自己抹消的な道徳理論の問題点は（あるとすれば）何か

秋葉剛史（千葉大学）

要旨

道徳理論は、我々がどのように行為すべきかについての指示を与える。そのため、日々の道徳的な意思決定において、自分の信じる道徳理論がどの行為を勧めているかを考えることは、我々にとってごく自然なことのように見える。しかしある種の道徳理論は、こうした直接的な意思決定法をそれ自身の基準から禁じる。なぜなら、その種の道徳理論に従おうとした場合、我々はかえってそうできなく（しにくく）なるからだ。このような一見逆説的な状況におかれた道徳理論は、しばしば「自己抹消的」と呼ばれる。そして少なからぬ論者によると、ある道徳理論が自己抹消的であるという事実は、場合によるとその理論を拒否するのに十分なだけの深刻な難点になる。しかし本稿ではこのような見方に対し、自己抹消性は少なくともそれ自体では、ある道徳理論を拒否するための説得的理由にはならないと主張する。そのため本稿では、従来この種の理論に向けられてきたいくつかの批判を検討する。より具体的に言うと、ここで考察されるのは、行為指導性の欠如、実践的指示の不確定性、心理的調和、望ましい道徳的熟慮、自己欺瞞、といった論点に関わる諸批判である。本稿の議論を通じて、自己抹消的な道徳理論に従来投げかけられてきたこれらの批判には十分応答可能であることが示される。

What are the Problems (if any) of Self-Effacing Moral Theories?

Moral theories generally tell us the right course of action. They deliver practical advices (or a series of commands) about how we should act, so it seems natural for those who sincerely accept a given moral theory to try to figure out, in each case where some practical decision is needed, what the theory recommends, and act accordingly. However, this is not what a moral theory always tells us to do. Certain moral theories in fact tell their followers not to consult them in daily decision-making. The reason for this is simply that if we consciously intend to act as the theories recommend, it would become difficult or even impossible for us to act in that manner. In recent literature, moral theories that satisfy this condition are called *self-effacing*, and have attracted some attention. Although quite a few authors seem to endorse the view that this character of self-effacement makes a moral theory highly problematic (or even unacceptable), in this paper I shall argue that this view is ungrounded. To do this, I will critically examine various objections to self-effacing moral theories found in the literature that concern the following points, respectively: lack of action-guidingness; threat of undermining psychological harmony and the desirable form of moral deliberation; an absurd requirement to have mutually contradicting beliefs; and an invitation to a kind of self-deception. It will be argued that none of these objections constitutes a serious threat to self-effacing moral theories as such.

道徳理論の主要な目的は、どのようなタイプの行為が道徳的に正しいかを明らかにすることだ。例えば功利主義なら、「幸福を最大化する」行為こそが正しいと主張し、カント的義務論なら、「定言命法と合致した」行為こそが正しいと主張するだろう。各々の道徳理論はこのような仕方で、どういったタイプの行為をなすべきかを我々に指示するわけである。そして、一般に道徳理論がこうした実践的な指示を発するものだとすれば、ある道徳理論を受け入れるときに我々がどのような態度で実践に臨めばよいかは自ずと決まってくるように思われる。すなわち我々は、いくつかの行為の選択肢があるときにはいつも、自分の信じる理論が正しいとする選択肢はどれかを割り出しそれを実行に移すよう心がければよい、とそのように思われる。こうした意思決定の方法は、ある道徳理論を信じる者にとってごく自然なものだろうし、当の道徳理論によって勧められるものでもあるだろう。

だが実のところ、話はそう単純ではない。なぜならある種の道徳理論に関しては、その理論が正しいと言う仕方で行為しようとすると我々はかえってそう行為できなく（しにくく）なる、ということが成り立つからだ。このことは、以下で見るように様々な事情に起因しうるのだが、いずれにしても、このような状況におかれた道徳理論は我々に対して、その理論の主張を意識しながら行為することを（それ自身の尺度から）禁じることになる。言ってみればそうした理論は、行為に臨む我々に「私の主張を気にかけてはならない」と命じ、我々の思考から自らすすんで消え去ろうとするのだ。本稿で焦点を当てたいのは、しばしば「自己抹消的 self-effacing」と呼ばれる、一見奇妙なこの種の道徳理論である。

一部の論者が主張するところでは、自己抹消的な道徳理論は単に奇妙なだけでなく、より実質的な困難を抱えている。彼らによると、ある道徳理論が自己抹消的であるという事実は、場合によってはその理論の拒否を正当化するのに十分なほどの深刻な難点になるのだ。じじつ関連文献の多くは、自己抹消性という性質が一般に道徳理論を受け入れがたいものにするという見方はほぼ前提とした上で、ある特定の道徳理論（e.g. 帰結主義や徳倫理の各バージョン）が実際に自己抹消的かどうかを検討することに主な労力を割いている[1]。

しかしこのような見方は、どの程度の根拠をもつのだろうか。自己抹消性という性質が道徳理論にもたらすという問題とは具体的にどのようなものであり、それらは本当にこの種の理論にとって深刻な難点になるのだろうか —— 本稿で考えてみたいのは、従来の研究であまり主題的に論じられてこなかったこれらの問題だ[2]。以下で見るように、実は自己抹消的な道徳理論は決して珍しいものではないから、これらの問題は、一般に我々が道徳理論とどう関わるべきかを考える上でも重要である。そして結論から言うと、（以上の流れから予想される通り）私は以下で、自己抹消性に関して提起されてきた諸問題は基本的に[3]、ある道徳理論を拒否するための説得的な理由にはならないと主張する（もちろんその理論が他の理由から受け入れがたいという可能性は残るが）。自己抹消性が困難をもたらすという印象の大部分は、道徳理論が我々の実践を導く際の多様な仕方を十分に考慮しないことから生じている、ということが本稿の議論を通じて明らかになるだろう。

1　例えば徳倫理については Keller 2007; Martinez 2011; Pettigrove 2011、帰結主義と義務論については Williams 1973: 128; Stocker 1976: 455f; Smith 2001 を参照。
2　ただしこれらの点にふれた議論がないわけではない。例えば Wiland 2007; Cox 2012 を参照。
3　ここで「基本的に」と留保した理由については、本稿の末尾（第 5 節）を参照のこと。

以下では、まず自己抹消性という特徴について改めて説明した後で（§1）、自己抹消的な道徳理論に対し最初に向けられる「行為指導性の欠如」という懸念をとり上げそれに応答する（§2）。続いて、その応答に対してさらに提示されうる三つの批判と（§3）、「自己欺瞞」という点に関連する一つの懸念（§4）に注目し、自己抹消的な理論をそれらから弁護する。

1. 道徳理論の自己抹消性

自己抹消性について改めて説明することから始めよう。後の議論の提示しやすさのため、ここでは特にD・パーフィットによる特徴づけ方（Parfit 1984: ch.1）を参考にする。前述のように、各々の道徳理論はある特定のタイプの行為を正しいものとし、そのタイプの行為がなされることを我々に要求する。これは言い換えると、一般に道徳理論は、あるタイプの事態を、我々が自らの行為を通じて実現すべき目標として定めるということだ。以下では、任意の道徳理論を「T」によって表し、道徳理論Tが実現されるべき目標として我々に課す事態を（「Aim of T」という意味をこめ）「A^T」と表そう。例えばTが功利主義なら、それが我々に与える目標A^Tは「幸福が最大になること」である。また、Tがある十種類の行為を禁じるような神の命令理論なら、その目標A^Tは「それら十種類の行為がなされないこと」である。（ここで「目標」は禁忌や義務に関する事態なども含む広い意味で使われていることに注意。この用法は以下でも同様である。）

さて道徳理論が一般にこのようなものだとすると、ある道徳理論Tを真なるものとして信じることのうちには、その理論の目標A^Tに対し一定の規範的ないし動因的conativeな態度をとることが含まれることになる。すなわちTを信じる者は、A^Tの実現は望ましい（A^Tは実現されるべきだ）という評価的（規範的）判断を受け入れ、A^Tを実現しようという意志やそれを実現したいという欲求などをもつ者、つまり、A^Tの実現に対し少なくとも潜在的な動機をもつ者である[4]。しかしこの潜在的動機が、Tを信じる者の日常的実践において顕在的な動機としてはたらくべきかと言えば一概にそうは言えない。というのも導入部で述べたように、ある種の理論Tにおいては、その目標A^Tを実現しようという意図に導かれて行為することがかえってA^Tの実現の妨げになる、ということがあるからだ。

本稿で関心を寄せるこの自己抹消的な状況を、もう少し丁寧に特徴づけておこう。そのために注意しておきたい点は三つある。第一に、自己抹消性を特徴づける上で重要になるのは、我々が一人称的な観点から行う意思決定の過程、つまり、自分はいかに行為すべきかを（通常いくつかの選択肢を前にして）考察し決定するという過程だ。以下ではこのような意思決定の過程を、短く「熟慮deliberation」とも表すことにする。第二に、この意味での熟慮を進める仕方には様々なものがありうるが、自己抹消性と特に関係するのは、何らかの道徳理論Tへの信念が顕在的・主導的な動機となって進められるような熟慮である。すなわちそれは、「当の理論Tが定める目標A^Tを、そのつど与えられた行為の選択肢と直接引き比べ、それらの選択肢の中でどれがA^Tの実現に最も寄与するか（あるいはその実現であるか）を割り出す」という熟慮の方法だ。以下では、

4　本稿ではAdams 1976: 467; Parfit 1984: 8に倣い、「動機motive」という語を、何らかの行為を生み出すものとしての心的な状態・態度を表す総称として用いる。なおここでは、何らかの道徳理論への信念（道徳的判断）と対応した動機づけの間には単に偶然的ではない結びつきがあると想定しているが、これは必ずしも、いわゆる「動機内在主義」をとることではない。

このような熟慮を「A^Tを直接的目標とした熟慮」と呼び、また、このタイプの熟慮に基づいて行為することを「A^Tを直接的目標として行為する」こととも表すとしよう。そして最後に、三つ目の注意点として、ある目標の実現にとってある熟慮の方法が適しているかどうかは、ほとんどの場合、我々について現に成り立つ様々な経験的事実 ── 我々の心理的・身体的な特徴や、社会的・生物学的な環境など ── に依存する。以下では、道徳理論が与える目標の実現に影響するこれらの経験的事実の総体を考え、それを「現実の経験的条件」と呼ぶことにする。

以上をふまえると、自己抹消性は次のように定義できる。すなわち、道徳理論Tが自己抹消的であるのは、Tが与える目標A^Tに関し、現実の経験的条件の下で、「我々がA^Tを直接的目標として行為すると、A^Tの実現はより困難になる」という条件が成り立つときである[5]。ただしここで「困難になる」は、長期的・全体的な傾向を表す。すなわちこの条件が意味しているのは、もし我々がA^Tを直接的目標とした熟慮に基づいて日々の実践を行うなら、A^Tはそうでない場合と比べ全体としてより実現されにくくなる、ということだ[6]。この定義が上の直感的説明と合致していることは、次のような考察からわかるだろう。一般に道徳理論Tは、A^Tを実現されるべき目標として我々に課す。それゆえTによれば、もしあるタイプの行為や習慣・傾向性がA^Tの実現を阻害するなら、我々はそれを行ったり身につけたりすべきでない。だが上の定義条件が成り立つ場合、「A^Tを直接的目標とした熟慮」はまさにそのような阻害要因になってしまう。よってTはそれ自身の基準に照らし、このタイプの熟慮を日常的に行わないよう我々に命じることになる。

一つ例を挙げておくと（他の例は次節以降で）、古典的な行為功利主義はしばしばこの意味で自己抹消的な道徳理論の典型だとされる[7]。周知のように、この立場が我々に与える目標は「関係者の幸福が最大になること」だ。しかし功利主義者自身も認めるように、我々が実際にこの事態を直接的な目標として、つまり、熟慮のたびに幸福計算を行うという仕方で日々の実践に臨む場合、この事態が実現される見込みは逆に低くなると考えられる。というのもその場合、我々は計算に時間をとられて行為の機を逸したり、計算自体に苦痛を感じたり、時間的・情報的な制約や各種バイアスの下で計算することによってかえって大きく選択を誤ったり、他の行為者との協調行動をとりにくくなったり、等々の事情から、幸福の最大化を直接目指さない場合よりむしろ少なくしか幸福を実現できないということになりがちだからだ。こういったわけで行為功利主義は、それ自身の尺度に基づき、この理論への信念を主導的な動機として行為しないよう我々に命じることになる[8]。

5　ただし本稿の末尾の第5節でも触れるように、ある種の道徳理論においては、この定義条件（括弧内の部分）が現実の経験的条件に依存せずに成り立つと考えられる。

6　この意味での自己抹消性は「いつ誰にとって」が問題になる相対的な特徴であり（Parfit 1984: 5）、かつ程度差を許容する特徴である（Martinez 2011: 279）。本稿で「自己抹消的」として念頭におくのは、現在の我々（標準的な人間主体）にとって、問題の熟慮を行った場合A^Tの実現されやすさが無視できない程度に低下するようなケースである。（これは曖昧だが、以下の議論の害にはならないだろう。）

7　Mill 1998: 69f; Sidgwick 1966: 405; Adams 1976: 467; Hare 1981: chs. 2-3; Railton 1984: 153-4; Wiland 2007.

8　では反対に、自己抹消的でない道徳理論にはどんなものがあるか。一部の論者はその例として、ある種の規則功利主義（Wiland 2007: 276）や、道徳法則についての意識から行為せよと説く（カント的な）義務論を挙げている（Keller 2007: 223）。ただし本稿の末尾で触れる問題も参照のこと。

2. 自己抹消性と行為指導性

さてそれでは、以上の意味で自己抹消的であるような道徳理論にはどのような問題があるだろうか[9]。この点で真っ先に浮かぶのは、「行為指導性 action guidingness」に関する次のような不満かもしれない[10]。一般に道徳理論は、我々の行為実践を一定の方向へと導くことができなくてはならない。物理学や化学などの理論とは異なり、道徳理論はその本性からして、いかに行為すべきかを我々に教えることを使命とした実践理論だからである。そしてある道徳理論が真に行為指導的であるためには、それは、我々が何らかの（ときに困難な）選択に迫られる具体的な意思決定の場面で、我々の決定を主導するような実質的根拠や観点を提供できねばならない。しかし自己抹消的な道徳理論は、明らかにこの要件を満たさない。それが意思決定の場面で与えることができるのは、せいぜい、「私の言うことを考慮するな」という無責任で投げやりな指示だけだからである。ゆえにこの種の道徳理論は、そもそも道徳理論としての存在意義をもたない。

だがこの批判は、（少なくともそれ自体としては）深刻ではない。この批判への応答としてはまず、それが依拠する大前提、すなわち「道徳理論が受容可能であるためには行為指導的でなくてはならない」という前提を斥ける道がありうる（cf. Parfit 1984: §17）。これは私見では必ずしも無謀な道ではないが、この路線を多少なりとも満足な仕方で擁護するには、かなり立ち入った考察 ── とりわけ、道徳理論の真理（適合）性と実践性についてのメタ倫理学的な考察 ── が必要であり、それに深入りする余裕は本稿にはないので、ここではより寛容な応答を探ることにしよう。以下で考察したいのは、道徳理論に行為指導性は必要だという前提は認めた上で、「自己抹消的な道徳理論であっても実際は十分に我々の行為を導きうる」と応じる道だ。この応答は次のように提示できる[11]。

ある道徳理論 T が自己抹消的であるとしよう。このとき前節の定義より、我々が A^T を直接的目標とした熟慮を日々の実践の中で行うことは、現実の経験的条件の下ではむしろ A^T を実現されにくくする。つまりこのタイプの熟慮は、A^T の実現に貢献しない。だがここで重要なのは、このタイプの熟慮が A^T の実現に貢献しないことが経験的条件によって決まるなら、同じ条件によって、他の可能なタイプの熟慮がどのくらい貢献するかも決まるはずであり、よって特に、それらのうちどのタイプが A^T の実現に最適であるかも決まるはずだという点である[12]。そしてそうだとすれば、理論 T はまさにそのタイプの熟慮を、A^T を直接的目標とした熟慮に代わる「代替的」な熟慮方法として我々に勧めることになると考えられる。

もちろん、ここで言う最適性はあくまで相対評価によるものであり、あるタイプの代替的熟慮

9　念のため注意すると、自己抹消的な理論は通常の意味で「自己矛盾」しているわけではない。自己矛盾した理論からは何らかの命題とその否定の両方が帰結するが、自己抹消的な道徳理論 T から「A^T は実現されるべきだ」という主張に加えてその否定が帰結する必要はないからである。また、自己抹消的理論が発するものとして本段落で挙げた「私の言うことを考慮するな」という指示は、一見すると遂行レベルでの矛盾をはらんだダブルバインド的指示（それに従うと従わなかったことになり、逆も然り）にもみえるがこれは単なる見かけにすぎない。なぜなら、T は自分の言うことのすべてを考慮するなと指示しているわけではなく、あくまで日常的な意思決定の方法について指図しているだけだからである。

10　Williams 1973: 128; Jackson 1991: 466; Wiland 2008: 393; Martinez 2011: 280; Pettigrove 2011: 192.

11　Cf. Railton 1984: 152f; Parfit 1984: 29; Cox 2006: 508; Wiland 2007: 279f.

12　こうした「最適」な熟慮がただ一つに決まらない（極大的によい熟慮が複数ある）ケースも論理的には不可能ではないだろう。だが、前述した習得コスト等までを考慮に入れるとそうしたケースはあまりなさそうなので、以下では最適な熟慮が一つに決まる場合をもっぱら念頭におく。

が A^T の実現に最適であるとしても、それは必ずしも A^T の「完全な」実現を保証するわけではない。とりわけ、我々がある熟慮を日常的に行えるようになるにはそれに適した傾向性を身につける必要があり、その作業には相応のコストが伴うから（cf. Wiland 2008: 390）、ある熟慮が A^T の実現にもたらす貢献度を考える際にはそのコストを差し引くといったことも必要になる。しかしそれらを総合した上で、あるタイプの熟慮を我々が行う場合に A^T が最もよく実現されると言えるなら、T は具体的場面での意思決定方法としてまさにそのタイプの熟慮を勧めていると言ってよい。つまり T は、たとえ熟慮において参照すべき目標を直接与えるのではないとしても、代わりに何を参照して熟慮すべきかを（経験的条件と共同で）決定することで、我々の日々の実践を間接的に導きうるのだ。

実際、道徳理論がこうした間接的な仕方で我々を導きうるという点は、早くから功利主義者たちによって強調されてきた[13]。例えばJ･S･ミルによると、我々の日常的な道徳的思考は、「幸福の最大化」という本来の目標によってではなく、むしろ常識道徳を構成する諸々の単純な規則（嘘をつくな、等々）によって導かれるべきである。なぜならそれらの規則は、人類が長い実験の歴史を通じて、その順守が結果として我々の幸福の最大化を生み出すことを確かめてきた当のものだからである（Mill 1998: 69-71）。つまり功利主義は、我々がその実現を見越して日々の熟慮方法を選ぶべきところの目標（＝幸福の最大化）を特定することで、間接的に、日常的な我々の意思決定に導きを与えるのだ。

あるいは別の例として、徳倫理のあるバージョン、すなわち「そのつどの状況で有徳な行為者がそうするであろう仕方で行為すること」を目標として我々に課す立場を考えてみよう（Hursthouse 1999: 28）。そしていま仮の話として、我々の多くは（残念なことに）そのつどの状況で発揮されるべき徳をほとんど備えておらず、有徳者のように振る舞おうとすると混乱してしまいかえって有徳者とは程遠い行動に出てしまうが、非常に鋭敏な羞恥心をもっているため、何をしたら人に笑われずに済むかを考えるとおおむね本物の有徳者と同じ判断に達することができる、と想定しよう。そうするとこの場合、「人に笑われない行為を選ぶ」という熟慮方針が、上の目標の実現にとって最適となるかもしれない。

賢明な読者は気づいているように、ある実践理論がこうした間接的な仕方で導きを与えるというのは、実は道徳理論に限らずごくありふれた現象だ。我々はしばしば、ある目標を実現するにはそれを直接追求するよりも何か別のことを目指した方がよいということを知っており、その知識に基づいて日々の実践に臨んでいる。例えば、我々各人にとってよいこととは「最大限の快楽を得ること」だと考える快楽主義者も、この目標の実現のためには快楽とは別のこと（e.g. そのつどの活動の独自性を認識すること）に当面の意識を集中する方がよいことを知っている[14]。あるいは、外国語を流暢に話すこと、感じのよい人として他人に映ること、眠りにつくこと、テニスの試合に勝つこと、ゴルフクラブの芯にボールを当てること、等々の目標 ── その意識的な追求がよい結果を生まないことが経験的に知られているような目標 ── に関しても、その実現にとって最適なアプローチがそれぞれの場合でどのようなものかは多かれ少なかれ正確に知られているだ

13　前註 7 の諸文献を参照。なかでもヘアの二層功利主義（Hare 1981）はよく知られているだろう。
14　快楽主義については Parfit 1984: 6; Railton 1984: 140-6; Martinez 2011: 279 を、また以下の他の例については、Railton 1984: 144, 154; Wiland 2007: 291 を参照のこと。

ろう（各種のノウハウ本はそういう知見であふれている）。そしてこれらの目標を実現しようとする場合、我々はまさにその種の知識を用いて、いかに行為すべきかの指針を得ることができている。だとすれば、ある道徳理論から（その自己抹消性のゆえに）直接的な導きを得られないとしても、そこに特段の問題はないのではないだろうか。

3. 間接的な導きにまつわる懸念

前節の議論によれば、道徳理論は自己抹消的であっても十分に行為指導的でありうる。なぜなら、たとえある理論 T が自己抹消的である ── ゆえにその目標 A^T を直接追求することは T 自体により禁じられている ── としても、T を信じる主体は、現に成り立つ経験的条件の下で目標 A^T の実現に最も貢献する熟慮の方法を選び出すことで、日々の意思決定のための指針を得ることができるからだ。しかし、このような仕方で自己抹消的な道徳理論に行為指導性を認めることに対しては、いくつかの批判がありうる。本節ではそうした批判の中から、関連文献に見られる三つのものをとり上げ検討しよう。

指示内容が確定しない

一つ目のありうる批判は、自己抹消的な道徳理論は我々に確定した指示を与えてくれない、というものだ[15]。我々が具体的場面で自己抹消的な理論から引き出すことのできる指示は、必然的に代替的熟慮の選択に依存している。つまりその指示内容は、理論が与える目標 A^T をそのまま参照できる（自己抹消的でない理論の）場合とは違って、どのような代替的熟慮の方法を我々が先に選んでおいたかに応じて決まる。しかしここには、誤りの余地がある。これまでの議論ではあたかも、ある目標の実現にとって最適な熟慮が現実の経験的条件によって客観的に決まるなら、我々は自動的にそれを正しく選び出せるかのように語られていた。だがもちろん、実際にはそんな保証はない。つまり、A^T の実現に最適だと思って我々が選んだ熟慮方法が実はそうでなかったという可能性は常に残るし、さらに言うとこの可能性は、問題の最適性が経験的条件に依存する仕方がきわめて複雑であることを思えばかなり現実的な懸念でもあるだろう。しかしそうだとすると、我々は、たとえ T により最終目標は与えられていたとしても、個々の具体的な場面で何を目指すように T が指示しているのかについては、実は永久に確信をもてないことになる。つまり T を頼りにしても、実際の行為の場面では何ら確定的な指示を得られないことになる。これでは到底、理論 T が我々の行為に導きを与えているとは言えないだろう。

だがこの批判は説得的でない。指摘すべき点は二つある。第一に、たしかに我々は全知ではないから、最適な熟慮方法がどれかについて思い違いをする可能性はゼロにはならないけれど、そのことから、我々がこの点に関してもちうる意見のすべてが等しく無根拠だとするのは誤りだ。我々は（ミルが示唆していたように）様々な経験や科学的知識の蓄積をもとに、ある目標にとって最適だと信じることがより理にかなっているような熟慮方法を割り出していけるし、その推測の精度を高めてもいける。つまりたとえ不可謬の確信ではないとしても、我々は理論 T が何を指

15　これを示唆するものとしては、Stocker 1976: 463, Wiland 2008: 390f を参照。

示しているかに関して十分に合理的な信念をもてるわけだ。さらに指摘の二点目として、こうした合理性以上のものを我々が望むべきでないと考えるべき理由がある。上の批判では、理論の指示内容を確実には知ることができないという点が自己抹消的な理論に特有の問題であるかのように論じられていたが、実は同様の問題は自己抹消的でない理論にも生じうる[16]。つまり、ある目標を熟慮の際に直接参照できる場合でも、与えられた行為の選択肢のうちのどれがその実現にとって最適かに関して我々が誤る可能性は十分にある。この可能性は、一般に道徳理論は十分に広い適用範囲をもった――それゆえ抽象的な――目標を与えようとするという事情からしてもある程度避けられないものだろう。（例えば、おそらく自己抹消的でないカント的義務論の場合でも、具体的な選択肢の中でどれが定言命法にかなっているかを確定することが容易でないケースはあるだろう。）最善の選択肢を選ぶという課題は、選択の対象が熟慮タイプではなく行為トークンであってもすでに十分難しいのであって、自己抹消的でない理論の場合も我々が望めるのはせいぜい合理的な選択だけだ。したがって自己抹消的な理論の場合にも、我々は十分な合理性をもってその指示内容を推測できればそれで満足すべきだろう。

何らかの望ましい事柄と衝突する

続いて二つ目の批判は、自己抹消的な理論による導きは何らかの望ましい事柄を損ねてしまう、というものだ。その理由としては、関連する二つのものが考えられる。

第一の理由は、我々の生における「心理的調和」という論点に訴えるものだ[17]。一般的に言って、我々自身の中で価値判断と動機づけが一致している状態は望ましいと言えるだろう。M・ストッカーが言うように、「我々は最低限、自分が主要な価値を認めているものによって動機づけられ、また自分の主要な動機が向かうものに価値を認めるべきだ。〔…〕このような調和は、よき生の一つのしるしなのである」（Stocker 1976: 454）[18]。だが――と反論は始まる――もし我々が自己抹消的な道徳理論によって導かれるべきだとするなら、こうした調和を手に入れることは難しくなる。なぜならその場合、我々はあるタイプの事態（A^T）に価値を認めながらも、それを実現しようという動機によって導かれてはならず、むしろそれとは別のタイプの（代替的熟慮において目指される）事態の実現への動機が支配的になるよう自らを律さねばならないからだ。つまり自己抹消的理論による間接的な導きは、我々の価値体系と動機体系の間に分裂を生じさせ、よき生の条件である心理的調和を実現困難にしてしまう恐れがある。

これに関連する第二の理由は、「道徳的熟慮」というものに関する次のような（一見もっともな）見方から生じる[19]。道徳的熟慮は本来、そのつどの状況で事物が特定の道徳的価値をもつことを可能にし、その価値を基礎づける特性――幸福を最大にする、有徳者の振る舞いと合致する等の、各道徳理論がそれぞれに価値構成的だと主張する特性――に対する注視と応答という仕方で行われるべきだ。というのも、ある熟慮はまさにその種の価値構成的な特徴と関係することによって

16　この点は、本稿での議論とは焦点の当て方は異なるが、Wiland 2008: 389f でも指摘されている。
17　Stocker 1976: 453-5, 461-3; Cox 2006: 508; Keller 2007: 222; Martinez 2011: 281; Pettigrove 2011: 193.
18　こうした調和の望ましさはその不在のケース、すなわち、自分が価値を認めないもの（薬物、児童ポルノ、等々）を欲してしまったり、自分が価値を認めるもの（健康、社会的活動、等々）を欲せなかったり、といったケースと比べてみれば納得しやすいだろう。
19　Pettigrove 2011: 193.　また Cox 2006: 510 にもこれと関連する議論がある。

道徳的熟慮になるのだし、また本来の道徳的熟慮は、そのつどの状況の中で行為の「理由」を構成するような特徴を見出し、その状況でなぜある振る舞いが正しいかの洞察をもたらすようなものであるべきだからである。しかし ── ここからが反論だ ── 自己抹消的な理論によって導かれる主体の熟慮は、こうした本来の姿とはかけ離れたものになってしまう。なぜなら自己抹消性の定義より、そのような主体が日々の実践で行う代替的熟慮は、道徳的価値を構成すると彼が信じる特性 ── 当の理論の目標 A^T に関わる特性 ── とは別の何かに目を向けるものとならざるをえないからだ。つまりその主体は日々の熟慮において、彼に行為の理由を与えるはずの事柄を洞察するどころか、絶えずそこから目を逸らしあさっての方向ばかりを気にかけることになってしまう。

しかし以上の批判も、自己抹消的な道徳理論にとって痛手にはならない。まず一つ目の理由に関して言うと、ある主体が価値づける事態と彼の動機づけが向かう事態が単に「同じでない」というだけで、そこに何か避けるべき不調和があるかのように語るのは明らかに大げさだ。もちろん、それらが同一でないケースの中には実際に望ましくない ── ストッカーの言葉では「分裂症的」な ── ケースもあるだろうが（註 18 を参照）、すべてのケースがそうである必要はない。とりわけ、自己抹消的な理論によって導かれる主体の場合のように、主体の動機づけが向かうタイプの事態が、彼が価値づけるタイプの事態の実現頻度や確率を高めるようなケースでは、その主体は前者のタイプの事態に対して少なくとも道具的な価値を認めることができる。つまりこのような主体においては、価値判断と動機づけの体系は無理なく調和するのであり、そこに避けるべき有害な衝突はない[20]。

また二つ目の理由にも次のように応答できる。ある自己抹消的な理論 T によって間接的に導かれている主体を考えよう。そして、この主体が目標 A^T の実現のため身につけている代替的熟慮は、A^T とは別のタイプのある事態「E^T」（「Ersatz aim of T」の意）を直接的目標として行われるようなものだとしよう。（例えば T が行為功利主義なら、その代替的目標 E^T はミルが言うように「常識道徳の諸規則が順守されること」になるかもしれない。）そうするとこの主体は、たしかに日々の熟慮の中で A^T ではなく E^T の実現をもっぱら気にかけることになるが、このことは必ずしも、彼が本当に重要な（価値構成的で理由供給的な）事柄から目を背けているということを意味しない。というのも、この主体はまさに A^T の実現を重視するからこそ E^T の実現を気にかけるのであって、彼の E^T への関心はあくまで A^T への関心から派生したものである ── ゆえに、もし A^T の実現に最善なのは別の熟慮方法だと信じていたなら彼は E^T への関心をもたなかった ── ということは十分ありうるからだ。こうした場合に主体が決して A^T を軽視しているわけではないことは、次の類比からもわかるだろう。ある戦闘機のパイロットが遠方の標的を射撃する際、その最善の方法は、当の標的を直接見るのではなく、機内にあるモニターを見てそこに映る枠の中に（標的を表す）印が入った状態で発射ボタンを押すことだとしよう。この場合、機内のモニターに最大限の注意を集中しているパイロットは、たとえ標的に直接目を向けてはいないとしても、それを軽視しているわけでは当然ない（むしろ明らかに重視している）。

20　ただしここで、この主体が価値づける事態の実現にとって最善の手段が、彼の忌み嫌う何らかの事態の実現を目指すことだったらどうするのか、という懸念が生じるかもしれない。これに対しては二つの答えが考えられる。第一に、自分が忌み嫌う事態の実現を目指すことは相応の心理的負担になるから、そうした手段が最善となる見込みはあまり高くない。また第二に、たとえそれが現時点での経験的条件の下で最善になるとしても、自己抹消的理論はそれを勧めないとも解釈できる（以下の第 4 節後半を参照）。

もっともここで批判者はさらに、本来の道徳的熟慮において求められているのは、道徳的に重要な特性の単なる「重視」ではなく、その種の特徴に対するリアルタイムの「応答」だ、という点を再度強調するかもしれない[21]。だが、自己抹消的な理論の支持者がこの主張を認める必要はないだろう。第一に、この主張は実質的に論点先取である疑いが濃厚だ。そもそも自己抹消的な道徳理論とは、それが道徳的に価値ありとする事態（A^T）を我々が日々の実践で直接意識すべきでないような理論のことだった。しかしここで批判者は単に、我々はそういう事態を直接意識すべきだと言い張っているだけのように見える。また第二に、上の批判者の主張は、道徳的熟慮や判断の本性に関するある特定のメタ倫理学的立場を前提しているように思われる。すなわち前提されているのは、道徳的熟慮や判断をある種の知覚に似た心のはたらき、つまり状況内の特性の直接的認知の一種として捉える立場だ（cf. McDowell 1998: ch.7）。言うまでもなくこのような立場は（見込みがない訳ではないとしても）かなり実質的なものであり、上の主張もその分問題含みなものになる。

不可能なことを要求する

最後に、三つ目の可能な批判は、自己抹消的な道徳理論による間接的な導きが機能するためにはある不可能事が達成されねばならない、というものだ。その不可能事とは、同一の主体が同時に矛盾した信念をもつことである。この点は、M・スミスのものを簡略化した次のような例を通じて明らかにできる（cf. Smith 2001: §2）。道徳理論 T の例として通常の功利主義を考えよう。この T は、もっぱら行為者中立的（ないし不偏的）な観点から各人の幸福に着目し、世界の可能的諸状況の望ましさをその合計の大きさに基づき順序づける。それゆえ T によれば、「誰のものであれ幸福の総和が最大になること」が目標 A^T である。しかし（実際にありそうな話として）、T は自己抹消的であり、現実の経験的条件の下でこの目標が最もよく達成されるのは、我々が T そのものではなく行為者非中立的な順序づけ体系、つまり、身近な人の幸福をそうでない人の幸福より重く見積もるような（身内びいき的に偏った）体系に基づいて熟慮する場合である、と想定しよう。そうすると、T を信じる主体が身につけるべき代替的熟慮は、「自分に近い者の幸福が最大になること」を目標（E^T）としたものになる。さてその上でいま、T を信じるあなたは、自分の行為（e.g. 重い荷物をもって運んであげること）によって次の二つの状況のどちらか一方だけを現実化できるとしよう：①あなたの友人の幸福が 5 単位上昇する；②まったく見知らぬ人の幸福が 6 単位上昇する。このときあなたは、この二つの状況のうちどちらが望ましい（より大きな価値をもつ）と判断すべきだろうか。一方で、T を信じる（公平無私な）者としてのあなたは「状況①＜状況②」と判断すべきだが、他方で、代替的熟慮に従って行為する（身内びいきする）者としてのあなたは「状況①＞状況②」と判断すべきだ。しかし明らかに、あなたは同時にこれらの判断に同意することはできない。こうして、自己抹消的な理論による間接的な導きはある不可能事を要求することになってしまう[22]。

だがこの批判も説得的でない。明らかにいまの議論では、あるタイプの熟慮をひとが実行する

21　Pettigrove 2011: 192; Martinez 2011: 280.

22　ここでの問題を、「T への信念」と「T が求める傾向性」の間の衝突から生じるものだとすれば、以下の註 24 と 28 で触れる批判は、この同じ衝突を逆の角度から問題視するものだと言える。

ときには、それに対応した命題を「本気で」信じていなくてはならないと想定されている。上の例で言えば、問題の主体が代替的熟慮に従って状況①の実現を状況②の実現より優先することから、彼は二つの状況の価値について「状況①＞状況②」と判断しなくてはならない、とされていたわけである。しかし実際には、我々はある命題を本気で真と信じることなしにもそれを実践上の前提 presupposition として採用することができ、その前提に基づく意思決定を行うことができる。例えばある会社の苦情処理係は、「カスタマーが言うことはすべて正しい」という命題を（真とは信じないまま）実践的前提として採用することで、自分の業務を円滑に進められるかもしれない（Sainsbury 2011: 140）。あるいは、心についての厳格な消去主義者は、心的語彙が実際には系統的に誤っていると確信しつつも、実践的には（例えば「悲しんでいる人を慰めよ」と命じられた場面では）その種の語彙を用いてどう行為するかを決められるだろう。また同様のことは、プトレマイオス天文学の天球仮説、ニュートン力学、理想気体の法則などを（厳密には偽だと信じつつも）用いて実践的な意思決定を行うようなケースにも言えるだろう。これらのケースが示すように、一般に我々がある熟慮方法をとるために、それに対応した命題を信じている必要はない。それゆえ、ある道徳理論 T の判定とは矛盾した行いを勧めるような熟慮を行う主体（上の功利主義者のように）も、T と矛盾した信念をもつ必要はない。

4. 自己欺瞞を命じるのではないかという懸念

前節では、「自己抹消的な道徳理論は間接的に我々の実践を導きうる」という主張に対するいくつかの批判を検討し、それらには十分応答可能であることを示した。しかしこの種の道徳理論への批判の余地はまだ残っている。本節で検討したいのは、自己抹消的な道徳理論はそれを信じる主体にある種の「自己欺瞞」を命じることになるのではないか、という懸念だ[23]。この懸念の内容は、少し長くなるが以下のような考察により与えられる。

ある自己抹消的な道徳理論 T が、ある主体の実践を導いているとしよう。我々はこれまで、そのような主体は、A^T の実現のための代替的熟慮を身につけているだけでなく、T を信じ続けていると暗に想定してきた。つまりその主体は、代替的熟慮の傾向性と同時に「T は真だ」という信念をもっている、と（前節の第二と第三の批判はまさにこの想定から生じていた）。だがもしかすると、これは T が誰かの行為を導く際の最良の仕方ではないかもしれない。もともとの選ばれ方からして、いま問題の代替的熟慮は、T の目標 A^T とは異なる ── そしてときにそれと食い違う ── 何らかの事態（E^T）を直接的な目標として進められるものだ。しかし第 1 節でも見たように、一般に T への信念をもつ主体は、少なくとも A^T の実現への潜在的な動機をもつ。それゆえ、問題の代替的熟慮を行うときに主体は、T への信念をいったん脇におき、その動機的な効力をいわば「割り引く」といったことをしなくてはならない。だがこれは、もっぱら代替的熟慮にいそしむべき主体にとっては余計な作業ではないだろうか。この主体はむしろ、A^T に特に価値はなく、この事態が実現されようとされまいと重要な違いはないと思っているとき、つまり、T を真ではなく偽だと信じているときの方が、代替的熟慮をうまく進められる ── 結果、本来の目標 A^T の実

23　Parfit 1984: 42; Cox 2012: §5.

現にもより近づける ―― のではないだろうか。

実際、この考えを支持するいくつかの根拠がある。その一つ目は、集中・注意力の問題だ。何らかの比較的単純なタイプの代替的熟慮 ―― 例えば第2節で示唆したような単純な規則に基づくもの ―― を行う場合でも、しばしば主体には、自分がおかれた具体的状況を注意深く観察し、その中で微妙な差異を見分けることが求められる。だがもし主体の中に、A^T が実現されるべきだという（T への）信念が残っていたとすれば、A^T への意識はときに彼の注意力を奪い、代替的熟慮の観点から集中して状況を観察することを妨げるかもしれない。こうした集中の妨害は、とりわけ速断が必要な状況では有害になる恐れがある。

第二に、T への信念を保持していることで、代替的熟慮の恣意的な中断への誘惑が生じる可能性もある。T を信じる主体にとって、A^T は定義からして、代替的熟慮における目標 E^T よりも本来は価値の高い目標である。それゆえ、A^T と E^T が食い違うと思われた場合には、主体はつい代替的熟慮を中断して A^T を直接追求したくなってしまう、というのはかなりありそうなことだ。（例えば第2節で見た行為功利主義者は、「嘘をつくな」という規則を破った方が大きな幸福を生み出せることが確実だと思えるような状況でこうした誘惑を感じるだろう。）しかしこの誘惑は、主体の行動から一貫性を失わせたり、結果的に誤った判断へと導いたりといった仕方で、結局は彼を A^T の実現から遠ざけるかもしれない。

第三に、D・コックスらが指摘する興味深い可能性がある。それは、主体が T への信念をもっていることで、代替的熟慮とそれに基づく行為が無価値なものと感じられるようになってしまうという可能性だ（Cox 2012: 299f; Wiland 2008: 386）。これはとりわけ、T による正しい行為の基準があまりにも高く厳格であるケースで起こりうる。例として、ある義務論は、非常に要求の大きな行い（e.g. 受けた恩は二倍にして返すこと）を我々が果たすべき義務として定めているとしよう。そして、ある時点でその理論を信じた一人の主体が、これよりも無理なく自分にも続けられる行い（e.g. 受けた恩をできる範囲で、せいぜいそれと同等に返すこと）を代替目標として設定し、普段はもっぱらその代替目標の実現に努めるようになったとする。さていま、この主体が当の理論の目標を改めて思い起こす機会をもったとしよう。このとき、もし彼がこの理論の高い要求を真正な義務として真に受けすぎるとすれば、彼はそれまで自分が行ってきた慎ましやかな善行（程々の恩返し）とその本来の義務の間にある大きな隔たりを強く意識せざるを得ず、それによって無力感を深めた結果、その慎ましやかな善行をばかばかしく思うようになる（そしてやめてしまう）かもしれない。この場合逆に、もしこの主体が当の義務論を偽だと信じ、自分の行いは非の打ちどころのなく善いと満足できていたら、彼の行いは結果的にその理論の目標により近づいていただろう。つまりこのケースは、ある道徳理論への信念がかえってその目標の実現から主体を遠ざけてしまう場合として理解されうる。（Cox 2012: 299f が論じるように、これと同様の議論は第2節で見た行為功利主義や徳倫理の理論に関しても可能だ。）

このようなわけで[24]、我々の現実の経験的あり方をふまえる限り、ある自己抹消的な理論 T の目

24　以上に加えた四つ目の理由として、愛情や友情に関する論点も挙げられうる。多くの道徳理論（特に帰結主義や義務論など）は、本来の善悪の基準は不偏的なものだと主張しながら、日々の熟慮においては愛情などの特別な人間関係に基づいて（偏った仕方で）思考することを勧めるかもしれない。だがそうすると、当の不偏的な道徳理論への信念は、この特別な関係を損ねてしまう可能性がある（先の第3節で見た第三の反論と Smith 2001 を参照）。これに対する応答は、後註 28 と Railton 1984: 146f を参照。

標 A^T が最もよく実現されるのは、主体が適切な代替的熟慮の傾向性に加え、「T は偽だ」という信念 ── 「T は真だ」という信念でなく ── をもつ場合だという可能性は十分にある。しかも今みた論拠の一般性からして、このことは、稀であるどころかむしろ大抵の自己抹消的な理論について成り立ちそうにみえる。だがそうだとすると、実のところ多くの自己抹消的な理論は、それを信じる者にある受け入れがたい指示を発していることになるのではないだろうか。ここで思い出すべきは、一般に道徳理論 T は、我々の習慣や心的傾向性ができるだけその目標 A^T の実現に適したものになることを求めていたという点だ。したがって、もし「T は偽だ」という信念をもつことが A^T の実現にとって最適であるなら、T はまさにその信念をもつよう我々に指示するはずである。だがこの指示は、すでに T を真と信じている者からすれば、自分の信じている理論を偽だと信じるようになれ、というものに他ならない[25]。この指示の受け入れがたさは明白だろう。

ただしこの指示の受け入れがたさは、その「実行不可能性」に由来するものとして捉える必要はない。たしかに、我々は信念を自由自在にもったり捨てたりできるわけではないから、あることを信じるようになれという指示は、文字通り解釈すると心理学的に不可能な要求でしかないように見える。しかし我々は、ある程度の範囲内でなら、何らかの因果的なはたらきかけ（e.g. 催眠術、洗脳、脳手術を受けるなど ── cf. Parfit 1984: 41）によって、自分の信念状態を意図したものに変化させることはできる。したがって、ここで理論 T が発する指示は（実行不可能なものとして解釈しないとすれば）、「T は偽だという信念が自分の中に生じるよう自らにはたらきかけよ」というものとして解釈できる。

しかしたとえ実行不可能ではないとしても、この指示はやはり別の意味で受け入れがたい。というのも、それは端的に言って「自己欺瞞」の指示に他ならないからである[26]。一般に、自分の目から見て偽であるような信念を他者に植えつけることは明らかに正しくない。だとすれば、そのようなことを自分に対して行うよう命じる道徳理論が正しいはずはない。もちろんここで、この手段によって得られる結果のよさ（A^T のよりよい実現）が引き合いに出されるかもしれないが、そもそも我々は道徳的善を実現するための道具ではないのだから、この不正な手段は正当化されないだろう。しかも、いま植えつけが問題になっている信念は、道徳的善悪という我々の行為全般の価値に関わるきわめて根本的な種類のものだ。その根本的な部分に関して自分を欺けと命じることは、自分の生の全体をまやかしの上に築けと命じることに等しく、到底受け入れられない（cf. Cox 2012: 299）。したがって、そのようなことを命じる自己抹消的な道徳理論はやはり拒否されるべきである。

応答

さて以上の批判は、自己抹消的な道徳理論を拒否するための説得的理由になっているだろう

25　ただし T は、すでに T を信じている者以外に対しては、T をはじめから信じさせないよう命じることになるだろう。つまり T は、シジウィックが「秘教的 esoteric」（Sidgwick 1966: 490）な道徳理論と呼んだものになるだろう。この意味での秘教性が特に深刻な問題ではないという主旨の議論は、例えば Railton 1984: 155; Wiland 2007: 295f にある。

26　しばしば「自己欺瞞」という語は、ある信念をもったままそれと矛盾した信念を自分自身に植えつける過程（ないしそれにより生じた状態）を指すために用いられる（いわゆる自己欺瞞の「静的パラドクス」が問題になるときにはこちらが念頭におかれている。Cf. 浅野 2012: 第 1 章）。だが本稿ではこの語を、ある時点でもっていた信念を放棄してそれと矛盾した信念を自分自身に植えつける過程（ないしそれに伴う状態）を指すために使う。自己欺瞞をめぐる議論において「パスカルの賭け」と呼ばれている過程は、この後者のタイプの過程を例示するものだ（cf. 浅野 2012：第 3 章（特に 121-2 頁））。

か。私の見るところその答えはここでも否である。それを示すためにまず略記法を導入しておくと、以下では、道徳理論Tは真だという信念を「B（T）」と表し、また、Tは偽だという信念を「B（¬T）」と表すことにする。そうすると上の批判は、大部分の自己抹消的な道徳理論Tについて次の二つが成り立つ、と主張するものとして整理できる：

主張1. 現実の経験的条件の下では、Tの目標A^Tが最もよく実現されるのは、我々が信念B（T）をもたずにB（¬T）をもつ場合である[27]。

主張2. もし道徳理論Tについて主張1が成り立つなら、Tはある時点で信念B（T）をもった主体に対して、その信念を捨てて信念B（¬T）をもつよう命じる。

そして批判者はこれらに基づき、一般に自己欺瞞を命じるような道徳理論は受け入れがたいのだから、上の1と2を満たすような理論Tも受け入れがたい、と論じていた。

この批判に対してはもちろん、個々の道徳理論Tごとに別個の対応を行うことも可能だろう。だがここでは、より一般的な応答を試みたい。そうした個々の対応の成否は、結局は以下でとり上げる論点に帰着することが見込まれ、個々のTに固有な事情に依存しない議論の方がその論点を際立たせやすいと思われるからだ。またここでは、上の批判へのそうした一般的な応答のうちでも、その最後のステップを拒否する――自己欺瞞を命じるような道徳理論が誤っているとは限らないと主張する――道には立ち入らないことにする。おそらくこの道をとることも不可能ではないが（cf. Parfit 1984: §16）、自己欺瞞の道徳的地位に関してはすでにかなり議論の蓄積があり（cf. Jenni 2003）、この限られた紙幅で多少なりとも明確な結論を出すのは難しいと思われるからだ。

というわけで我々の問題は、上記の主張1と2が妥当かどうかという点になる。この点を考える上で重要なのは、この両者に共通して現れる主張1の読み方だ。いま注目すべき部分が際立つように要約すると、主張1が述べているのは、A^Tの実現にとって、「現実の経験的条件の下ではB（¬T）をもつことが最適だ」ということである。しかしこの主張には、実のところ二つの可能な読み方がある。すなわち一方で、この主張は、我々の「経験的条件」を現時点でのそれとして厳密に捉えた上で、短期的・一時的な最適性を主張するものとして読まれうる。つまりこの場合、1が述べるのは、「我々の経験的なあり方（心理的・身体的特性や社会的・生物学的環境など）がいま現在のままである限りはB（¬T）をもつことが最適だ」ということになる。この読みの下では、1は我々の経験的あり方が現在のそれとは異なるような状況に関しては何も述べていないわけだ。しかし他方、主張1は我々の「経験的条件」というものをより長いスパンで緩やかに――現時点でのあり方を一つのフェーズとして含むが、それと異なったものへも移り変わりうるものとして――捉えた上で、長期的・総合的な最適性を主張するものとしても読まれうる。この読み方の下では、主張1は、「我々の変化していく（しうる）経験的あり方に照らしてトータルでみるならばB（¬T）をもつことが最適だ」ということを意味するものになる。

この区別を念頭において見直してみると、たしかに主張1は、いま言った一つ目の意味で読ん

27　より正確には、我々がこの信念状態に加え適切な代替的熟慮の傾向性を同時にもつ場合、であるが、この点はいまの論点に関わりがないので以下では省略する。

だ場合には真になるかもしれない[28]。すなわち、上で挙げられたいくつかの論拠が示唆するように、我々のいま現在のあり方だけを見て評価する限りは、A^T の実現にとって最適なのはB（¬ T）をもつことだと言えるかもしれない。しかしここで重要なのは、上の批判が全体として成り立つには、主張1は前段落の二つ目の意味でも真にならなくてはならないという点だ。このことは、a）上の批判が多義性の誤謬に陥るのでないとすれば、主張1はもう一方の主張2が真になるのと同じ意味で真でなくてはならず、b）主張2が真になるのは、その前件である主張1を前段落の二つ目の意味において読んだときだけである、という二つの事実から帰結する。このうちa）についてはよいとして、b）が成り立つ理由についてはもう少し説明の必要があるだろう。すでに見たように、一般に道徳理論Tは、その目標 A^T の実現に最も貢献するような心的傾向性をもつことを我々に勧める。しかしもちろん、Tが我々にもつよう勧めるのはあくまでも長期的・総合的に見て A^T の実現に貢献するような傾向性であり、ある傾向性が単に特定の条件下でその実現に貢献するというだけでは、Tがその傾向性を我々に勧めるための条件として十分でない。つまり、主張2の前件（＝主張1）が前段落の一つ目の意味で読まれたときには、その前件は必ずしも後件の成立を保証しない。その保証がある（＝主張2が真になる）のは、主張1を前段落の二つ目の意味で読んだ場合に限るのである。

だがこの二つ目の意味で読んだ場合、主張1は真にならないと考えるべきもっともな理由がある。つまり長期的・総合的に見て、我々がB（T）の代わりにB（¬ T）をもつことが A^T の実現にとって最善であるということはまずありそうにない。その理由はひとことで言えば、そのような信念状態にある主体は、必要が生じた場合にも代替的熟慮の方法を適切に選び直すことができず、それは A^T の実現にとって致命的に不利だから、というものだ。

この点の説明のため、ある自己抹消的な理論Tを信じる主体が、現時点である代替的熟慮の方法dを身につけているとしよう。こうした主体に関して、いま注目すべき点は二つある。第一に、彼はときに、d以外の熟慮方法を選び直すべき（それが合理的であるような）状況に直面しうる[29]。例えばそうした状況としては、以前この主体がdを選んだときに成り立っていた経験的条件が何らかの仕方で変化し、もはやそこからdの最善性が帰結しないようなものになる（そして彼がそれに気づく）ケースがある。あるいは経験的条件そのものは不変でも、何らかの経験的発見により A^T の実現にとってdが最善だと信じることがもはや合理的でなくなるケースや、さらにはもっと単純に、主体が何らかの誤った経験的信念や誤った推論によってdを選んでしまっていたようなケースもありうる。おそらく、熟慮方法の見直しが必要になるこうしたケースはさほど稀でもないだろう。

しかし注目すべき第二点として、これらのケースにおいては、理論Tに関して主体がどういう信念をもっているかが、その後の A^T の実現されやすさを決定的に左右することになる。すなわち一方で、もし主体がB（T）を保持しているなら、彼はdに代わる新たな熟慮方法を選ぶために必

28　ただし実はすでにこの点に関しても、大いに異論の余地がある。というのも第一に、行為功利主義の場合にそう考えられているように、ある道徳理論Tにとっての代替的熟慮が複数の規則を含むような場合、規則同士が衝突する状況が生じることが予想され、それらの葛藤を解決するためには本来の目標 A^T（幸福の最大化）を参照することが不可欠になるはずだからだ（cf. Mill 1996: 71; Hare 1981: ch.2）。また第二に、仮に A^T の実現にとって最善の代替的熟慮が愛や友情といった偏向的な傾向性に基づくとした場合も、A^T が代表する不偏的な視点をもっている方が、愛情の過剰や不足の際でも安定した振る舞いが可能になるということは十分ありうる（Railton 1984: 146-8）。

29　もちろん第2節でも触れたように、熟慮方法の選び直しと身につけ直しには相応のコストがかかるだろう。ここで考えているのは、そうしたコストを差し引いてもなおそれらを実行すること（これは場合によると教育を通じた世代間作業になるだろう）が A^T の実現によりよく寄与するようなケースである。

要な参照点（= A^T）をもつことができる。しかし他方、主体がB（T）を捨てB（¬ T）をもつようになっているとすれば、彼はもはや A^T を目標と捉えていないのだから、その実現に適した熟慮方法を彼が選び直せる見込みはごくわずかしかない（そもそも熟慮方法を選び直すべきであることに気づかない恐れもある）。よって主体がB（¬ T）をもつ場合には、ある種の状況において A^T が実現される確率は著しく低下することになる。そしてこうした（おそらくさほど稀でない）状況への対応能力を奪うことになる以上、B（¬ T）をもつことが総合的・長期的にみて A^T の実現にとって最善であるということはまずありそうにない。むしろ A^T の実現のためには、我々はやはりB（T）を保持しているべきだと考えられる[30]。（ここでの事情はちょうど、ある生物種が現在の特定の生態環境に対してどんなに適応的だとしても、多少の環境上の変化に順応できないならば、種の存続という目的にとって総合的・長期的に有利だとは言えないのと類比的だ。）

もっともここで批判者からは、次のような応答があるかもしれない。たしかにいま論じられたように、信念B（T）を保持していることには A^T の実現に対し一定のプラスの効果があるかもしれない。だがここで我々は、この信念は同時に様々な仕方で――代替的熟慮への集中を妨げたり、代替目標を無価値なものだと感じさせたりといった仕方で――A^T の実現に対しマイナスの影響も及ぼすということを忘れるべきでない。この影響は、やはり決して無視できないものなのではないか。つまり、B（T）をもつことによるマイナスの効果は場合によると、長期的に見てもそのプラス分を上回ることになり、その結果、A^T の実現にとって最適なのはそれと反対の信念B（¬ T）をもつことになる（＝主張1が14頁で述べた二つ目の意味で真になる）、という可能性は十分あるのではないか。

繰り返しになるが、筆者の見立てでは、B（T）の代わりにB（¬ T）をもつことの不利性は前述のようにかなり致命的で、いま述べられた可能性も実際はあまりありそうにない。だが、これはたしかに経験的と言えば経験的な問題なので、以下では議論の「保険」として、もし仮にその可能性を認めたとしても、上の批判を拒否する余地はなお十分に残る、ということを示しておきたい。より正確に言うと以下で示したいのは、たとえ現実の経験的条件下では長期的に見てもB（¬ T）をもつことが A^T の実現にとって最適であるとしても、理論Tが我々にB（¬ T）をもつよう命じていると解釈する必要は必ずしもない（主張1が先の二つ目の意味で真になるとしても主張2を認める必要はない）、ということだ。

このことを示すため、まず次のように想定してほしい。いまあなたは、Aという事態を実現されるべき目標と捉えている一方で、ϕ という行為をすることはどうしても避けたいと考えている。しかしあるときあなたは不本意にも、「現実の諸条件の下では、ϕ をすることがAの実現のための最善の方法だ」という判断――例えば、「現在の収益率が続くなら、その海外支店をたたむことが会社の存続にとって最善だ」という判断――に至る[31]。このとき、もしあなたにとって、目標Aを

30　主張1の不成立を示そうとする以上の議論は要するに、Tが主体に対しB（¬ T）をもつよう指示することはない、というものだ。しかしここで次のような疑問があるかもしれない――たとえいま問題の主体がB（¬ T）をもつ（= A^T を道徳的に価値ある目標とは捉えていない）としても、主体が A^T の実現を（道徳以外の観点から）望んでさえいれば、彼はdに代わる適切な熟慮方法を必要な場合に選ぶことができるはずだ。それゆえ、Tが主体に対し、B（¬ T）をもちながら A^T の実現を望むよう指示すること（したがって当然B（¬ T）をもつよう指示すること）は十分ありうるのではないか――と、このような疑問である。しかしながら、この主体が A^T の実現を望むとすれば、彼の熟慮の過程はやはりその欲求によって影響を受けることになるだろう。そしてこの欲求は、本文でみたような理由（特にはじめの二つ）から、代替的熟慮の実行を妨げるだろう。よって実際のところ、もし理論TがB（¬ T）をもつよう指示するとすれば、Tは A^T の実現を望まないよう指示することになると考えられる。

31　ここでは、「ϕ ＝その海外支店をたたむこと」、「A ＝会社が存続すること」である。

放棄することはもちろんφをすることもできない相談だとしたら、あなたはどうすればよいだろうか。その一つの答えは、この判断を真にしている現実の諸条件（e.g. 現在の収益率が続くこと）それ自体を変化させること、である。つまりこの場合あなたは、問題の諸条件に然るべき変化をもたらすこと（e.g. 収益率を改善すること）で、φを行うことがAの実現にとってもはや最善でないという状況をつくり出すよう試みることができる。要するに上の判断は、φを指示するものとしてではなく現状改革を促すものとしても解釈できるのだ（むろん一定の範囲内ではあるが）。

そしてこれと同様の解釈は、一見自己欺瞞を命じているように見える道徳理論の場合にも可能である。すなわち、たとえある理論Tを信じる主体が「現実の諸条件の下では、B（¬T）を自分に植えつけることがA^Tの実現のための最善の方法だ」という判断に至ったとしても[32]、彼はこの判断を、必ずしも自己欺瞞の指示として理解する必要はない。むしろ彼はこの判断を、関連する経験的条件を変化させよという勧め、典型的には、自身の道徳的な思考能力を改善すべしという自己変革の勧めとして解釈することもできる[33]。

以上より、本節で見た自己欺瞞に関する批判は説得的でない。なぜなら、それを構成する第一の主張は批判の成立に必要な意味では真でない公算が高く、また、たとえ第一の主張がその意味で真だとしても、第二の主張を拒否できる可能性は依然残るからである。

5. おわりに ── 非経験的な自己抹消性

本稿では、自己抹消的な道徳理論が直面するとされるいくつかの問題をとり上げ、それらが本当にこの種の理論にとって難点になるかどうかを検討してきた。ここまでの議論から得られたのは、それらは実際には深刻な問題を提起しないという結論である。

だが実のところ、道徳理論の自己抹消性に関しては以上で見てきたのとは若干異なる種類の問題がある。本稿の第1節で見たように、道徳理論の自己消去性は、「我々がその理論の与える目標を直接的目標として行為すると、その実現は逆に困難になる」という条件によって定義される。そして本稿ではもっぱら、この定義条件が、我々についての諸々の経験的真理に依存して成り立つようなケースを考察してきた。だが何人かの論者が指摘するように、この定義条件はときとして、そうした経験的真理なしでも成り立ちうる。つまり言ってみれば、「論理的・非経験的な自己抹消性」と呼びうるケースがあるのだ[34]。例えば、ある種の状況において「親切な行為をすること」を目標として課すような道徳理論を考えることはできるだろう。しかしこの目標を実現しようという明示的意図は、当の目標の実現をいわば論理的に排除してしまうように思われる。というのも、ある他者Cの苦境に居合わせたとき、真に親切な行為者は「Cを何とか助けたい」といったCへの直接的配慮に導かれて手を差し伸べるのに対し、上の目標に直接導かれて行為する者は、同じ手助けをするにしても「親切な行為に該当することをしたい」といった考慮を経てそれをすることになる ── そしてこれは親切な行為ではない ── からである。もちろんこれと同様の議論は、

32　ここでは、「φ＝信念B（¬T）を自分に植えつけること」、「$A = A^T$」である。

33　もちろん、こうした現状改革がいつでも可能であるとは限らない（その可能性の有無は、Tの性質や主体のおかれた偶然的状況などに左右されるだろう）。しかしここでのポイントは、そうした改革の可能性がアプリオリには排除されない以上、上記の判断をそうした現状改革の勧めとして解釈できる可能性も一概には排除されないという点だ。

34　Cf. Stocker 1976: 462; Cox 2006: 511; Keller 2007: 225f; Martinez 2011: 279.

友情、愛情、勇敢さ、謙虚さ、等々についても成り立つ。だがそうだとすると、我々は上のようなタイプの目標を与える道徳理論とどう付き合えばよいのか。そうした理論を受け入れつつ、その目標の実現を排除しないことはどうやって可能なのか[35]。これらの点についての考察は、また別の機会に譲ることにしたい[36]。

参考文献

Adams, R. 1976. Motive Utilitarianism, *Journal of Philosophy* 73: 467-81.

浅野 光紀 2012.『非合理性の哲学―アクラシアと自己欺瞞』、新曜社 .

Cox, D. 2006. Agent-Based Theories of Right Action, *Ethical Theory and Moral Practice* 9: 505-15.

Cox, D. 2012. Judgment, Deliberation, and the Self-Effacement of Moral Theory, *Journal of Value Inquiry* 46: 289-302.

Hare, R.M. 1981. *Moral Thinking: Its Levels, Method and Point,* Oxford UP.

Hursthouse, R. 1999. *On Virtue Ethics*, Oxford UP.

Jackson, F. 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection, *Ethics* 101: 461-82.

Jenni, K. 2003. Vices of Inattention, *Journal of Applied Philosophy* 20: 279-95.

Keller, S. 2007. Virtue Ethics is Self-Effacing, *Australasian Journal of Philosophy* 85: 221-31.

McDowell, J. 1998. *Mind, Value and Reality*, Harvard UP.

Martinez, J. 2011. Is Virtue Ethics Self-Effacing?, *Australasian Journal of Philosophy* 89: 277-88.

Mill, J.S. [1871] 1998. *Utilitarianism*, fourth edition, edited by R. Crisp, Oxford UP.

Parfit, D. 1984. *Reasons and Persons*, Oxford UP.

Pettigrove, G. 2011. Is Virtue Ethics Self-Effacing?, *Journal of Ethics* 15: 191-207.

Railton, P. 1984. Alienation, Consequentialism, and the Demands of Morality, *Philosophy and Public Affairs* 13: 134-71.

Sainsbury, R.M. 2011. Fiction and Acceptance-relative Truth, Belief and Assertion, in F. Lihoreau (ed.) *Truth in Fiction*, Ontos: 137-52.

Sidgwick, H. [1907] 1966. *The Methods of Ethics*, seventh edition, Dover.

Smith, M. 2001. Immodest Consequentialism and Character, *Utilitas* 13: 173-94.

Stocker, M. 1976. The Schizophrenia of Modern Ethical Theories, *Journal of Philosophy* 73: 453-66.

Wiland, E. 2007. How Indirect Can Indirect Utilitarianism Be?, *Philosophy and Phenomenological Research* 74: 275-301.

Wiland, E. 2008. On Indirectly Self-defeating Moral Theories, *Journal of Moral Philosophy* 5: 384-93.

Williams, B. 1973. A Critique of Utilitarianism, in J. C. Smart & B. Williams, *Utilitarianism, For and Against*, Cambridge UP: 77-150.

Williams, B. 1995. Acting as the Virtuous Person Acts, in R. Heinaman (ed.), *Aristotle and Moral Realism*, UCL Press: 13-23.

35　この点に関連する議論は、例えば Williams 1995; Hursthouse 1999: pt. II などにみられる。

36　本稿の執筆にあたり、匿名の査読者から多くの啓発的で有益なコメント・質問をいただいた。記して厚く感謝する。また本稿は、平成 27 年度文科省科研費の交付をうけた研究（課題番号 15H06088）の成果の一部である。