

直観的な道德判断における抽象性と具体性の問題

太田紘史

1. 直観的な道德判断の可変性

我々は、何が善で何が悪か、何が正当で何が不当かといったことについて、直観的な判断をなすことができる。もちろん各々の直観的な道德判断が正しいか誤っているかは別にして、少なくとも我々は陰に陽にそういった判断に基づいて、日々の出来事に対して道德的な価値づけを与えて意思決定をしている。そして直観的な道德判断は、日々の生活だけではなく、倫理学の歴史においても重要な位置を占めてきた。とりわけそれは、規範理論を構築するための基礎的な足場として訴えられることがある。しかしそれは、規範理論の構築においてはむしろ夾雑物であり、退けられるべきものとして扱われることもある。

我々は、直観的な道德判断を基礎として規範理論を構築すべきなのだろうか。この認識論的な問いを考えるうえでは、また別の問いを考察する必要がある——そういった判断は、どれくらい堅固なものなのだろうか。直観的な道德判断もまた、人間の心が生産するものの一つであるかぎり、それがどれくらい頑健な認知的生産物なのかということが問われてしかるべきである。もしもそれが頑健なものでなければ、それが規範理論の構築において持つ価値はかなり低くなるだろう。これに関連するまた別の問いは、そういった判断はどのような条件下で生みだされるときに適切なものと見なせるのか、というものである。直観的な道德判断は、状況しだいでは度合はおろか方向性すら異なりうるというくらいに、可変的なものかもしれない。すると、仮に直観的な道德判断を規範的な基礎として利用するとしても、考慮に入れられるべきなのはどの状況での判断なのだろうか。

これらの点は、半ば概念的、半ば経験的な問題である。とくに最近の哲学者、倫理学者、心理学者、神経科学者は、直観的な道德判断に焦点を当てる経験的な研究を行ってきた。興味深いのは、そういったいくつかの研究が、直観的な道德判断の可変性を実際に示唆している点である。人々が善悪や規範的価値にまつわる特定のシナリオないし思考実験において直観的な道德判断を下すとき、その判断は様々な要因によって左右される。その要因には、判断状況における情動状態や(Lerner, Goldberg & Tetlock, 1998; Schnall, Haidt & Jordan, 2008; Wheatley & Haidt, 2005)、判断対象となるシナリオの言語的表現や呈示順序(Petrinovich & O'Neill, 1996; Wiegmann, Okan & Nagel, 2012)といったものが含まれる⁽¹⁾。

このような経験的事実は、道德哲学にとって認識論的に重要なものとなるだろう。実際

Sinnott-Armstrong(2006)は、道徳直観の可変性（彼の言葉では「信頼不可能性」）を根拠にして、直観主義を反駁しようとしている。また、例えば情動状態のあり方によって道徳判断が変化しうるのだとすると、情動と道徳判断の関係性をどのように理解するかによって、規範理論の基礎とすべき道徳判断の選択も変わってくるだろう。

だが、こういった議論や考察においては、そもそもそういった可変性ないし信頼不可能性がどの程度のものであるのか、そしてそれがどういった心理的背景から生じているのかについて検討することが必要である。というのも、認識論的に影響があるくらいの実質的な可変性が存在するのかといった点は、そういった心理的背景についての理論的理解なしには論じられないからである。私は今回、このような方向の検討を示したい。

今回私が考察対象とするのは、「抽象／具体パラドクス」である。我々の直観的な道徳判断は、呈示されるシナリオが抽象的に表現されるか具体的に表現されるかによって変化しうるのが、経験的知見として示唆されている。私は、この現象の背後にある心理的背景についての理論的枠組みを提案することを通じて、その認識論的な含意を探りたい。

本論文は以下のように進む。第2節では、抽象／具体パラドクスの代表的事例を列挙し紹介する。第3節から第5節にかけて、これらの事例を説明しようとする案を批判的に検討する。それを踏まえて、第6節で私なりの説明を提案し、第7節でその認識論的な面での含意について述べる。

2. 抽象／具体パラドクス

まず代表的な事例をいくつか紹介する。いずれも行為の道徳的責任を問うものであり、ある種の直観的な道徳判断を問うものである。

Nichols & Knobe(2007)は、決定論的世界において道徳的責任が成立しうるかどうかに関して、人々がどのような直観的判断を下すかを調査した。呈示される問いは二つある。一方は、決定論的世界において秘書の女性と一緒にいるために妻と息子を殺した男性について描写したうえで、この男性が完全な道徳的責任を負うことが可能かを問うものである（具体条件）。他方の問いは端的に、決定論的世界で人が完全な道徳的責任を負うことは可能かを問うものである（抽象条件）。結果として、具体条件では大半の人（72%）が道徳的責任を帰属したものの、抽象条件では大半の人（86%）が道徳的責任を帰属しなかった。すなわち、決定論的世界における道徳的責任の帰属可能性は、抽象的な表現よりも具体的な状況の認識によって高まるように見える。

また Nahmias, Coates & Kravav(2007)も、決定論的世界での道徳的責任について人々がなす直観的判断を調査した。ある問いは次のようなものである——架空の決定論的世界の住

人（「エルタ人」）が悪いことをすれば、その行為に関してその行為者は非難に値するか（抽象条件）。別の問いは次のようなものである——愛人と結婚できるように妻を殺したエルタ人のスミットは、妻を殺すというその意思決定に関して道德的責任を持つか（具体条件）。これらの問いに対して被験者は、非難に値することに同意するかどうかを答えた。結果として、やはり抽象条件よりも具体条件において当の同意が有意に多く得られた⁽²⁾。この結果からも、決定論的世界における道德的責任の帰属は、具体的な状況の認識によって惹起されやすいように見える。いずれの研究においても、抽象的に問われた場合には非両立論的な帰責判断がなされる傾向があり、具体的なシナリオを通じて問われた場合には両立論的な帰責判断がなされる傾向があったと、一見思われる。

以上は、決定論的な状況での行為に対する判断を調査したものであるが、他方で特定の脳状態において不可避となる行為に対する判断を調査したものもある。

De Brigard, Mandelbaum & Ripley(2008)は、脳の状態が引き起こす行動について、その道德的責任がどのように判断されるかを調査した。一方の条件では、被験者は次のような人物のシナリオを与えられる。デニスはある脳状態にあり、医師にこう告げられる——その脳状態が彼の行動を特定の仕方で行き起こしており、そして同じ脳状態の人ならば同じように行動するだろう、と。このシナリオに関して、被験者は判断を求められる——デニスは、自身の脳状態によって引き起こされる行動に対して道德的責任を持つか（抽象条件）。他方の条件では、被験者は次のようなシナリオを与えられる。デニスはある脳状態にあり、医師にこう告げられる——その脳状態が女性達をレイプするという彼のかつての行動を引き起こしたのであり、同じ脳状態の人ならば同じように行動するだろう、と。被験者は次を問われる——デニスは、自身の脳状態によって引き起こされたレイプに対して道德的責任を持つか（具体条件）。二つの条件の差異は、行為を具体的に描写するか否かの差異だけである。結果として、道德的責任があるとする判断が、具体条件のほうで有意に高い傾向で得られた。

このように直観的な道德判断は、抽象的に問われる場合と具体的なシナリオを通じて問われる場合とで、有意に異なる。我々は、このような抽象／具体パラドクスをどのように理解したらよいだろうか。以下では、これまでに提案されてきた三種類の説明を検討する。

3. 二種類の認知プロセスによる説明

Sinnott-Armstrong(2008)は、二種類の認知プロセスを導入することで抽象／具体パラドクスを説明しようとする。

彼は、意味論的記憶とエピソード記憶がそれぞれ、具体的状況の表象とそれから要約の

表象をなすという見解に訴える。Klein, Cosmides, Tooby & Chance(2002)の示唆によれば、エピソード記憶は自身の性格についての個別的な記憶表象を持つのに対して、意味論的記憶はそれらから要約された、自身の性格についての一般的な特徴づけに関する表象を持つ。人が自身の性格がどのようなものかについて問われるときには、意味論的記憶が持つ一般的な特徴づけの表象が利用されるとともに、エピソード記憶がその特徴づけに一致しない様々な反例（自らの行為に関するエピソード）を提供する。というのも、意味論的記憶はせいぜい要約された情報を符号化するに過ぎず、それと一致しない様々な個別的な情報はエピソード記憶によって符号化されているからである。

Sinnott-Armstrong は、このような二種類の記憶の差異と類比的に、具体的状況認識と抽象的状況認識の二つのシステムを提案する。抽象／具体パラドクスが生じるのは、二つの認知プロセスの独立的な作動と競合によるものだという。具体的なシナリオが与えられるときには具体的な認知プロセスが作動し、より強い仕方で帰責をなすような判断が生産される。他方で、抽象的なシナリオが与えられるときに働くプロセスが導く帰責判断は比較的弱い。もちろんこれら二つのプロセスの間には相互作用があるだろうが、それらが導く判断傾向の違いが、シナリオの具体性に応じた帰責判断の非対称性を結果するのだという。

この提案に対する疑問は、二つの認知プロセスがいかにして当該の判断における非対称性を結果するのかが不明であるというものである(Mandelbaum & Ripley, 2012)。すなわち、なぜ具体的状況認識のプロセスのほうが、他方よりもより厳しい帰責判断を下し、その逆ではないのだろうか。仮に、二つの認知プロセスが道徳判断においても上記のような仕方で競合するとしても、なぜ具体的な認知プロセスのほうが強い帰責を導くのかについての説明が必要である。

この点は、ちょうど二種類の記憶の場合と同じように、二つのプロセスが含む記憶表象の差異によって説明できるかもしれない。具体的プロセスは、不自由な状況下での様々な行為の事例について、責任を帰属しない形で符号化している。抽象的プロセスは、それらから抽象した一般的な特徴づけとして、《不自由な状況下での行為に責任はない》という内容を符号化している。しかし具体的プロセスが符号化するものには、必ずしもこの一般的特徴づけに一致せず、(おそらく比較的少数ながら)いくつかの反例となるような事例も含まれている。それゆえ、一連の実験の具体条件では、具体プロセスにおいてそういった反例となる事例の表象が活性化されることで、相対的に高い傾向で帰責を導く。他方で抽象条件では、抽象プロセスが上記の一般的な特徴づけに従って、相対的に低い傾向で帰責を導く。このような説明が与えられるかもしれない。

しかし、このような説明が可能になるためには、そもそも具体的プロセスにおいて、不

自由な状況下での殺人やレイプの事例が、責任を帰属される形で符号化されていなければならない。先述の二種類の記憶が衝突するのは、一般的な特徴づけの記憶表象と特定の事例の記憶表象が整合しないことによる。抽象／具体パラドクスで同様に考えたいのならば、一般的な特徴づけの記憶表象に整合しないような事例の記憶表象を指定せねばならない。

だが、そういった記憶表象が存在するかどうかは疑わしい。我々の日常的な道德判断が決定論的な状況での殺人や特定脳状態におけるレイプといった事例に関わることは、ほとんどないであろう。さらに、仮にそういった記憶表象が存在するとしても、なぜその記憶表象においてその殺人やレイプの事例が責任を帰属される形で——帰属されないのではなく——符号化されているのかということが問題になる。この問題はしかし、決定論的な状況での殺人は具体的な形で表現されるとより強く帰責されるのはなぜなのかという、当初の問題に帰着する。すなわち、具体条件における判断傾向を説明するために、記憶表象における個別事例の帰責上の身分に訴えたとしても、なぜその事例がそのような帰責上の身分を持つのかについて、説明が必要となるのである。そしてこのような道德判断の傾向性に関する疑問は、我々の当初の疑問に他ならず、問題を先送りにしたに過ぎない。

このように、二種類の認知プロセスを指定すること自体は、実験結果で見られるような非対称性のパターンを説明するうえでは——とくに具体条件での帰責の傾向性が高まるという点について——不十分であると考えられる。他方で以下に見るように、情動による説明はその点に焦点を当てるものである。

4. 情動による説明

道德判断の変化や差異を説明するうえでは、しばしば情動がその要因として提案される。抽象／具体の差異についても、情動による説明は次のような仕方で提案されうる——具体的な状況の認識は情動を惹起し、その情動が直観的な道德判断に影響する。

この考え方は Nichols & Knobe(2007)によって実際に提案されている。彼らはさらに、そういった情動による影響が干渉的なものだ主張している。すなわち、人々は本来非両立論的な道德的責任概念の適用をするのだが、具体的な状況の認識によって情動が惹起され、その情動が責任概念の適用に干渉的に作用し、結果として両立論的な帰責がなされるのだという。

本稿では、このような情動による因果的効力が干渉的なものかどうかを問わず、むしろ端的に、情動が抽象／具体パラドクスでも因果的効力を持つかどうか、そしてそれが説明上どう関わるかという点に焦点を当てたい。

たしかに、情動が道德認知において何らかの因果的効力を有することは、他の様々な研

究から知られている。いくつかの実験的研究によれば、行為を悪と見なす道德判断は嫌悪情動の操作によって増長される(Lerner et al., 1998; Schnall et al., 2008; Wheatley & Haidt, 2005)。これらも含めて一般に、否定的な情動は否定的な道德判断を導くというパターンが成立しているのかもしれない。例えば、Nichols & Knobe や Nahmias et al. の具体条件では殺人に対する怒りが惹起され、また De Brigard et al. の具体条件ではレイプに対する怒りが惹起されており、これらの怒りが帰責の度合いを高めたのかもしれない。

しかしこのような説明に対しては、二つの疑問を挙げることができる。

第一に、帰責における情動の因果的役割は判然としない。というのも、情動に障害を持つ被験者において決定論と道德的責任にまつわる実験を行ったところ、健常者と同様のパターンで応答をなすということが判明しているからである (Cova, Bertoux, Bourgeois-Gironde & Dubois, 2012)⁽³⁾。それゆえ抽象／具体パラドクスでの情動の役割についてははっきりせず、むしろ情動ではなく認知的な要因による説明が求められる。

第二に、たとえ帰責において情動が因果的役割を持つとしても、それで説明が完結するわけではない。というのも、情動が何らかの認知的要因を介して作用しているという理論的余地が残っているからである。一方で、情動はいかなる認知的要因も介さずに道德判断に影響しているのかもしれない。だが他方で、情動は何らかの認知的要因の促進や抑制を通じて道德判断に影響しているのかもしれない。前者のような情動だけによる説明を受け容れることは、何らかの認知的要因の貢献を踏まえた説明をすべて退けた後になされるべきである。

以上の点を踏まえ、我々は認知的要因による説明の可能性を探るべきである。先ほどの二種類の認知的プロセスによる説明はその一種であるが、以下で見るのはそれとはまた異なり、一種の規則の表象による説明として提案されているものである。

5. 規則の表象による説明

Mandelbaum & Ripley(2012)は、規則の表象に訴える説明を提案している。それによれば、我々は《規範が破られるときには常に、その責任は行為者にある》という無意識的信念を有している（略して NBAR = Norm Broken, Agent Responsible）。ここで言われる規範とは、最大限に広く解される規範である。それには、自然界のあり方（物体は無から生じない）や社会的慣習（ビジネスでの服装マナー）など、世界のあり方に関して期待されていることに加え、もちろん道德的規範も含まれる。何かが悪いこととして認識されれば——正確には何であれ規範に違反するものとして認識されれば——、その認識は NBAR との組み合わせにより、特定の行為者に責任があるはずだという帰責判断を傾向づける。もちろん帰

責対象となる行為者が見つからないこともあるだろうが、適当な対象が見つかる限り、我々はそれを責任ある行為者と見なし帰責することになる。このようなある種の規則の表象とそれに沿った判断に訴えれば、以下のような形で、抽象／具体パラドクスは説明されるといえる。

第一に、決定論的世界での道德的責任についての判断は、次のように説明されるという。一方で、人は本来非両立論的な直観を持つので、抽象条件では帰責しない。他方で具体条件では、殺人という有害な行為の認識が、NBAR を介して、行為者への帰責判断を導く。ここではもちろん、抽象条件と同様に非両立論的な直観も働く。だが、よく根づいた NBAR のほうが、相対的にあまり根づいていない非両立論的な直観に勝り、帰責判断を結果することになる。こうして抽象条件と具体条件の間で非対称性が見られることになる。

第二に、特定の脳状態における行為の道德的責任についての判断は、次のように説明されるという。一方で被験者は、精神的に病んでいる行為者には責任がないと本来信じているので、抽象条件では帰責しない。他方で具体条件では、上記と同様に NBAR を通じて帰責判断が傾向づけられる。やはり（決定論の場合と同様にして）NBAR が他方の信念に勝って帰責を結果し、二つの条件の間に非対称が見られることになる。

しかし私の見るところ、この説明に対して、以下のような指摘ないし疑問が与えられてしかるべきである。

第一に、この仮説は規則の表象を中心とした説明であるものの、そのような規則の表象を引き起こす認知的要因が何なのかは一切述べられていない。そしてそれは、他でもなく行為の認識であると考えるのが自然である。というのも、一連の研究における抽象条件と具体条件の差異は、殺人やレイプのような行為が特定されているかどうかにあるからである。たしかに NBAR の規則には、行為に帰責するという点が含まれているものの、行為の認識がそれに先んじるということを要求してはいない。だが、抽象条件と具体条件として用いられている素材を分析する限り、心理的に関連しているのは行為の認識であると考えるのが自然であり、仮に NBAR のような規則の表象が関わっているとしても、それだけでは説明として不十分である。

第二に、他方で NBAR の必要性が疑われる。というのも、上記のとおり説明上中心となるべきは行為の認識であって、その認識から帰責へいたる経路には他にもさまざまなものが考えられるからである。とりわけ、NBAR のような非常に一般的な規則の表象を措定するよりも、むしろ特定の種類の行為に相対的な規則の表象や、あるいはそもそもそういった表象なしに規則的なパターンとして判断を形成する傾向性を措定することでも説明は可能である。

このように一方で、この仮説で指定される説明項の充分性に問題があり、他方で指定される説明項の必要性に疑問を突き付けることができる。これらの点を踏まえ以下では、行為の認識を中心とする説明の枠組みを提案したい。

6. 行為の概念化と非推論的写像による説明

まず、行為の概念化の重要性について、より正確に指摘したい。

いずれの実験的研究においても、抽象条件と具体条件の差異は行為の表現における差異である。Nichols & Knobe の抽象条件では、行為がどのようなものかについて触れられておらず、具体条件では行為が殺人として具体化されている。また Nahmias et al. の抽象条件では、行為は「悪いこと (something bad)」として描写されているのに対して、具体条件では行為はやはり殺人として具体化されている。そして De Brigard et al. の抽象条件では「特定の行動」と表現されているのに対して、具体条件では「レイプ」と表現されている。いずれにおいても条件間で行為の表現に差異が存在する。

それゆえ、これらの条件間で生じた帰責の非対称性の背後に指定すべきは、行為の概念化に関連した心理的プロセスだと考えるべきである。すなわち、行為（とりわけ有害な行為）が認識されるかどうかによって、我々の道徳判断が左右されているのである。

この説明は、以下のような経験的予測をなす。仮に Nahmias et al. の条件の変種として、『10 月 15 日の正午ちょうどに何らかの悪い行為をなしたエルタ人のスミットは、その意思決定に関して非難に値するか』という記述を与えたとしよう。これは厳密に言えば、行為そのものを具体化した記述ではない。これが抽象条件に比して具体化されている部分は、行為の種類や、あるいはその意図性や帰結ではないし、またもちろん社会的規範との兼ね合いなどでもない。それはある意味でたしかに具体化を伴うものの、行為のあり方に関わるような具体化ではないのである。私の見解からの予測では、このような仕方での（いわば些末な）具体化は、被験者において直観的な道徳判断の傾向性を変えないであろう。

では、行為の概念化はいかにして道徳判断を左右しうるのだろうか。私は、それは行為に道徳的身分を与えるような心理的推移によるものだと提案したい。すなわち、例えば行為がいったん殺人として概念化されると、その行為の概念化の仕方に相対的な仕方で、その行為は道徳的に悪いものとして判断される。ここで、特定の問いを与えられるとき（例えばその行為者が道徳的責任を有するかどうか尋ねられると）、その判断に沿った仕方で、関連する信念が形成されて明示的に（ときに特定の度合いで）表明されるのである。

ここで、これはある種の道徳推論であると思われるかもしれないが、そうではない。たしかに道徳判断からその明示的な表明へ至るまでには、問いの形式に合わせた信念の変換

がしばしば必要になるだろう。例えば、道德判断は「この行為は悪い」というものであったとしても、問いは「この行為は悪いものか」というものではなく、「この行為に対して行為者は道德的責任を持つか」というものかもしれない。この場合には、最初の道德判断には含まれない《道德的責任》という概念の適用が要求される。このような場合には——とりわけ最初の道德判断との関連が認知主体にとって不明瞭であるような場合には——、ある種の推論が関わると考えるのがもっともらしい。

だが、そこでの心理的推移が推論的であるかどうかに関わらず、重要なのは、当の行為の概念化から道德判断へ推移する過程が推論的か否かという点である。そして私の考えでは、このような推移はむしろ、行為の概念化から道德判断への非推論的な写像であると考えべきである。というのも、少なくとも一連の実験的研究において被験者は、明示的な理由づけやその帰結を要求されているわけではなく、端的に、関連する言明への同意の度合いを答えるように要求されているからである。

ここで、この非推論的な写像を可能にする計算論としてどのような理論的な余地があるのかについて、三つの次元に沿った素描を以下に記しておきたい。

第一の次元は、この写像に規則の表象が伴うか否かである。一方の考えでは、特定の規則が内的に表象され、それが適用されている。他方の考えでは、特定の規則が内的に表象されることはないが、特定の規則的なパターンが例化されている。

前者の考えにおける規則の身分は、言語における統語論的規則の身分と類比的である。統語論的規則は心的に表象され、それは文を生産したり特定の文字列を文として判断したりするために適用されると考えられている。だが、このような規則の適用が推論と呼ばれることはない。それは端的に、規則の非推論的な適用である⁽⁴⁾。実際に言語学では、文字列が文法的に正しい文であるかについての直観的判断が用いられることがある。そこでは統語論的規則がその判断において適用されていると想定されるが、しかしそのゆえにその判断が推論を経たものだと思なされたりはしない。これと同様に、行為に道德的身分を帰属するような心理的推移もまた、ある種の非推論的なプロセスとして働くのかもしれない。

後者の考えにおける規則的なパターンは、コネクショニストネットワークの規則的な振る舞いと類比的である。コネクショニストネットワークは、ノード間の重みづけをある仕方で行うことによって、それ自体は規則を表象しないままに、規則的な仕方で行う振る舞うことがある。ここで、あるシステムが規則的な仕方で行う振る舞うのであれば、それは実質的には規則に従っているものとしても記述でき、それはそのシステムが規則を表象していることと変わりはないと言われるかもしれない(cf. Dennett, 1987)。だが一般に、それらは切り離されうる事態である。月が特定の自然法則に従って周回運動をなすからといって、月が

その自然法則を表象するわけでもなければ、何らかの表象が月の運動において因果的効力を持っているわけでもない。同様にコネクショニストネットワークも、規則の表象の適用を介さずに、特定の規則的なパターンで入力状態から出力状態へ推移する。このようなネットワークの推移と類比的な形で、行為への道徳的身分の帰属は進行しているのかもしれない。

第二の次元は、規則表象あるいは規則的パターンの一般性の度合いである。まずは、規則表象の場合に話を限ろう。当該の表象された規則は、何であれ意図的行為に責任を結びつけるような、非常に一般的なものかもしれない。あるいはそれは、殺人と責任を結びつける規則、レイプと責任を結びつける規則といった具合に、中間的なレベルの規則の一群かもしれない。このような分化は様々でありうる。例えば《殺人》と《意図的に死なせる》という二つのものは独立に概念化されており、一方の概念だけを含んだ規則の表象が存在するのかもしれない。そして規則の表象でなく規則的パターンにおいても、同様にして一般性の度合いを考慮することができる。

こうした点は、理論的な可能性にとどまらず、経験的にも探究可能である。例えばある実験的研究では、認知主体において《殺人》と《意図的に死なせる》という概念が道徳判断に異なる仕方でも貢献することが、トロツク問題のバリエーションを被験者に与えることを通じて示されている(Sinnott-Armstrong, Mallon, McCoy & Hull, 2008)。もちろんこの研究では、ある行為が道徳的に許されるかどうかについて被験者は判断しており、これは我々が目下問題としている帰責判断とは異なっている。だが重要なのは、規則（ないし規則的パターン）の一般性について、この種の経験的研究から積極的な示唆が得られるということである。

最後の次元は、因果的な駆動要因である。以上のような計算論は、純粋に認知的な仕方でも表現されている。しかし、情動がそれらの計算論的過程の駆動要因になりうることは、理論上排除されない。ここまで述べてきたような認知的要因に関わる因果的相互作用は、情動的な反応によって実現されているかもしれないからである。事実、先述の通り、情動が道徳判断にしばしば因果的に関わるということ知見は存在しており、このような知見の蓄積は情動による道徳判断の駆動を支持することになるだろう。ただし、そのような情動による影響の存在がいかなる認知的説明をも排除するというのもまた誤りであり、さらなる認知的要因による説明が求められるべきである。

7. 直観的な道徳判断の規範的身分について

以上の理論的枠組みは、心理的背景についていくつかの余地を残しており、各次元につ

いて確定していくことそれぞれが、何らかの認識論的な含意を持つだろう。しかしいずれにせよ、行為の概念化という主要な点は変わらない。ここで、この点が直観的な道德判断の規範的身分に関して持つ認識論的含意について、簡単に述べておきたい。

私は本稿の冒頭で、二つの懸念をとりあげた。第一に、直観的判断の可変性ないし信頼不可能性が、それを規範理論の基礎として用いることに対して否定的な含意を持つかもしれない。第二に、状況によって判断が異なるとき、それらのうちどれを選択して規範理論の基礎として用いるべきかが不明確となる。どちらの点においても、規範理論の構築における直観的判断の認識的身分が何かしら問題含みのものになりうる。

だが私の提案する心理的背景についての理論的理解は、こういった懸念をむしろ払拭する方向に働く。

抽象／具体パラドクスは、しばしば次のような現象として理解される——条件間で、同じ内容に対して異なった態度の道德判断が得られる。だが私の提案する理論的理解を通じて、このような見方はむしろ退けられるべきだということが示唆される。というのも、抽象条件と具体条件とされているものは、少なくとも認知主体が認識する限りにおいて、同じ内容を表現しているわけではないからである。抽象条件では行為が特定されていないのに対して、具体条件では行為が特定されており、認知主体はこれらの内容上の差異に可感的な仕方では道德判断を生みだしているのである。

この点から、直観的な道德判断の可変性ないし信頼不可能性それ自体が、表面的なものにとどまるということが示唆される。我々はむしろ安定的に、異なった状況や出来事の認識に応じて異なった直観的判断を下すよう傾向づけられているのである。これは、上記の第一の懸念を払拭することになるだろう。というのも直観的な道德判断は、同じ内容について異なる態度を示しているわけではないからであり、それゆえ当初意図された意味における可変性を示してはいないからである。そしてこれはまた、第二の懸念を払拭することにも貢献する。そもそもの懸念は、同じ状況について異なる態度の判断が下されるがゆえに、どちらの状況での判断に認識的な重みを与えるべきか、というものであった。しかしそもそも、抽象条件と具体条件とされるものの間で同じ主題が取り扱われているわけではないということ——少なくとも認知主体の認識の範囲では——を踏まえれば、もはやそのような判断の選択においてジレンマが生じることもないだろう。

経験的研究の進展の度合いは加速度的であり、他にも道德直観の可変性や信頼不可能性を示すとされる事例が次々に指摘されている。だが、私が今回論じたように、その心理的背景の理論的理解を通じて現象そのものの理解もまた大きく変わり、それとともに道德直観の規範的身分に対する見方も変わってくる。我々に必要なのは、経験的探求の重要性と

同じくらいに理論的研究の重要性を認識し、それを進めていくことであろう。

註

- (1) ここで、こういった研究は、思考実験や知的な明確化に慣れていない人々に対して行われている限り、その種の結果はとるに値しないという見方もあるかもしれない。だが第一に、哲学者による直観的な道德判断もまた、そういった道徳的に無関連らしき要因によって左右されうることが判明している(Tobia, Buckwalter & Stich 2012; cf. Schwitzgebel & Cushman, 2012)。また第二に、どのような仕方であれ哲学者の直観的な道德判断と非哲学者のそれが異なるときに、前者がいわば「本来の直観」によるものとして優先される根拠は何なのだろうか。ここで、哲学者が思弁的な推論に長けているということは答えにならない。なぜならば、目下の関心対象は、直観的判断というある種の非推論的な判断であるからだ。
- (2) 同意が得られた頻度は、決定論的世界が物理的因果により表現された場合は(抽象)50%と(具体)79%、決定論的世界が心理的因果により表現された場合は(抽象)70%と(具体)81%であった。
- (3) この点は Mandelbaum & Ripley(2012)によって指摘されている。
- (4) この考えは、いわゆる道德文法仮説の一部にも含まれるが(Mikhail, 2007)、生得性や発達に関する見解にまで踏み込んでいないという点でそれとは異なる。

文献

- Cova, F., Bertoux, M., Bourgeois-Gironde, S., and Dubois, B. (2012). 'Judgments about moral responsibility and determinism in patients with behavioral variant of frontotemporal dementia: Still compatibilists', *Consciousness and Cognition*, 21(2), 851-864.
- De Brigard, F., Mandelbaum, E., and Ripley, D. (2008). 'Responsibility and the brain sciences', *Ethical Theory and Moral Practice*, 12(5), 511-524.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press/Bradford Books.
- Klein, S., Cosmides, L., Tooby, J., and Chance, S. (2002). 'Decisions and the evolution of memory: Multiple systems, multiple functions', *Psychological Review*, 109, 306-329.
- Lerner, J., Goldberg, J., and Tetlock, P. E. (1998). 'Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility', *Personality and Social Psychology Bulletin*, 24, 563-574.
- Mandelbaum, E. and Ripley, D. (2012). 'Explaining the abstract/concrete paradoxes in moral psychology: The NBAR hypothesis', *Review of Philosophy and Psychology*, 3(3), 351-368.
- Mikhail, J. (2007). 'Universal moral grammar: Theory, evidence, and the future', *Trends in Cognitive Sciences*, 11(4), 143-152.
- Nahmias, E., Coates, D., and Kvaran, T. (2007). 'Free will, moral responsibility, and mechanism: Experiments on folk intuitions', *Midwest Studies in Philosophy*, 31, 214-242.
- Nichols, S. and Knobe, J. (2007). 'Moral responsibility and determinism: The cognitive science of folk intuitions', *Nous*, 41, 663-685.
- Petrinovich, L. and O'Neil, P. (1996). 'Influence of wording and framing effects on moral intuitions', *Ethology and Sociobiology*, 17, 145-171.
- Schnall, S., Haidt, J., and Jordan, A. (2008). 'Disgust as embodied moral judgment', *Personality and Social Psychology Bulletin*, 34(8), 1096-1109.
- Schwitzgebel, E. and Cushman, F. (2012). 'Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers', *Mind & Language*, 27, 135-153.
- Sinnott-Armstrong, W. (2008). 'Abstract + concrete = paradox', in S. Nichols & J. Knobe (Eds.), *Experimental Philosophy* (pp. 209-230), New York: Oxford University Press.
- Sinnott-Armstrong, W., Mallon, R., McCoy, T., and Hull, J. (2008). 'Intention, temporal order, and moral judgments', *Mind & Language*, 23(1), 90-106.
- Tobia, K., Buckwalter, W. and Stich, S. (2012). 'Moral intuitions: Are philosophers experts?', *Philosophical Psychology*, 1-10, 2012.
- Wheatley, T. and Haidt, J. (2005). 'Hypnotically induced disgust makes moral judgments more severe', *Psychological Science*, 16, 780-784.
- Wiegmann, A., Okan, Y., and Nagel, J. (2012). 'Order effects in moral judgment', *Philosophical Psychology*, 25(6), 813-836.

[東京大学大学院総合文化研究科／日本学術振興会特別研究員]