究もある（教育を断続的に受けさせれば能力は維持されるが、そのためのコストがさらに発生してしまう）⁽¹⁰⁾。もしこのように教室外の実生活における領域一般的な課題に対してあまり有効でないなら、知性改善論として成り立つ範囲がかなり限定されてしまうことが危惧される。

とはいえ、脱バイアスならびに批判的思考力育成は大いに試みるに値するものである。まず、それがきわめて有効な集団も存在するだろう。そうした集団は、簡単な批判的思考力のコースを受講するだけで、各種バイアスの脅威を適切に理解してくれるだろうし、さらにその後は自己教育を通じて信頼のおける信念形成プロセスを実現できるようになるかもしれない。その場合、かかるコストもわずかで済む。次に、批判的思考力が領域特異的でありうるという点も必ずしも悪いことばかりではない。それが際立った力を発揮するような範囲や分野がうまく特定できれば、望ましい結果が得られるかもしれないからである。最後に、低コストで効果的な教育手法がこれから見出される可能性も残っていると思われる。だが以上がいずれも現状では未解明な要素を多分に含み、今後もさらなる探究と実践

を必要としていることは間違いない。

そこでバイアスの影響を排除することを重視しつつも、実行のコストが小さくて済むような、別の方向を模索しておくことできないものだろうか。達成困難でコストのかかる脱バイアスを個人に求める手順を踏まず、外的な制度などの力によって信念形成プロセスの信頼性を向上させることができるなら、それも知性改善の望ましい方略なのではないか。こうした考察を背景に登場する立場こそ、認識的パターンナリズムにはかならない。

4. 認識的パターンナリズムの可能性

認識的パターンナリズムとは、大枠では、ある個人の信念形成プロセスに対する外部からの干渉がその信頼性を向上させるのであれば、その干渉は本人の同意なく許される場合がある、とする立場のことである⁽¹¹⁾。これが「認識的」なのは干渉が信念形成プロセスの信頼性に関わるからであり、また「パターンナリズム」と呼べるのは本人の同意なき干渉が許されうるとするためである。本人の力や教育で脱バイアスを実行するのは容易ではないが、外部からの干渉によりバイアスの発生を未然に防ぐことができれば、個人の信念形成プロセスの信頼性は高まる。そうした干渉は、脱バイアスを試みるよりもはるかに低コストだろうし、本人の健康など非認識的に価値ある状態を損なうものでない限りは、何よりも真なる信念の獲得を促してくれるがゆえに正当化される、というのである。

こうした主張の実質を理解するには具体例が参照できればよい。幸いにも社会には認識的パターンナリズムの実践として捉えられる制度の例がすでに存在するので、ここではそのうち、司法の場で信念形成プロセスに対する干渉が行われている場面を見ておこう⁽¹²⁾。陪審員制度が採用されている米国では、被告に関する情報への陪審員のアクセスに制限を設けている。被告の性格や前科などは陪審員が知っておくべき事柄であるように思えるが、現実には、陪審員はそうした情報に過剰な重みづけを与えてしまい、適切な司法判断が下せなくなるという傾向が見られる。この傾向もまた「後光効果」などのバイアスのせいで生じるが、そうしたバイアスの影響を未然に防ぐべく、陪審員が被告の性格や前科などの情報に触れることのないように、アクセスできる情報の範囲を制限するというやり方が採られている。これは情報への自由なアクセスを制限する点で、信念形成プロセスへの干渉として位置づけられるものである。なるほどこれなら陪審員はシステム2を働かせて脱バイアスを果たすという困難な課題を負わずに済むし、またコストもほとんどかからず、ここでの信念形成プロセスへの外部からの干渉は、本人たちの同意なく行われるものの、適切な司法判断を下すという目的に照らして、十分に正当化されると考えられる。

さらに認識的パターンナリズムでは、こうした既存の実践の正当化だけでなく、その適用

の拡大も主張される。たとえば、病気の診断にミール考案の予測モデルを使用してもらうことを考えてみよう。医師にバイアス盲点がある場合、人間の診断よりも予測モデルの方が通常は正確だと医師が理解はしていても、自分の診断に限ってはバイアスを免れていると考えてしまうため、予測モデルを使用してくれないということが起こりかねない。そこで、予測モデルの使用を本人の同意によらず法的に義務づけるなどすればよいだろう。それにより、診断の正確さが高まるのはもちろん、見込まれるコストもさほど大きくないため、ここでも認識的パターナリズムは正当化できると考えられるのである⁽¹³⁾。

こうした領域をはじめ、認識的パターナリズムの適用が正当化されうる領域は他にも見出していけるだろう。どこがそうした領域で、何がそこで有効な方略となるか。そしてそれはどのように制度化可能で、どのくらいのコストが生じるのか。こうした点を明らかにしていくことが、認識的パターナリズムにとって今後は重要な課題のひとつとなるだろう。

だが当然の疑問も生じてくるだろう。はたしてそのような適用拡大はどこまで認められるのだろうか。とりわけ本人の同意なしに信念形成プロセスに干渉することは、自分の信念をどう形成するかを自分では決定させない点で、自律性の侵害に当たるのではないか。もしこのような自律性侵害の懸念が大きければ、結局のところ認識的パターナリズムの実践はごく狭い範囲に限定されるべきであり、したがってその適用拡大という主張は実質的に成り立たなくなるのではないだろうか。本稿では認識的パターナリズムに突きつけられるすべての問題を扱う余裕はないが、最後にこのもっともな疑問を取り上げ、それに対してどのような応答が可能なのかをごく簡単に検討して締めくくりたい。

まずこの疑問に対しては、そもそも自律性の侵害というなら、誤信念を生み出し、自分の利益にかなうような意思決定を妨げるかもしれないという点で、バイアスの方こそが我々の自律性の基盤を掘り崩す恐れがある、といった応答が可能だろう。外的な制度の力でそうしたバイアスの影響を取り除いたり発生を未然に防いだりすることは、むしろ自律性の十全な発揮を支援することであるように思われる。かりに外部からの干渉に伴う自律性侵害が多少はあるにせよ、それを補って余りある見返りがここには期待できるのではないか。少なくとも自律性の発揮と侵害のバランスがうまくとれている限りは、ここでの疑問を斥けることができると考えられるのである。

一方でこうした応答とは別に、バイアスに対して外部から干渉することが、そもそも本当に自律性の侵害に当たるのか、という点からの異議を申し立てることもできるだろう。システム1の領域で生じるバイアスは、個人の身体の内側にありながらも本人の内的アクセスが届かず、しばしば意図せぬ結果を引き起こす。このような意味で、システム1ないしバイアスは、我々にとってむしろ外部に位置づけられるべきものといえるかもしれない。

もしそうなら、認識的パターンリズムの実践は、当人への直接的な干渉というよりも、個人の外部にあるものへの干渉として捉えられるべきものであり、その点ではそれが直ちに自律性の侵害に相当するとはいいい切れないのではないか、というのである⁽¹⁴⁾。

この異論のように、個人ないし自律性というものの範囲をいわば切り詰めていくことで認識的パターンリズムを擁護することも、ひとつの手としては可能だろう。しかし、システム1やバイアスのみならず、そもそもある意味で信念形成プロセスが一般に外部性を帯びていることに注目すれば、もう少し別の応答の仕方ができるかもしれない。その外部性とは、個人の内部に限定できない歴史性や共同性である。我々はシステム2を働かせてさまざまなやり方で信念を形成するが、それは多くの場合、思考ツールを使用して行われる。そうした思考ツールには、暗算や内語のように内化して使用するものもあれば、ミールの予測モデルのように身体の外に置いたまま使用するものもあるのだが、いずれも通常は歴史上のどこかで先人たちに案出され、改良されながら世代間で継承されてきたものである。個人の外で展開するこうした歴史的な共同事業に支えられているという意味で、個人が思考ツールを使用して信念を形成するプロセスは、実際は多くの外的な要素にも支えられているといつてよい。この外部性なしでは、我々はそれなりに信頼のおける仕方での信念を形成することは難しいのである。だが必ずしもそれが我々の自律性を脅かすとはいえない。むしろそうして形成される信念ないしは知識が個人の利益に貢献するなら、ここの外部性はかえって我々の自律性に資するとさえいえるのだ。

おおよそ同じことが、認識的パターンリズムの提供する外的制度についてもある程度成り立つのではないだろうか。適切に与えられたなら、それもまた個人の信念形成プロセスを外から支える足場のようなものとして働き、プロセスの信頼性を高めて個人にとって有益な知識を生み出すことに資するからである。ただここの主張するには、少なくとも認識に関わる範囲での自律性の実質について再考し、必要に応じてその概念を改訂することが必要となるだろう。自律性とは、個人の外部から認識を支える足場が、たとえ本人の同意がなくとも、適切に提供されているおかげで、ようやく十分に発揮されうるものにほかならない。このように自律性が捉え直されれば、認識的パターンリズムがそれを侵害するのではないか、という懸念にも新たな視点から応答可能になるのである。

このように認識的パターンリズムは、その適用範囲の明確化はもちろん、自律性のような基本的概念の再考を含む理論的な整備にも、重要課題のひとつとして取り組まねばならない。その見通しの検討は別の機会に譲らざるを得ないが、少なくとも以上により、人間の手に届く知性改善論の選択肢として真剣な検討に値する可能性を認識的パターンリズムが有しているということは、十分に示されたと思われる。

註

- (1) 本稿では便宜上、個人の頭の中だけではなく、情報の収集や紙の上での計算など個人の外部で生じる過程をも含む広い意味で「信念形成プロセス」の語を用いる。
- (2) Kitcher(1992)を参照。ここでの信頼性主義についてはさらに Goldman(1992)なども参照。知性改善には信念形成プロセスの信頼性だけではなく、速さや頑健性なども評価の要素として考慮すべきだが、本稿では Ahlstrom-Vij(2013)や Bishop & Trout(2005)と同様の方針を採り、最重要と思われる信頼性に絞って議論を進める。なお、自然主義的な立場から書かれた認識論全体の邦語教科書として、また以下の内在主義の規定についても、戸田山(2002)を参照。
- (3) そもそも内在主義は認識に関わる助言を提供することを目標とするものではなく、それゆえ知性改善にも関わらないとする見解もあるが(Bishop & Trout, 2008, pp.1055-1056)、ここではあくまでも知性改善をも目標とするものとして内在主義を解釈しておく。
- (4) 以下での内在主義批判の詳細については Kornblith(2012)を参照。この著作でコーンブリスもまた最終的には二重過程理論に向かっている。
- (5) 二重過程理論およびバイアス研究についての一般向け解説として Kahneman(2011)を参照。二重過程理論の検討も含め、認知科学における合理性論争を概観した有益なサーベイとして太田・小口(2014)を参照。
- (6) 正確には、信念形成プロセスの信頼性に関して中立的な「ヒューリスティクス」と、状況次第で誤信念を生む傾向を指す「バイアス」とを区別すべきだが、本稿では便宜上「バイアス」と統一する。
- (7) ただしシステム2も完全ではなく、熟慮的な思考がミームによって汚染されているために、不合理な仕方での信念形成が行われてしまう可能性がある。そこで、脱バイアスのみならずミームの評価および浄化という手続きも必要となるという興味深い主張が Stanovich(2004, chap.7)に見られる。
- (8) そうした議論の例として山(2011)がある。なお、バイアスをめぐる心理学研究が認識上の規範や助言、ないし知性改善に直接的に結びつくとする主張として、Bishop et al.(2005)をも参照。
- (9) 教育を通じた脱バイアスには莫大なコストがかかるため実行可能性に乏しいとする主張は、とくに認識的パターナリズムに近い論者に共有されている(Ahlstrom-Vij, 2013, chap.1; Trout, 2005)。
- (10) 批判的思考力の限定性や上述のバイアス盲点など、脱バイアスを阻みかねない種々の障壁については、Lilienfeld, Ammirati & Landfield(2009)のサーベイを参照。そこでは他に、認知における個人差・文化差の問題や、脱バイアスが逆効果となってバイアスを強化する可能性などが指摘されている。
- (11) ここでの叙述は Ahlstrom-Vij(2013)および Trout(2005)の主張の主要部分を私の観点から要約したものである。なお彼らはバイアスに陥りがちな人間の認知能力を外部からの干渉によって補強するべきだと考える点では共通しているが、何点か相違もある。前者は本人の同意なしの干渉であればパターナリズム的だと捉えたうえで、認識に限定したパターナリズム擁護論を提出している。これに対し後者は本人の意志に反しなければ制度的干渉は必ずしもパターナリズム的ではないと規定したうえで、個人の福利などより一般的な領域における制度的干渉を擁護している（とはいえそこで論じられているのは「リバタリアン・パターナリズム」の実践として Sunstein & Thaler(2008)で捉えられている事例である）。
- (12) 以下の情報アクセスの制限の事例は、もともとゴールドマンの論文で扱われているものである(Goldman, 1991; cf. Ahlstrom-Vij, 2013, pp.29-32)。なお「認識的パターナリズム」という語の初期の使用も、彼のこの論文に見られる。
- (13) Ahlstrom-Vij(2013, pp.170-172)ではこうした議論に加え、予測モデルを信用させるべく使用者の態度変容を促すような認知的不協和を生じさせるという試案が提出されている。必ずしも認識に関わるものではないが、ある種のバイアスをうまく逆用して望ましい意思決定を促すという同様の方略は、Trout(2005, pp.425-433)や Sunstein et al.(2008)でも提案されている。
- (14) ここまでの応答については Trout(2005, pp.408-416)の議論を参照。

文献

- Ahlstrom-Vij, K. (2013). *Epistemic Paternalism: A Defence*, London: Palgrave Macmillan.
 Bishop, M. A. & Trout, J. D. (2005). *Epistemology and the Psychology of Human Judgment*, Oxford: Oxford University Press.

- (2008). ‘Strategic Reliabilism: a Naturalistic Approach to Epistemology’, *Philosophy Compass*, 3/5, 1049-1065.
- Cherniak, C. (1986). *Minimal Rationality*, Cambridge, Mass: MIT Press. (2009, 中村直行他訳, 『最小合理性』, 勁草書房.)
- Goldman, A. I. (1991). ‘Epistemic Paternalism: Communication Control in Law and Society’, *The Journal of Philosophy*, 88(3), 113-131.
- (1992). ‘Epistemic Folkways and Scientific Epistemology’, in H. Kornblith (Ed.), *Naturalizing Epistemology: Second Edition*, (1994, pp.291-315), Cambridge, Mass: MIT Press.
- Johansson, P., Hall, L., Sikstrom, L. & Olson, A. (2005). ‘Failure to Detect Mismatches between Intention and Outcome in a Single Task’, *Science*, 310, 116-119.
- Kahneman, D. (2011). *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux. (2012, 村井章子訳, 『ファストとスロー——あなたの意思はどのように決まるか? (上・下)』, 早川書房.)
- Kitcher, P. (1992). ‘The Naturalists Return’, *The Philosophical Review*, 101, 53-114.
- Kornblith, H. (2012). *On Reflection*, Oxford: Oxford University Press.
- Lilienfeld, S. O., Ammirati, R. & Landfield, K. (2009). ‘Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare?’, *Perspectives on Psychological Science*, 4, 390-398.
- Nisbett, R. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, N.J.: Prentice-Hall.
- 太田紘史・小口峰樹 (2014). 「思考の認知科学と合理性」, 信原幸弘他編, 『シリーズ新・心の哲学 I 認知篇』 (111-164 頁), 勁草書房.
- Pronin, E., Lin, D. & Ross, L. (2002). ‘The Bias Blind Spot: Perceptions of Bias in Self versus Others’, *Personality and Social Psychology Bulletin*, 28, 369-381.
- Sieck W. R. & Arkes, H. R. (2005). ‘The Recalcitrance of Overconfidence and Its Contribution to Decision Aid Neglect’, *Journal of Behavioral Decision Making*, 18, 29-53.
- Stanovich, K. E. (2004). *The Robot’s Rebellion: Finding Meaning in the Age of Darwin*, Chicago, IL: University of Chicago Press. (2008, 棕田直子訳, 『心は遺伝子の論理で決まるのか——二重過程モデルで見るヒトの合理性』, みすず書房.)
- Sunstein, C. R. & Thaler, R. H. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*, New Haven, CT, and London: Yale University Press. (2009, 遠藤真美訳, 『実践行動経済学——健康、富、幸福への聡明な選択』, 日経BP社.)
- 戸田山和久 (2002). 『知識の哲学』, 産業図書.
- Trout, J. D. (2005). ‘Paternalism and Cognitive Bias’, *Law and Philosophy*, 24, 393-434.
- 山祐嗣 (2011). 「批判的思考と適応」, 楠見孝他編, 『批判的思考力を育む——学士力と社会人基礎力の基盤形成』 (66-86 頁), 有斐閣.

〔関西大学准教授・哲学〕