

演繹的道德推論の射程

村上 祐子[†]

立教大学[†]

1. 序

適切な意思決定のためには様々な状況を勘案して論理的に推論すればよい結論が出てくるだろうと考えるのはごく自然である。だが、「これはよい」「こうすべきだ」といった価値判断を含む結論を導き出すためには、価値観の記述と事実記述が前提に混在する中で、論理的推論をすすめるなければならない。また、実際におこっている状況だけではなく、さまざま起こりうる状況を想定したうえでどのようにすべきなのかを考察していくことになる。また、このような価値判断の体系における道德原則が存在するのか、という問いに答えるには道德的推論が必要となる一方で、あたかも物理学や幾何学の理論研究のように、道德原則と道德的推論体系が得られれば道德的命題が得られるはずであるという研究プログラムも展開されてきた。

このような価値判断にかかる推論（実践推論）をどのように理論化するのか？という問いは哲学の最初期からの課題であり、多くの哲学者が取り組んだ結果、「事実と価値」、「事実と行為」、「道德的状态にかかる記述と忠告」といった理論構築に向けた重要な概念的区別がなされてきたが、未だ決定的理論は出ていない。一方で、理論模索の過程で提案された理論をどのような数理モデルに載せるのか？という問いは善悪の概念そのものの探求を含むという意味で哲学の数理プロジェクトの一環と考えられる。

価値判断を含む推論にかかるこれまでの提案は、義務論理を代表とする演繹的アプローチと機械学習のような過去データを利用する帰納的アプローチに大別されるが、現在のトレンドは後者である。演繹的道德推論は極めて限定された意味で用いる限りでは非常に強力である一方で、我々が普段用いる価値概念の多様性が暴露される。

当発表では演繹的道德推論研究史を概況する。特に第2節と第3節では、形式化以前の哲学的・倫理学での分析を中心に、現在の道德的推論にかかる議論にも有益であると考えられる点を扱う。第4節では20世紀の形式化の試みにおける概念分析の精緻化について述べる。第5節では、道德推論の現状の課題と哲学的論点との関連を述べる。

2. 事実命題と価値命題

ヒュームは因果にかかる帰納的推論の分析を行う中で、事実と価値は異なる体系にあることを主張した。

道德的状态にかかる記述と道德的勧告は、物理学における実験データと理論的予測に類似しているように見える。だが、物理学とは異なり、道德的状态の記述データは現状を反映してはいるが、この現状が道德的に正当化されるとは限らないため、これらのデータを基にして勧告を行うことはできない。

では、道德原則（例えば法律の条文）を公理のように扱って、道德推論を行えば定理として韓国が得られるかと言えば、そうではない。道德原則は特定の状況・文脈に依存しないものであるからこそ原則としてのステータスを持ちうるが、文脈に依存しなければ価値判断は不可能だからである。

したがって道德推論で可能となることはせいぜい、状況に付随する別の記述を結論として得る程度のことであり、論理的であればこそ、結論に新しい情報は付け加わらない（もっとも、数学の定理のように、人間側が見落とししていた記述をあたかも新たな発見とみなす可能性はある）。

3. 形式言語に適した命題概念

20世紀前半に、論理的推論を形式言語で記述する試みが始まった。この時のターゲットは数学的推論の形式化であり、演繹的推論を形式化することにより、演繹的に導出可能な命題の集合（定理）と真である命題の集合との関連を明確にしようとしたことだった。したがって当初はこの企てに帰納的推論が含まれなかったし、扱う命題も数学的命題を範型に、日常言語で記述される際に現在時制直説法となる文のみが理論化の範囲とみなされた。だがこの数学的推論の形式化の企ては同時に、日常言語による推論をも形式化する動機づけとなった。そこで扱われたのが人間が自然言語の中で行う推論であったが、数学的推論の形式化を目指したラッセルとホワイトヘッドの「プリンキピア・マテマティカ」の体系が不適当なのは明らかである。この延長上に構築されたのが、のちに部分構造論理と呼ばれることになる論理体系群であり、道德推論の演繹的形式化体系はいずれもこの類だ。

このような反実的状况を表現する仕組みの一つが「可能世界」と呼ばれるものである。この考え方によって、価値評価について演算子として形式言語に含め、評価としてあらゆる世界からなんらかの「道德的理想世界」が想定されているというモデル化が行われたのが義務論理体系である。命題演算子「…ことは道德的義務である」が適用された命題（義務）は評価点から到達可能であるすべての道德的理想世界で命題「…」が真であるときに真とされる。これにより、義務と許可の双対性が明示されたが、ベースとして古典命題論理を用い、しかも論理的真命題がすべての道德的理想世界で真であることから、

「論理的真は道徳的義務である」という直観に反する帰結が得られる。

4. 事実と行為、

「論理的真は道徳的義務である」という道徳推論の数理モデル化における不適を解消するために、(1) 古典論理を用いず、部分構造論理を用いる (2) 論理的・数学的真理が成立しないような「不可能世界」を想定する (3) 命題演算子の真理条件を変更する、などのテクニカルな解決策が提案されたが、その背後にある概念的提案のうち重要なのが事実命題と行為命題の概念的区別であった。

道徳的義務など価値判断の対象となるのは事実ではなくあくまでも行為であると考え、例えば論理的・数学的真理が価値評価を受けることは不適切である。また、事実命題に関しては古典論理を適用しても構わないが、行為に関しては事実命題とは異なる論理的關係を考える。行為命題間の関係や行為命題と事実命題の関係を表現するために行為命題にフラグとして命題演算子をつける。行為命題が真であるための条件は、その行為がどのように評価されるかによっていくつかのタイプに類別されるため、この点に関しては哲学的分析が下処理として必要となる。ひとたび行為のタイプ付けができれば、どのような状況下でその行為が成立するか記述可能となる。

そのうえで、特定の文脈で成立している行為について価値評価を付与することが可能となる。価値評価のシステムとしては、ゲーム論の方法が適用可能である。

5. 自動化された道徳推論と展望

昨今の帰納的・統計的推論のブラックボックス状況に関して、状態のシステムへの信頼が実質的問題となっている。見たこともない相手から突然「こうすべきだ」「こうなっている」という勧告が得られたとして、それを受け入れるためには少なくとも信頼醸成にかかるプロセスが必要であろう。たとえばその勧告の背後にある前提と推論手段を検証したり、その相手のプロフィールを調べたりすることになる。このために説明能力を必要とすると言うのが主流の議論であるが、この点に関しては議論の余地があるように思われる。というのも、人間の理性能力・行為主体としての能力の根幹に責任能力を前提する際に、言語的説明能力を養成しているからだ。だが、これは人間と機械のインタフェースの問題である。理性的かどうか判断する人間にとって解読可読な情報提示の方法が自然言語によるものであり、情報の受け手に対して説得力を持つ必要があるという前提は、人間のみが行為者でありえた状況下においてのみ成立する、機械による「説明」ないし情報処理の在り方は自然言語によるものではない。非言語情報による推論が可能であるという方向に推論の理論を拡大すれば、機械と人間の双方を要素とする分散システムによる情報処理を一元的に扱うことができる可能性があり、そこでの情報伝達のありようは言語に限定される必要がない(Barwise-Perry, Barwise-Seligman)。

演繹的道徳推論の限界は、ほかの形式言語による演繹的方法と同様に、きわめて厳密な言語使用に対応したモデル化であるだけに、我々が日常生活で用いているようなあいまいな言語使用を反映しにくいことによる。われわれが「よい」というときにはどのような意味で言っ

ているのかは文脈により大きく異なるので、形式言語による記述以前に、日常言語表現の下処理の必要があり、演繹的方法を適用するにはこの部分がボトルネックになっていると考えることができる。

帰納的道徳推論は本質的に、過去の価値判断に関する記述を参照しながら推論を行うことになる。この過去の価値判断データが道徳的に正当化されるのか、という点に関する検証を伴わない限り、人工因習ともいべき既存の差別の強化は必至である。このような価値判断データのクリアランスに関する研究が今後期待される。

さらに、道徳的予測に相当する勧告については、前提条件の記述方法およびどの種の義務概念などの価値判断基準を用いるかによって、前提の解釈のずれが発生し、推論が正しく行われたとしても結果となる勧告そのものが不適切なものとなる可能性が大きい。

これらの考察をふまえると、帰納的道徳推論の下処理として個々の前提集合に対して正当化可能性を検証するために演繹的方法を用いる可能性が開ける。すなわち収集されたデータについて道徳推論を適用し、不適切な結果をもたらした場合にはデータを廃棄する方法があり得る。もっともこのような前処理は、社会的に受容可能な道徳推論システムのごく一部をなすものであり、研究開発の初期の段階から、システムの各階層において道徳的正当性を検証しながら開発を進めなければいけない。

参考文献

- Richardson, Henry S., "Moral Reasoning", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral/>>.
- Henderson, Leah, "The Problem of Induction", The Stanford Encyclopedia of Philosophy (Spring 2019 Edition), Edward N. Zalta (ed.), forthcoming URL = <<https://plato.stanford.edu/archives/spr2019/entries/induction-problem/>>.
- Jon Barwise and John Perry, Situations and Attitudes, CSLI, 1999.
- Jon Barwise and Jerry Seligman, Information Flow: Logic of Distributed Systems. Cambridge UP, 1999.