

合意による道徳を擁護する二つの方向性

杉本俊介(大阪経済大学)

デイヴィッド・ゴティエが主著『合意による道徳』(1986)において Why Be Moral?問題に挑戦してから30年以上が経った。この本のなかでゴティエは、合意による道徳を構築しようと、(1)道徳から自由な領域、(2)相対的譲歩のミニマックス原理、(3)ロック主義的但し書き、(4)制約された最大化、(5)アルキメデスの点、という5つのアイデアを中心に議論を展開している。

合意による道徳では、完全競争市場が道徳から自由な領域とされ、この市場が失敗することで道徳が合理的な応答として導入される。すなわち、道徳原理(相対的譲歩のミニマックス原理)が「合理的な諸個人のあいだでの完全に自発的な事前合意の対象として」導入される。しかし、なぜ合意した後もこれらの原理を合理的な諸個人は受け入れる必要があるのだろうか。これがゴティエにとっての Why Be Moral?問題である。ゴティエはこの問題に対して(4)制約された最大化というアイデアに訴えている。

ところが、このアイデアには多くの疑問が投げかけられてきた。ゴティエが置いている諸々の前提は現実に成立しないのではないか。ゴティエの非標準的なゲーム理論をどう理解したらよいか。制約された最大化は本当に合理的か。制約された最大化を循環せず定義できるか。他のタイプの性向も比較検討する必要があるのではないか。そこで、「ゴティエの仕事を批判する文献のあいだで支持を増やしていることがあるとすれば、この提案[制約された最大化]が数多くの理由で失敗しているという点である」とも言われてきた(Buchanan [1990] 239)。もはや合意による道徳を擁護することはできないのだろうか。

私の知るかぎり、合意による道徳を批判的に継承する試みは二つの方向で行われている。一つは、ピーター・ダニエルソンの『人工的な道徳』(1992)に代表される機械道徳への実装化の方向性である。ダニエルソンはコンピューター・プログラムによって制約された最大化を実装し、ゴティエの議論の前提や非標準的なゲーム理論がどこまで現実的で説得的かを検証している。最近では W・ウォラックと C・アレンの『ロボットに倫理を教える』(2009)がこの方向性を推し進めている。

もう一つは、マイケル・トンプソンの『生と行為』(2008)に代表される行為論的な方向性である。トンプソンはゴティエが制約された最大化を一種の「性向」(disposition)として捉え、その性向が合理的であることからそれに従った行為が合理的であるとしている点に注目する。デレク・パーフィットが合理的不合理性という反例を挙げて問題にしている点である(Parfit [1984] 12-13)。だが、トンプソンによれば、この種の二層理論での〈性向と行為〉は規則功利主義の〈規則と行為〉やジョン・ロールズの「二つのルール概念」の〈実践と行為〉とも共有する形式をもち、彼独自の「素朴な行為論」によって適切に捉え直すことが可能だとする。

本発表ではそれぞれの方向性を Why Be Moral?問題への解決という点から検討していく。まず、合意による道徳の概要と制約された最大化に対する問題点を示す。次いで、ダニエルソンやそれに関連した機械道徳への実装化の試みを紹介し、それが1回限りのゲームという考えを放棄した点で Why Be Moral?問題の解決としては役立たないことを論じる。さらに、トンプソンの「素朴な行為論」を紹介し、それが性向に着目することで1回限りのゲームという考えを保持し Why Be Moral?問題の解決として魅力的であることを明らかにする。そのうえで、この方向性における課題を示したい。