トロッコ問題からロボットとの共生道徳への第一歩

トロッコ問題からロボットとの共生道徳への第一歩

柴田正良：金沢大学名誉教授

「道徳的行為者のロボット的構築による＜道徳の起源と未来＞に関する学際的探究」
（基盤研究(A) 19H00524）による研究会
場所：中京大学名古屋キャンパス
Aug. 7, 2022

1．普遍的観点（神の視点：「より良い」可能世界を求めて）

この「より良い」という基準が何であれ、普遍的観点を採るならば、ある行為の「正しさ（正当性)」の評価は、その行為がなされた可能世界全体を「より良い」ものとするか否かにかかっている。その際、行為者が誰であろうと、また行為者の立場がどのようなものであろうと、その行為がなされたか否かだけが問題である。

例えば、その基準が功利主義によって与えられた場合、（単純な「トロッコ問題」の想定では）５人より１人だけ死亡する方が、「快さの量」において「より良い」であろうから、現場に居合わせて線路のポイントを切り替えることのできる人は、誰であれその行為において正当化される。ここで比較されているのは、トロッコがそのまま走って５人が死ぬＡ世界全体の「快さの総量」と、ポイントを切り替えて１人だけが死ぬＢ世界全体の「快さの総量」である。

この普遍的観点は、言わば「神の視点」からの評価なので、行為者があなただろうと他人だろうと、あるいは行きすがりの観光客だろうと、その鉄道会社の従業員だろうと関係がない。要は、誰による行為だろうとそれが世界全体を「より良く」するか否かである。したがって、この「より良さ」を実現するためなら、あなたが線路に身を投げようと、近くの他人を線路に突き落とそうと、選ぶところはない。

これは、およそ受け入れがたい議論だろうか？

2.　個別的始点（古典的自由主義者たちの視点）

もしも道徳原理が一つではないなら、あるいは可能世界相互の「より良さ」の比較が通約不可能な複数の基準によるものならば、当然のことながら、トロッコ問題に関しても「唯一の正しい」解というのは存在しないだろう。あるいは、トロッコ問題のジレンマの異様さは、普通われわれの目から隠されている複数原理の不一致がこの問題においてはあからさまに強調されるから、余計に感じられるのかもしれない。

いずれにせよ、以下では、功利主義や義務論などの普遍的観点相互の争いを辿るのではなく、個人的観点からトロッコ問題へのアプローチを考えてみよう。その際、個人的観点としては、自分の「自由」こそが最高の価値であるとする一種の古典的な自由主義を前提にしよう。ここでの「自由」とは、ある意味で利己的な欲求実現を意図通りに実行する自由のことであり、これが各行為者の行為の始点である。すると、道徳的な考慮は、他人の自由、つまり他人の欲求実現との衝突とその調整の必要のゆえに初めて出現することになる。

さて、このような（ロボットを含む）自由主義的な（ある意味で利己的な）行為者にとって、本来的に、あるいは存在と共に、課されている義務というものはあるのだろうか？　唯一あるとすれば、J,S.ミルがかつて唱えた「他者危害の原則」（Principle of haram to others）を少し読み替えたくらいのものであろうか。この原則は、公権力が個人の自由を制限できるのは（その人以外の）他人に危害が加えられるのを阻止す

る場合に限られる、と主張する。「公権力」を「行為者」に読み替えて敷衍すれば、それは、「どの行為者も、他者に危害が加えられるのを阻止する場合を除いては、他者の自由を制限することはできない」、ということになる。この主張を＜「他者の自由を制限することができる」→「他者危害がある」＞と単純化して表現し、論理的同値であるその対偶を取れば、＜「他者危害がない」→「他者の自由を制限することはできない」＞となる。元の形に沿って言い直せば、「他者に危害が加えられるのを阻止するのでないならば、どの行為者も、他者の自由を制限することはできない」ということだ。他者に危害を加えていない行為者の視点からすれば、これは、「他者に危害を加えていないのだから、自分の自由は、他者から制限されない権利をもつ」ということになる。そして、これこそが、古典的自由主義の真骨頂である「愚行権」だ。愚行権は、危害が他ならぬ自分自身にのみ生じる場合は、いかに浅はかなことであろうと、その行為をなすことが許される、と高らかに（？）宣言する。「放っておいてくれ、オレがタバコを吸っても誰にも迷惑はかからないよ」という主張が真なら、この人の喫煙は道徳的に許されるし、仲間は「タバコは健康に良くないぜ」と説得を試みることはできても、「喫煙は道徳的に許されない」とは主張できない。「喫煙」が「自殺」でも、本質的には同じである。

したがって、愚行権は他の行為者の自由を「許容性」の点で最大限に尊重する一つの形だが、義務の形で他者の自由を最大限に尊重する主張は、この古典的自由主義の立場にはないのだろうか？　それは、「他者の自由が侵されている場合には、可能な限り全力でそれを救え」と各行為者に命ずる義務だろうが、それはカントの言うような意味での「完全義務」だろうか？　自分の自由の確保を最も重視する行為者は、同じ道徳共同体のメンバーから、自分の自由を最大限に尊重してもらうことを望むだろう。メンバー間での「基本的に同一の義務と権利」の所持という公平の原則からすれば、メンバー相互の自由の尊重は、例外なしにすべてのメンバーに課される義務だと考えてよいだろう。もしこの「互恵的な自他の自由尊重」が道徳共同体の存立条件から演繹されるのならば、これが唯一、「完全義務」と言えるのかもしれない。その導出の試みは興味ある皆さんにお任せするが、しかし、「完全義務」そのものは無理に望むほどのものでもないだろう。むしろ、濃淡・強弱の差こそあれ、すべてが「不完全義務」、つまり「努力義務」だと考えた方が、新しい道徳規則、新しい道徳共同体の創作という点からすれば好ましいのではないか。それでなくとも、ロボットはいざ知らず、われわれ人類は、あまりに多くの本能的な「義務傾向」を進化の末に受け継いでいるのだから。

「他者の自由の危機をできる限り救え」という義務は、たとえ完全義務ではないとしても、最もそれに近い不完全義務、すなわち最も強く推奨される努力義務であろう。「危機にある自由」を「危機にある命」と置き換えれば、その自然さはさらに納得がいくかもしれない（ロボットの場合は、「命」ではなく「複製不可能な個性」かもしれないが）。すると、これより弱い形での他者の自由への関与は、せいぜい「ある程度は推奨される」という範囲の事柄であり、それらを無視しても非難されずに「許される」ことになる。

例えば、極端に言うと、現在のわれわれの感覚からすればかなり受け入れがたいことだろうが、自立した我が子を他人と比べて特別扱いする義務はないし、路上で苦しんでいる浮浪者に手を差し伸べる義務もない。また、自分が属している会社や学校などの集団、及びそのメンバーが窮地に陥っていても自分の時間や財力をそのために用いる義務もない。それらにどう関与するかは、誤解を恐れずに言えば、究極のところ、行為者の「趣味の問題」である。

3.　ジレンマの種と構造

そこで、行為者のこの（強くはあれ）「努力義務」がトロッコ問題においてどのようにジレンマに陥るのかを、典型的な幾つかの要因に的を絞ってスケッチしてみよう。安心してもらっていいが、この問題に「正解」はない。示すことができるのは、ジレンマに「馴れる仕方」だけである。しかし、その前に、この問題に直接は関わっていない幾つかの、われわれの「直観的な義務傾向」を確認しておこう。

第一に、実際のところ人類は、「利他的傾向」をかなり強くもった祖先の末裔だということを示す進化遺伝学の見方がある。それを筆頭に、かなり多くのことが遺伝的に、つまり本能レベルで仕込まれているように思われる。例えば、自分の命と引き換えに我が子をひたすら守ろうとする遺伝的傾向の持ち主は、それをもたない遺伝的傾向の持ち主よりも、自分の遺伝子を後世に伝えるチャンスが大きいだろう。自然淘汰を生き延び続ける血統には、それに貢献したはずの多くの傾向性が本能レベルに存在する、と言うことができよう。これらの傾向性の中には、美しさと健康を備えた配偶者や、肉体的な強さを持った配偶者を求める、というものもあるだろう。また、近親相姦をぜひとも避ける、というものもあるに違いない、等々。

そうした生物的な傾向性に加えて、「嘘をつかない」とか、「約束を守る」とか、「誠実である」とか、「家族の絆を大事にする」とか、「友情を裏切らない」とか、「自分の属する集団のルールに従う」といったような、枚挙にいとまがないほどの数多くの、尊重すべき社会的な傾向性もあるだろう。これらは、人間が社会的な存在であることの証でもある。

それらとは別に、トロッコ問題では、「傍観者と未必の故意」の要素がかなり深く絡んでいるように思われる。哲学的行為論のややこしい専門的議論に入り込まないこととすれば、行為するとは、通常、自分の身体を動かすことによって世界の一部に因果的変化を引き起こすことだ、と理解されている。したがって、典型的には、「もし私があのように行為しなかったとすれば、あのような結果は生じなかったはずだ」、という直観、すなわち「自分は結果に対する必要条件の（決定的な）一つだ」という実感を、行為者はもっているに違いない。行為は行為者にとって格別の出来事であって、それ以外の世界の因果的過程は、傍観者としての行為者の前で勝手に進行していく絵巻物にすぎない。そして、傍観者は通常、「仮に私がどのように行為したとしても、あの結果は生じたはずだ」という実感をもっている。夏の大雨に打たれるがままの私は、どうやって雨が降るのを止められようか？

しかし、出来事の推移を傍観していることは、ある場合には、「何もしない」という立派な行為をなすことに他ならない。友人がいじめられているのを傍らで眺めていただけの私、電車内で暴力がふるわれれているのを止めずにいた私、また、会計の誤魔化しをわざと見逃した監査役の私。これらは、何もしないことが何かをなしている場合の卑近な例であって、そこには、結果を生じさせてもよいとする「未必の故意」が働いていた、とされるだろう。こうした場合、傍観していることの方が、身体的には「何もしない」のだから、新たに因果過程に介入するよりは楽だろう（後で咎められても、「えっ、そうとは知らなかったので」と言い逃れもできる？）。つまり、傍観するのではなく、そこで新たに行為に着手するには、一般に、「あらかじめ準備した意図的行為」よりも大きな心的コストがかかると思われる。傍観の態度から一転して新たに行動を起こす場合、傍観者は、不意を突かれてそうなることが多いだろう。しかし、出来事の自然な流れに任せていれば、受け入れがたい結果が生じる。未必の故意が生じうる所以だ。そこで、行為者は意を決して、結果を変えるために、この予期せぬ時点で行為に踏み切る。したがって、多くの場合、仮にあらかじめ準備された意図

的行為の結果と未必の故意による結果が同じものだとしても、未必の故意の方が軽い責任で済むのではないだろうか？

3. トロッコ問題に直面して：

（1）数の問題

あなたは通りすがりの観光客で、５人を救い１人を犠牲にするために、ポイントを切り替える。それでも、あえて出来事の流れに介入したことで、多くの人はこの行為に抵抗を感ずるだろう。

では、これが 50 人を救うための何らかのスイッチの切り替えだったら、どうか？　あるいは 30 万人を救うためだったら？　救われる人数が多くなれば、この行為は正当化されると感じられるだろうが、では、何人を救うことならいいのか？

逆に、あなたは通りすがりの観光客なのだから５人を見殺しにしても、許されるように感ずるだろうが、見殺しにされる人の数が、50 人なら？　あるいは 30 万人だったなら？　明確な答えはないだろう。

もっと別の状況では、あなたは、１人を救い１人を犠牲にするために、「特別の理由なしに」ポイントを切り替える。多くの人は、この行為に抵抗を感ずるだろう。しかし、あなたの自由な決定によって見殺しにする対象を変えることは、ただそのまま見殺しを放置することよりも悪いことなのだろうか？　ここにも、明確な答えがあるとは思われない。

（2）義務（約束）の問題

あなたはその鉄道会社の従業員で、「事故の際はできるだけ多くの人の命を救うこと」、というのが会社との契約の一部である。これに従って、あなたは１人を犠牲にして５人を救うためにポイントを切り替える。あなたが自分の行為を正当化するための義務を負っていたことに、多くの人は安堵するだろう。自分の手で無実の人の命を奪うことに怖じ気づき、あなたがこの義務を果たせなかったとしたら、１人の命を救ったにもかかわらず、あなたはどれほどの罪に問われるのだろうか？　それは、他の業務命令を単なる怠慢によって履行しなかった場合と、どれほど違うべきなのだろうか？

あなたは、会社との通常の契約に加え、副社長の命の警護という密命も同時に受けていた。そして、ポイント切り替えの直前に、業務命令の履行によって犠牲になるのはその副社長だと分かった。あなたは、業務命令を優先させるべきなのか、あるいは密命を優先させるべきなのか？　この義務や約束の葛藤に、あらかじめ会社との間で決着がついていないなら、明確な答えはないように思われる。

（３）近親者たちの問題

偶然にも現場を通りかかったあなたは、５人のうちの１人が自分の息子であるのを発見した。あなたは、息子の命を救うためにポイントを切り替える。多くの人は、息子を思うあなたの行為に同情するだろう。もし息子が５人ではない方の１人であったら、あなたは「何もしない」ことによって息子の命を救うだろう。この場合の方が、多くの人の共感が得られるかもしれない。

５人のうちの１人は息子などではなく、かなり縁遠い関係者、あるいは道で挨拶を交わすだけの知り合いであったが、あなたはその人を救うためにポイントを切り替える。近親者であることの程度が下がっていけば、多くの人は、そ

れがあなたの行為を正当化することは少ないと感ずるだろう。だが、どれくらいの関係性なら、一般に「さもありなん」と認められるのだろうか？

あなたは、鉄道会社の従業員で、例の業務命令と密命を受けているが、いま現場にいる１人はあなたの息子である。そこで、あなたは息子の命を救うために業務命令を無視し、ポイントを切り替えない。あなたの行為は、どれくらいの賛同を得られるのだろうか？　さらに、よく見れば、息子に見えたのは副社長で、息子は５人の中にいる。あなたが「業務命令を尊重しかつ息子の命を救う」ためにポイントを切り替えるのと、ポイントを切り替えずに「息子の命を犠牲にするのみならず業務命令に反するが密命には従う」のと、どちらが正当化されるのだろうか？　ここにも明確な答えがあるようには思えない。

あるいは、通りすがりのあなたは、たまたま見つけた近親者を救うためにポイントをそのままにするが、犠牲になる人の数が５人ではなく、50 人ならあなたの「何もしない」という行為は非難を免れるだろうか？　あるいは、それが 30 万人だったなら？　近親者の範囲と犠牲者の数との間に、明確な比較関係などあるのだろうか？　そのようなものがあるとは思われない。

（4）マインドコントロールなどの問題

あなたはある宗教団体から強力なマインドコントロールを受け、「ポイントを切り替える」という行為そのものが自分を地獄に突き落とす、と強固に信じ込んでいる。いま、あなたは問題の１人が現場を離れ、ポイントを切り替えても誰も犠牲にならないことを知っているが、５人どころか 50 人が死ぬことになろうとも、ポイントを切り替えない。その中に自分の息子がいても同じだろう。あなたが「何もしない」理由は、ある意味で利己的な非合理的信念だと言えるだろうが、あなたを非難すべき理由は、正確には何なのだろうか？　その理由は、鈍感な独裁者に対する断罪理由とどこがどう違うのだろうか？　その独裁者は、一見して病的ではないが、実は利己的な欲望と勝手な誇大妄想によって他国を侵略し、何百万もの人間を殺戮しても意に介さない。宗教に由来する行為への負の傾向性･･･

もちろん、トロッコ問題にまつわる道徳的考慮のリストはこれで完全なものではないが、教訓は明らかである。それは、さまざまな道徳的考慮に優劣の順序をつけてジレンマから完全に逃れることはできない、ということだ。大もとの道徳原理にしてからが、互いに通約不可能な仕方で複数存在する。加えて、それらの一つの原則に属する複数の考慮の間ですら、多くの場合、キレイな優劣順序をつけることができない。しかも、決定的に重要なことは、トロッコ問題のように劇的な芝居仕立てではなく、ごくありふれた日常の道徳問題においても、同じようなジレンマ構造があるということだ。このジレンマ構造は、たまたまわれわれの中の様々な道徳的考慮、雑多な道徳的傾向性が表に出てこないときには、われわれの目から隠されているというにすぎない。

つまり、基本的には、いかなる道徳的課題もジレンマ構造をもっている、と考えるべきなのだろう。したがって、ロボットと共生する道徳において、たとえ最も重要な規範を「他者危害」の原則だけに置くようなシンプルな道徳システムを構築したとしても、それ以外の無数の道徳的考慮、道徳的傾向性の間での調整が不可欠である。しかし、悪いことばかりではない。そこには、超越論的な根拠づけも、自然科学的な還元も、宗教的な強迫もないのだから、オープンな議論が可能であろう。そしてそれらの道徳的提案は、暫定的なルールや法律の形であらかじめ明瞭に示すことができる。

ここから先の道のりは終わりなき提案と修正と再提案の連続であろうが、ロボットや人類を含めた＜われわれ人間＞の光溢れる希望の船旅となることを信じたい･･･