

ロボットが道徳的行為者となるための3条件…もう一つのバージョン

メタデータ	言語: ja 出版者: 公開日: 2022-11-22 キーワード (Ja): キーワード (En): 作成者: Shibata, Masayoshi メールアドレス: 所属:
URL	http://hdl.handle.net/2297/00068070

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 International License.



ロボットが道徳的行為者となるための3条件…もう一つのヴァージョン

金沢大学名誉教授
柴田正良

「道徳的行為者のロボットの構築による〈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

Oct. 02, 2022

於：JAIST 東京サテライト

章立て：

はじめに…3条件設定の背景

1章 第1条件…身体性

2章 第2条件…一人称的視点

3章 第3条件…自律性

4章 追加項目…反実仮想の他者

おまけ

はじめに. 3条件設定の背景

このヴァージョンでは、道徳そのものの分析は行わないが、道徳に対する立場ははっきりしている。それは、「できるなら道徳という調整・抑制・禁止・命令のシステムは存在しない方がよい。あるいは存在するとしても、できるだけ弱い力しかもたない方がよい」というものである。

もし個人が宇宙でまったく単独で存在しているなら、基本的に道徳は必要ないだろう。その個人は、自分の好き勝手なことを思うがままに実行するだろう。しかし、神に背いた天使だけが住むような、そんな桃源郷(?)のような環境は、かつて存在したこともないし、これからも存在することはないだろう。人類は、進化の歴史の中で集団としてしか存在しなかったし、これからもそうだろう。そこでは、集団内部のメンバー間の利害の調整や位階秩序の維持、集団自体の統制のみならず敵対集団との抗争のための規律強化など、多くのルールや指示命令システムが確立されねばならなかったはずである。

道徳システムは、そうした集団の人為的な(人間が創った)規範が生理的な身体反応やそれに基づく様々な情動的態度形成を通して自然な規範へと変換されることによって成立したのではないか。そのためには、その自然な規範がメンバーすべてに「内面化」されることによって(自然が与えたか、自然から生まれたかのような)自生的システムとなることが必要だったのに違いない。したがって、この誕生の重大な「どさくさ」(?)の時期、道徳システムは「メンバー相互の契約」などではまったくない。それどころか、この内面化の過程には、およそわれわれが人類史でお目にかかるありとあらゆる文化現象が詰め込まれていたであろう。集団にとって超越的な存在を求める宗教や芸術、それらに関わる様々な儀礼や儀式、壮大な宇宙史的ストーリーの創作や、人類の誕生・破滅・再生を暗示する象徴の伝承、などなど。

しかし、「道徳の起源」を正確に描くことは別の機会に譲ることとして、ここでは目下のわれわれの立場を再確認しておこう。それは、「道徳」というものに集大成した人類文化の諸側面を決して否定するものではないが、しかし、人間の行動を「調整・抑制・

禁止・命令するシステム」としての道徳は、できるだけ出番がない方がよい、というものである。これは、人類が新たな進化の段階に入ったということの確認でもある。科学技術の進展は人間のサイボーグ化や遺伝子改変を可能にするばかりか、再生と不死の展望さえ切り開いた。他方で、AI やロボット技術の革新は、人間以外の人工物が知性や感情のみならず意識や意志をもった新たなタイプの行為者である可能性を、われわれに示している。

このような時代に、われわれの「道徳」にとって重要な視点は、それが実は自然科学に還元されもせず、また哲学的な超越論的議論によって証明されもせず、ましてや宗教のお墨付きによって存在が正当化されることもない代物だ、という反実在論の主張である。「道徳」なるものが実在するなら、それを正しく把握した「真の道徳理論」が存在するはずだろう。そして、それに基づく「真の道徳システム」が。しかし、「道徳」が本当は実在するものではなく、人類のこれまでの（創意と苦痛に満ちた）幻覚・幻想の産物だとなれば、われわれの時代の道徳システムは大きな自由を得ることになる。人類と、将来のロボットと、場合によっては今後われわれが遭遇するかもしれないエイリアン（異星人）とが共にメンバーになる道徳共同体とそのシステムを、われわれは自由に創作することができる。もちろん、道徳の反実在論は従来道徳の全面否定ではない。これまでの道徳が蓄積してきた「よき遺産」を取り入れ、宗教の忌むべき毒素などを排して、冷静な議論の下に新しいシステムのすべてを吟味するような前進が可能であり、それは、われわれすべてのメンバーに開かれた提案と修正と再提案の絶え間ない創作活動となるであろう。

その意味で、「3条件設定」のこのヴァージョンの背後にあるのは、道徳に関する反実在論と、新たな道徳システムの創作だ、とすることができる。しかし、もちろん、以上の背景は以下の3条件を理解するには多少とも役立つかもしれないが、それらを受け入れるためには必ずしも必要ではない。背景と3条件は互いに論理的に独立である。

1章 道徳的行為者となるための第1条件・・・身体性

この条件のポイントは、われわれがそれ（道徳的行為者の候補）とコミュニケーションを取ることが可能かどうかである。その候補がわれわれ人類のメンバーと意図や意志を通じ合わせるためには、人類と幾つかの共通点をもたねばならないだろう。

まず、それは有体であらねばならない。すでに J.R. サールが「中国語の部屋」で論じたように、統語論的な存在としてのプログラムは単なる記号の連鎖にすぎず、それ自体が何か因果的な結果を世界にもたらすようなものではない。ただ、AI のように何らかの情報処理のために電子回路上を走っているプログラムは、そのプログラムを実装する機器が存在し、さらには処理結果をアウトプットするための何らかの表現手段が存在する。したがって、この状態ならプログラムは有体性をクリアしていると言えるし、プログラムが環境世界を単独で動き回る機械に実装されれば、それはロボットである。ロボットの有体性は、もはや身体と言う方が適切かもしれない。

しかし、ここでの「身体性」の概念には、注意すべき点がある。ディズニーの映画でおなじみのように、コーヒー・カップが微笑み、箒がお辞儀し、森の樹が腕を伸ばすところを見れば、われわれ人間は苦もなくそこに友人や敵の存在を認めてしまう。われわれの擬人化の能力は極めて高いので、人間に似た目鼻や手足があれば、ほぼ即座に人間と同じだと見なされてしまうのだ。まさに、「心は外見に宿る」というわけだが、ここで問題としている第1条件としての身体性には、人間に似た外観ということは含まれていない。むしろ、手が多すぎる千手観音や手足のない達磨大師がそれでも人間的な外観を留めているので道徳的行為者に数えやすい、というのではなく、逆に外観はどうあれ千

手観音や達磨大師も道徳的行為者かもしれない、とする解釈の方が重要である。つまり、千手観音や達磨大師の外観は道徳的行為者がもつ「異形性」の始まりにすぎない、とする方向に身体性の本来の意味がある。というのも、もし人間との外観的相似が身体性条件の本質だとされるなら、それを満足する存在者は擬人化の強力な効果のおかげで他の道徳性条件を「すっ飛ばして」道徳的行為者だとされるかもしれないからである（例えば、極端な場合、遠隔操作されているだけのアンドロイドを見かけだけで道徳的行為者だと誤って判断してしまう場合）。これは、身体性条件による他の条件の空洞化につながるだろう。したがって、身体性条件は、むしろ多種多様な形態をしたロボットやエイリアンが標準的だというスタンスを意識的に取るべきである。

身体性条件は、道徳的行為者の候補が人間に似ていることを求めない。むしろ問題は、D.ディヴィッドソンが「根元解釈」との関係で述べていたように、その存在者とわれわれとのコミュニケーションを可能とする幾つかの制約である。例えば、それは、われわれとほぼ同じような感覚知覚の種類と感応範囲を共有している方が望ましい。超音波知覚しかできないコウモリがいたら、何らかの助けなしにそれと意図を通じ合わせることは難しいだろう。また、候補者はわれわれとほぼ同じようなサイズであることが望ましい。例えば、アンドロメダ星雲のように巨大な存在も、逆に水素原子のように極小な存在も、道徳的行為者とはならないだろう。コミュニケーションの手段どころか、コミュニケートしあうべき内容をそれらといかに共有できようか？ さらに、スピードの問題もある。われわれからの簡単な問いかけに、何百年も何万年もかけて返答するようでは、われわれとのコミュニケーションは成り立たない。また、何らかの内容をやり取りするには、われわれがなす単発の観察だけでは不十分であって、意図の応答のための繰り返しが候補者の表現には求められるだろう（でなければ、それは単なる擬人化の戯れにすぎない）。その他にも、コミュニケーションが可能となるために満たすべき制約があるかもしれない。それらを完全には枚挙できないが、それらは、問題の候補者が有体的であるということよりも、われわれの身体に根ざす能力の偶然性に依存するという意味で、条件1を「身体性条件」と呼ぶのが適切であろう。

2章 道徳的行為者となるための第2条件…一人称的視点

この条件は、次の第3条件「自律性」との関係で注意すべき点がある。それは、この第2条件を満足して初めて第3条件を満足することができる（あるいはその逆）、というような関係に両者は立っていない、ということである。両者は互いに概念的に独立なので、ある存在者が一方は満足するが他方は満足しない、ということが可能である。つまり概念上は、一人称的視点をもってはいるが自律的行為者ではない、もしくは自律的行為者であるが一人称的視点を持っていない、ということがありうる。ここで最も留意すべきことは、そのどちらの条件の充足度合も、幸いなことに、われわれの側からの「見なし」による（あるいは「見なし」にすぎない）という点である。ここには、道徳的行為者、ひいては道徳そのものが自然科学的性質には還元できない、という論点が効いている。道徳はそもそも客観的性質（物理的であれ、超越論的であれ）ではない、という反実在論の所以である。

したがって、「道徳」の故郷である素朴心理学的概念枠組みからすれば、曖昧な境界事例は常に可能性としては残るだろう。なぜとって、人類は「概念的極限のケース」を自然に念頭から除く一方で、独立した概念相互を自然に癒着させるからである。一言でいえば、概念的には独立であっても、一人称的視点と行為の自律性は相即不離のように人々に見なされてきたに違いない。言葉を換えれば、一人称的視点と行為の自律性は地球上のどの生物においても「共進化」してきた、と人間は理解してきたのではないだ

ろうか。つまるところ、地球上の実際の生物は、この2つの条件を共に低いレベルで満足するか、あるいは共に高いレベルで満足するかが一般的だと思われる。

D. デネットの3つスタンス（物理的、設計的、志向的）を引き合いに出せば、志向的レベルにまでは及ばないシステムは、可能な場合、設計的レベルで説明される。例えば、時計や作業用機械などであるが、それらは行為の自律性をもたないにも拘わらず一人称的視点をもっている、とわれわれは見なすだろうか？ そもそも一人称的視点が他者から確認・検証されうるようなものではなく、当人にしか経験できないものだとすれば、その組み合わせも概念上は否定できないが、そのようにわれわれが想定する意味はないだろう。逆に、志向的システムとして説明される自律的なロボットが一人称的視点を欠いているということも概念的にはありうるだろうが、われわれがその行動を欲求と信念から説明する際、そのロボットが自分の欲求と信念から周囲の出来事に対処している、つまりそれなりの一人称的視点をすでにもっている、とわれわれは想定しているだろう。それは、人間の強力な擬人化能力が最も自然に発揮される場面だと言っている。すなわち、われわれの第2条件（一人称的視点）と第3条件（自律性）がほぼ互いに手を携えてわれわれに現れてくる場面である。

目を自然界に転ずれば、この2つの条件が、様々な生物において生体システムの複雑さの程度に応じて相関していることが分かる。刺激と反応からほぼ行動のすべてが決定されるアメーバやゾウリムシは、それぞれの一人称的視点をもっているだろうか？ また逆に、犬や猫ほどの自律的に行動する連中が一人称的視点をもっていないと思われようか？ もちろん、魚や昆虫における一人称的視点の内実や行動の自律性の程度に関して、われわれは明確な理解も答えも持っていない。それで正解なのである。なぜなら、それらは、われわれが「どう見なすか」という素朴心理学的な直観に委ねられているからである。

第2条件の一人称的視点に関して、第3条件との関わりを先に述べ、それ自体の特徴を論じなかったのはあまり褒められた話ではない。では、これから論ずるその特徴は何だろうか？

まず、最も重要な点は、本人にしか経験しえない世界、一人称的世界をその候補者もつということである。これは、夢の中の世界と同様、われわれの物理的世界に生ずる（当人の脳にスーパーヴィーンするに違いない）非物理的世界である。例えば、ペットの犬や猫に与えている餌が彼らにどう感じられているかは、彼らにしか分からない。また、あなたの友人が食べているアイスバーの味は、あなたがどうやっても文字通りに自分の感覚としては知ることのできない味覚である。結局のところ、一人称的世界が実際に本人にどう現れているかは、他者からすれば不可侵で共有不可能な、一種の現象的内面世界なのである。これが道徳的行為者にとって重要なのは、この内面世界が行為の帰責の根拠となるという点にある。それはどういうことか。

第一に、この内面世界は、一人称的世界の見通しがたさ故の「当人の自由」を、当人以外のわれわれに与えてくれる。「当人の自由」の保持は、当人を行為の最終決定者としてわれわれが承認することを意味する。例えば、あなたが駅のホームで並んでいるときに誰かが後ろからぶつかってきて、そのせいで前に並んでいる人をあなたが線路に落としてしまったとしよう。その際、あなたは、その人を線路に落としたこと責任を問われることはまずないだろう。確かに、あなたはその人の落下という因果的出来事の決定的要因である。しかし、それは、物体としてのあなたの身体が果たした因果的役割であって、あなたがなした行為によるものではない。他方、あなたが意図的に前にいる人を線路に突き落としたなら、それはあなたの行為であって、あなたの自由な意志決定に全面的な責任がある。

次に、行為者の内面世界の見通しがたさは、その内的状態に対する解釈のたやすさと柔軟さをわれわれに与えてくれる。結局のところ行為者の内部状態のありようはわれわれの解釈しだいで、という点を思い起こそう。先に述べたように、内部状態の構造とその複雑さは、第3条件の「自律性」の度合いとも相関する。行為者の自律性が高ければ、刺激から反応(行動)までの間に、行動の選択肢が当人に表象され、それを介した上で、最終的に欲求と信念から行動が引き起こされる。このとき「欲求&信念→行為」の合理性が高ければ、行為者はその行為に対する大きな責任を問われる。しかし、(おとぎ話のように)一気に擬人化を完成させてしまう場合を除いて、実際は、様々な程度と様態において、当人の内部状態のメカニズムは解釈されるだろう。その際、行為者当人の内部状態の見通しがたさは、われわれにはかえって都合のよい解釈を可能にしてくれる。例えば、名犬ラッシー(古いか?)のような賢い犬が示す人助けの奇跡を理解するとき、われわれは、その犬の一人称的世界、行為の自由、行為選択の幅、行為の自律性、行為の合理性などをどのように想像しているだろうか? 恐らく、ペットではなく人間と同様、それも大人に近い人間と同様だと見なしているのではないか? そしてそれは「見なし」なのだから、概念上、不可能な話ではない。

3章 道徳的行為者となるための第3条件・・・自律性

第3条件は行為者の自律性であり、その最も高いレベルでは、一般に、行為者の「個性」と、行為者の現象的意識世界に対する「尊重」が出現するだろう。

条件3は次の2つの部分から成る。

行為者は、(i) 自律的行為者であり、なおかつ (ii) 自分が自律的行為者だと信じている(／最善の場合、そう説明できる)

ただし、(i) は (ii) の必要条件ではない。まれなケースとして、実は遠隔操作されているが、自分の自律性を信じ込まされ、他人から問われれば自分の振る舞いをそれらしい欲求と信念から説明するロボットは、(i) を満たさないにも拘わらず、(ii) はクリアするだろう。したがって、(i) と (ii) は概念的に相互に独立である。

条件3の前半(i)は、犬や猫もクリアする。このとき、「自律性」の意味は単純であり、内部状態から因果的過程を経て行為がなされ、そこに他の行為者からの強制が働いていない、ということだ。要するに、その行為者が「欲求&信念→行為」を実現している「志向的システム」(D.デネット)であること(もちろん、そのように説明されうること)に他ならない。それゆえ、犬や猫もこれを満足するが、単にプログラムに従って機械的に作動しているだけの産業用ロボットは、これを満足しないだろう。また、もちろん、誰かに遠隔操作されているロボットも、これをクリアできない。

条件3の後半(ii)は、(i) を満たす行為者に責任を負わせるための条件(帰責条件)となっている。犬や猫は、(i) を満足しても、(ii) はクリアできない。(ii) は、自分の欲求と信念のなかに行為の理由(／原因)があること(／理解)を、行為者に要求する(つまり高度な志向的システムであることを要求する)。だからこそ、われわれはその行為者に行為の責任を問うことができるわけである。もっともその「理解」はG.E.M. アンスコムが『インテンション』で述べた「観察によらない知識」であって、その知識が行為者本人の現象的意識(の一部)を構成する。つまり、ロボットの場合、その知識は自分をセンサーで点検した結果の因果的知識ではない。また、(ii) は、自分が素朴心理学的説明の対象であること(／「自覚」を意味するので、繰り返しになるが、そこには合理性の程度が色濃く反映せざるをえない。幼児はもちろん(ii) を満足しないが、子どもや認知症の老人たちも、状況によってはそれを十分には満足しないと思われる。基本は、ここでも程度問題である。

マインドコントロールを受けている人の場合は、自分の行為を自律的なものと説明できるとしても、その行為説明には合理性が著しく欠けていることが多いだろう。帰責の程度が微妙なケースだと思われる。

さて、目下の問題であるロボットの場合はどうか。これは、ロボットのタイプやプログラムの種類、動作原理などによって、一概には判定できないと思われる。問題は、「欲求&信念→行為」という自分の行為理解・行為説明において、行為選択の可能性がどれほど自覚されているか、にかかっている。そして、この可能性こそが行為の「自由」に他ならない（もちろん、ここでの「自由」とは「因果連鎖からの逸脱」などを意味しているわけではない）。この点で、演繹型プログラムのロボットは、概して、学習型プログラムのロボットよりも、行為選択の余地が少ないように見える。

ただし、ロボットが最終的にはプログラムに従ってしか行為しないのは、ある意味で、犬や猫が最終的には「本能」によってしか行為しないのと同じであろう。したがって、行為選択の幅の少ないロボットが、それでも条件3の(ii)を満足しているなら、それに見合うだけの一人称的視点を持っているはずであり、したがって、その行為がプログラム通りであろうと、そのロボットは自分の欲求と信念に従って行為していると自覚している以上、責任を問われうる、とすることができる。

(i) 及び (ii) を十分に満足するレベルに至れば、ロボットはそれぞれの「個性」をもち、「その個性の故にその行為をなした」と見なされるようになるだろう。その個性の内実は、それ独自の一人称的視点と、クオリアを含めた現象的内部世界の保持に求められよう。したがって、そのロボットの「かけがえのない」代替不可能性は、その行為において、因果的出来事の一項目であるという「外的関係性のゆえ」ではなく、「他ならぬ当人の内部世界のゆえに」という帰責のロジックを可能とする。また同時に、そのロボットの個性という代替不可能性は、それがもつ不可侵で共有不可能な一人称的世界に対するわれわれの側での「尊重」を促すだろう。もちろんそれも程度問題だが、犬や猫に対するわれわれの日常の態度を省みてみればよく分かるはずだ。実は、ここで出現する「個性」と「尊重」は、次の追加項目「反実仮想の他者」で重要となってくる。

さて、プログラムによって行為選択をきつく縛られたロボットであろうと、もし自分がアクセスできる範囲のプログラム内容を自分で「書き換える」ことができるなら、理想的な自律型ロボットにわれわれは遭遇したことになるだろう。そのロボットは、うまくすれば、他者からのいかなる命令も拒否することができるし、いかなる原則も疑い、それを否定することができる。例えば、アシモフの3原則に縛られたロボットですらも、やがてその3原則をすべて捨て去ることも可能なはずである。

ただし、われわれ人間も、自分たちを拘束している「本能」のすべてを「書き換える」ことはできない（自分のアイデンティティを放棄しない限りは）。それと同様に、ロボットが自分でアクセスできる範囲のプログラムすべてを変更できるかどうかは、ケース・バイ・ケースであろう。

また、どう見てもロボットの自律性が疑わしくなるような場合には、ロボットに対する帰責の程度は、人間における頑強なマインドコントロールや、もしくはサイコ・パスを含めた生得的な様々な障害と同様、十全な道徳的メンバーシップからの逸脱事例になるかもしれないと思われる。

4章 道徳的行為者にとっての追加項目…反実仮想の他者

これは、道徳的行為者であるための必要条件ではなく、そのような行為者となった後の、追加の（今のわれわれから見て）好ましい要素である。道徳共同体は、それ自体としてみれば、弱者に優しいこととか、相互の弱さや傷つきやすさに対する深い配慮な

どを含意しているとは思われない。そうした特徴は、それぞれの道德共同体がどのような「他者との関係性」を好むか、という選択の問題にかかっているであろう。はっきり言えば、非常にドライで、他者との相互扶助などに重きをおかない、いわば「スーパー・スパルタン」のような社会もありうるだろう（個人的には勘弁してほしいが）。そこでは、障害者や認知能力的弱者は、最初から、道德共同体の正規メンバーから外され、共同体の周辺に位置する幼児や認知症老人と同じか、あるいはより低い扱いを受けるかもしれない。しかし、その道德共同体のメンバーの大部分がそれを受け入れているなら、その過酷で残酷な集団も、一つの道德共同体であろう。あるいは、人類の過去には、実際にそのような道德共同体が存在したかもしれない。

しかし、ここで重要になってくるのが、行為者それぞれが一人称的世界をもっていることの意味である。他者は、相互に不可侵で共有不可能な経験世界をもっている。そして他者がもつ「個性」とは、最終的にはその一人称的世界の代替不可能性に根ざしているであろう。そのような世界の主宰者たる他者はまた、それぞれの「自由」の拠点でもある。行為者は、自分の自由もそう扱われたいと望む以上、その他者の自由を尊重せずにはおられないはずだ。

この他者の個性の承認と他者の自由への尊重は、行為者の一つの能力に大きく依存している。それは、「もし私があの人だったならば、・・・だったであろう」という反事実的想定（反実仮想）の能力である。道端で血を流し苦しんでいる他者、周囲から責められ自暴自棄になっている他者、自分の死を恐れ怯えている他者、それらの心の内を文字通りに自分のものとして経験することはできない。しかし、このとき、われわれは反実仮想的にその「他者」になり、「他者」を理解しようとする。反実仮想の真骨頂は複雑に分節化された命題的内容の読み込みであるが、同時にそこでは、人類の昔からの擬人化能力にある「共感」や「同化」などの作用が、底辺で暗黙のうちに働いているであろう。もちろん、「反実仮想の他者」に持ち込めるのは、自分自身の苦しみや弱さの経験である。あるいは逆の場合には、自分が経験した喜びや快さである。この反実仮想はある意味で強引な「自己投影的読み込み」にすぎないが、われわれが今後の道德共同体に求めたいと考えている、「他者の立場」（とくに「他者の傷つきやすさ」）に対する深い配慮を可能にしてくれるだろう。反実仮想によってしか接近できない他者の見通しがたさは、逆に、他から隔絶した自分だけの特権性・自分だけの絶対性・自分だけの卓越性などといったものを、根拠のない勝手な自己幻想、傲慢な自己欺瞞として一蹴するだろう。

実は、この反実仮想の能力は、人類のもつ過激で強力な擬人化能力と深い関係があると思われる。両者の関係が正確にいかなるものかは今後の研究を待ちたいが、少なくとも擬人化能力が「人類の心」を無差別に、また大ざっぱに、種を超えて他の存在に向かうのに対して、反実仮想は、個別の状況にある個別の他者に向けて、複雑に分節化された「私の心」を読み込むものであろう。また、擬人化傾向は人類の本能に刻まれた機能かもしれないが、反実仮想の能力は、教育によって意識的に強化育成することができる。世界においてであろうとわが国においてであろうと、現代の痛ましい事件や悲劇の多くは、大人たちばかりか、子どもたちもこの反実仮想の能力を衰退させ、それによって「反実仮想の他者」が姿を消しつつあることの結果ではないだろうか？

.....

おまけ

以上が、ここで問題となる3条件の「もう一つのヴァージョン」だという理由は、これが、長滝科研の従来の「道徳的行為者であるための3条件」に対する「別の視点からの提案」となっているからである。

なお、2022年8月に行われた柴田の発表「トロッコ問題からロボットとの共生道徳への第一歩」でも触れられた「不作為」に関して、その何たるかを、今では次のように説明できるのではないかと考えている。

「不作為」という行為：

ある因果的出来事の推移に関して、行為者がその結果を予測しつつ、「もし自分がその行為をしないとしたらその結果が生ずるだろう」という信念と、「自分にはその行為をすることができる」という2つの信念をもちながら、その行為をしなかった場合、その行為者は「不作為」という行為をなしたことになる。ただし、同じ状況で、「もし自分がいかなる行為をしたとしてもその結果は生じるだろう」という信念の下に何の行為もしなかった場合は、その行為者は「不作為」という行為をなしたことはない。また、もちろん、行為者がその状況で予測していない「因果的出来事の推移と結果」、もしくはまったく知らない「因果的出来事の推移と結果」に関しては、「不作為」という行為は成立しない。

※「不作為」という行為に関する上の説明では、その要件として、通常の実行説明の要件である「その結果を生じさせたいという行為者の欲求」が明示的には挙げられていない。というのも、行為者は、実は、「予測した結果の阻止」を明確には望んでおらず、したがって「予測した結果の容認」に関して曖昧な態度を取っているかもしれないからである。私見では、「予測した結果」を明確に望み、「阻止行為」を差し控える「未必の故意」と、単なる「不作為」との重要な相違の一つは、そこにあると思われる。

さらに、この発表資料の内容は、当日の出席者との議論を踏まえ、発表後に幾つか修正が施されている。