

令和元年6月12日現在

機関番号：13901

研究種目：基盤研究(B) (一般)

研究期間：2016～2018

課題番号：16H03341

研究課題名(和文) 科学に基づいた道徳概念のアップデート

研究課題名(英文) Updating the concept of morality based on science

研究代表者

久木田 水生 (Kukita, Minao)

名古屋大学・情報学研究科・准教授

研究者番号：10648869

交付決定額(研究期間全体)：(直接経費) 8,300,000円

研究成果の概要(和文)：本研究では「道徳性」について、様々な科学の分野で得られている知見と、哲学・倫理学分野で伝統的に受け入れられている見解との相違点を明らかにした。具体的には哲学・倫理学においては、個人の自律性や理性を重視した道徳性理解が主流であるが、実際には感情的・身体的要因、社会的環境的要因が道徳的判断において重要であること、それらと合理的要素が複雑に入り組んで道徳的判断が可能になるということが明らかになった。それを踏まえて、私たちは「機械倫理」と呼ばれる、人工知能やロボットに道徳的判断や道徳的行為を行わせる研究分野において、「倫理的な機械を作る」という問題のフレーミングそのものに困難があることを見出した。

研究成果の学術的意義や社会的意義

現在、「道徳の衝突」(異なる文化や宗教の間での道徳観の対立)が様々な問題を引き起こしている。一部の学者は「道徳は科学が扱うべき」と論じる一方で、「道徳は科学では扱えない」という意見もある。本研究は道徳性についての科学的知見が現在のところどのようなもので、いまだ科学で扱えていない部分はどこなのかを明らかにすることで、道徳の衝突の状況において有益な対話の基盤を提供することが期待される。

研究成果の概要(英文)：In this research, we clarified the gap between the findings regarding "morality" obtained in various fields of natural science and the views that have traditionally been accepted in philosophy and ethics. Specifically, in philosophy and ethics, the mainstream view emphasizes individual autonomy and reason, but science has revealed that the emotional and physical factors, as well as social and environmental factors, also play important roles, and that these factors and rational factors are intricately intertwined to enable moral judgments. Based on this, we found a difficulty in the very framing of the problem as "making machines ethical" in the research field called "machine ethics," in which they attempt to make artificial intelligent systems and robots that are capable of moral judgments and/or moral actions.

研究分野：哲学

キーワード：道徳性 機械倫理 人工道徳

様式 C-19、F-19-1、Z-19、CK-19（共通）

1. 研究開始当初の背景

(1) 心理学、進化生物学、ゲーム理論、神経生理学、人類学、動物行動学などの自然科学の分野で「道徳性」をテーマにした研究がここ数十年の間に非常に盛んになっている一方で、哲学や倫理学の分野ではこういった研究成果を反映した研究はあまりされていなかった。

(2) 2006年頃から人工知能やロボット工学の分野において、「機械倫理」や「人工道徳」と呼ばれる、人工知能やロボットに道徳的な判断や行為を行わせる研究が行われるようになっていた。ここでの道徳性や倫理の理解は、多くの場合、伝統的な哲学・倫理学における道徳性の理解に基づいていた。

2. 研究の目的

(1) 自然科学の諸分野で行われている道徳性についての研究を総合して、道徳性の概念を再考する。

(2) 新しい道徳性の概念に基づいて、道徳的判断や道徳的行為のモデルを作り、そこから「人工道徳的行為者」のデザインを考える。

(3) 道徳性に関して、科学で明らかになっていること、科学ではまだ十分に扱えていないことを腑分けして、道徳の衝突によって引き起こされる問題に対する適切なアプローチを考える。

3. 研究の方法

(1) 科学の諸分野での道徳性についての研究を調査する。文献を調べると同時に、様々な分野の専門家を招聘して現在進行中の最新の研究について情報提供をしてもらい、異分野の専門家の間で情報の共有とディスカッションを行う。

(2) 哲学・倫理学分野における道徳性についての理論を調査する。

(3) (1)と(2)の成果を踏まえて、哲学・倫理学の分野での道徳性についての理論と、近年の科学分野で明らかになっている道徳性についての知見とを比較・対照し、どのような食い違いがあるかを調べる。

(4) 近年の人工知能やロボット工学の進展を踏まえたうえで、道徳性のモデルを考え、人工知能やロボットに道徳的に判断・行為させるにはどのような実装の道筋が適当なのかを、考える。

4. 研究成果

(1) 近年の科学的知見と伝統的な哲学・倫理学的道徳理論との間に明らかになったギャップを以下に挙げる。

(1-1) 動物の道徳性：本研究において私たちが最初に考えた問いの一つは、道徳性が人間に特有のものかということである。西洋の哲学・倫理学の伝統では、道徳性を高度な理性的活動と考える傾向があり、人間以外の動物を道徳的配慮の対象と考えることはあっても、動物を道徳性の主体と考えることはなかった。しかしながら道徳性の期限と思われる共感、利他行動、協力行

動、不公平忌避などは多くの動物においてみられる現象であった。しかしながらそれが人間の道徳性、人間の倫理的行動と本当に比較しうるものなのかどうかについては議論の余地があると思われた。

(1-2) 子供の道徳性：従来の合理性に基づく道徳理論では、生まれたばかりの子供はもちろん道徳性の主体とは考えられない。情動の発達、知性の発達後に、社会の規範を学んでいくことで子供は道徳的主体になるのである。しかし心理学の分野での研究では、言語習得以前の子供でも、他者が協力的であるか否かなどについての判断を行っているということが分かっている。ここから道徳性にはほとんど生得的ともいえる要素もあるということが分かった。かといって教育を受けずに十分な道徳性を身に着けることは出来ないということも事実であるので、道徳性においてどこまでが生得的でどこからが獲得されるものかは今後の研究課題である。

(1-3) 情動か理性か：道徳的行動が情動に基づくのか理性に基づくのかという問題は道徳哲学においても長い歴史を持つ論争であるが、主流の理論においては理性の役割が強調されているこの問題に関しては、近年の自然科学者の間でも意見が分かれているが、道徳的判断・道徳的行為における情動の役割の重要性（そして理性の無力さ）が強調される傾向にある。一方で、心理学者のポール・ブルームのように理性の重要性（そして情動の有害性）を強調する学者もいる。少なくとも情動と理性の両方が道徳的判断や道徳的行為において何らかの役割を果たすということは言えるが、情動と理性がどのように相互作用しているかについてはまだ十分な研究がなされていない。

(1-4) 環境の影響：伝統的な道徳哲学においては人間が自律的な行為者であるということが前提されている。しかしながら人間はそれほど自律的に意思決定して行為をしているわけではないということを示す知見が、近年、心理学において豊富に得られている。テクノロジーの力が増大し、現代の社会においては私たちの行為がテクノロジーのサポートなしで遂行されることはほとんどなく、さらにテクノロジーはますます自動化の度合いを増している。このような状況においてこの問題はさらに深刻さを増している。このことは道徳的実践において賞罰といった慣行に大きな影響を与えうる重大性を持っている。このような科学的な知見をどのように社会に反映させていくべきかということは、これから倫理学者、法学者、工学者が取り組むべき問題である。

(2) (1) で明らかになったことに基づいて私たちは新しい道徳概念の構築と、道徳性のモデルの構築に取り組んできた。その結果、道徳的行為は「自律的で合理的な個人」に帰されるという従来の道徳哲学・倫理学において主流の捉え方は維持するのが難しい、と結論するに至った。特に現在では、複雑なテクノロジーによって思考や行為がサポートされ、あるいは「媒介」される度合いが大きい。行為は、その行為の主体と見なされる人間とテクノロジーの間の複雑な相互作用の結果として遂行される。またその行為は環境、状況、過去の経験などから（明示的にも非明示的にも）大きな影響を受ける。従って道徳的行為が個人の自律的で合理的な意思決定の結果として導かれるという考えは、現代において科学的知見や社会技術的環境とそぐわなくなっている。私たちはこれに代わって社会的な道徳性、あるいは技術によって媒介された（あるいは拡張された）道徳性という概念を提唱する。しかしその時、行為の責任という難しい問題が生じてくる。一般に道徳的行為（ここでは道徳的に良い行為と悪い行為の双方を指して「道徳的行為」と呼ぶことにする）は他者に対して利益や危害を与える、あるいは規範に合致したり違反したりす

るものである。それが他者に危害を与える場合、あるいは規範に違反する場合には道徳的な非難の対象となる。しかし媒介された道徳性において、誰が非難を受けるのか、誰が行為に対して責任を持つのかというのは難しい問題になる。部分的にはこれが理由で、哲学者と倫理学者は道徳性についての従来の概念に固執している。私たちは責任についての新しい概念と実践を今後、慎重に考えていかなければならないと考えている。

(3) 機械に道徳的に判断し行為する能力を持たせるという「機械倫理」の試みに対する私たちの研究の含意は以下のとおりである。当初は私たちは新しい道徳的判断・道徳的行為のモデルを考え、それを機械倫理のプロジェクトに役立てるという意図を持っていた。しかしこの研究機関において明らかになったことは、単独のエージェントに道徳性を帰することが、道徳性に関する古い考え（現在の社会や、近年の科学的知見にそぐわない考え）に基づいているということである。従って私たちは機械倫理のプロジェクトに対しては、「機械を道徳的にする」という問題のフレーミングそれ自体がミスリーディングなものである、ということである。私たちが目指すべきは機械と人間からなる複雑なシステムが全体としてできるだけ望ましい帰結を引き起こすような振る舞いを示すようなデザインを考えることである。

5. 主な発表論文等

〔雑誌論文〕（計6件）

(1) Arisa EMA, Hirotaka OSAWA, Reina SAIJO, Akinori KUBO, Takushi OTANI, Hiromitsu HATTORI, Naonori AKIYA, Nobutsugu KANZAKI, Minao KUKITA, Kazunori KOMATANI, AND Ryuichiro ICHISE, "Clarifying Privacy, Property, and Power: Case Study on Value Conflict Between Communities", Proceedings of the IEEE, Vol. 107, Issue: 3, 575-581, March, 2019. DOI: 10.1109/JPROC.2018.2837045. (査読有)

(2) 久木田水生「人を評価する人工知能が人間同士の関係に与える影響とその倫理的含意」、『三田評論』、1230巻、pp. 40-46、2019年2月。(依頼論文) (査読なし)

(3) 久木田水生「遠隔戦争の論理と倫理」、『 α シノドス』、Vol. 257+258、2018年12月20日。(依頼論文) (査読なし)

(4) 久木田水生「二一世紀の可愛い機械のトルコ人たち」、『ユリイカ』、第50巻、第9号、207-214、2018年6月27日。(依頼論文) (査読なし)

(5) 久木田水生「ICTがもたらすコミュニケーションの変容」、『中部哲学会年報』、第49号 27-40、2018年4月30日。(依頼論文) (査読なし)

(6) 久木田水生「麦とペッパー：テクノロジーと人間の相互作用」、『人工知能』、32巻5号 653-659、2017年。(依頼論文) (査読なし)

〔学会発表〕（計7件）

(1) 久木田水生「コミュニケーションの過去・現在・未来：ヒューマン・エージェント・インタラクションのもたらす新しいコミュニケーションとその倫理的課題」、電気学会倫理委員会全

国大会シンポジウム、2019年3月12日、北海道科学大学。（招待講演）

（2）久木田水生「人はなぜ道具を使うのか：遺伝子とミームの複雑なダンス」、進化経済学会2018年オータムカンファレンス、2018年9月29日、名古屋工業大学。（招待講演）

（3）久木田水生「テクノロジーがデザインする人と人の距離：その倫理的含意」、社会情報学会学会大会、シンポジウム「AIが“媒介”する社会」、2018年9月8日、島根大学。（招待講演）

（4）久木田水生「共生か寄生か？ ——人とロボットの哲学と倫理」、静岡哲学会第40回大会シンポジウム「情報技術の展開と哲学」、2017年10月3日、JR静岡駅ビル パルシェ7F会議室 第1会議室。（招待講演）

（5）Minao Kukita, ``Buridan's Asimo: Difficulties in mechanisation of moral competence,`` Society for Philosophy of Technology, June 14-17, 2017, Technische Universität Darmstadt. (査読あり)

（6）Minao Kukita, ``When HAL kills, stop asking who is to blame,`` CEPE/Ethicomp 2017, June 5-9, 2017, L'Università degli Studi di Torino. (査読あり)

（7）Minao Kukita, ``Morality, Emotions, and Ethical Symbol Grounding,`` 10th International Conference on Applied Ethics, Oct. 28-30, 2016, Hokkaido University. (査読あり)

〔図書〕（計2件）

（1）ウェンデル・ウォラック、コリン・アレン著、岡本慎平、久木田水生訳と解説『ロボットに倫理を教える』、名古屋大学出版会、2018年12月。

（2）久木田水生、神崎宣次、佐々木拓著『ロボットからの倫理学入門』、名古屋大学出版会、2017年2月。

〔産業財産権〕

○出願状況（計0件）

○取得状況（計0件）

〔その他〕

ホームページ等

<http://www.is.nagoya-u.ac.jp/dep-ss/phil/kukita/morality-mod-science.html>

6. 研究組織

(1) 研究分担者

研究分担者氏名：戸田山和久

ローマ字氏名：Kazuhisa Todayama

所属研究機関名：名古屋大学
部局名：情報学研究科
職名：教授
研究者番号（8桁）：90217513

研究分担者氏名：村上祐子
ローマ字氏名：Yuko Murakami
所属研究機関名：立教大学
部局名：理学部
職名：特任教授
研究者番号（8桁）：80435502

研究分担者氏名：神崎宣次
ローマ字氏名：Nobutsugu Kanzaki
所属研究機関名：南山大学
部局名：国際教養学部
職名：教授
研究者番号（8桁）：50422910

(2) 研究協力者

研究協力者氏名：中村美知夫
ローマ字氏名：Michio Nakamura

研究協力者氏名：三輪和久
ローマ字氏名：Kazuhisa Miwa

研究協力者氏名：服部宏充
ローマ字氏名：Hiromitsu Hattori

研究協力者氏名：大澤博隆
ローマ字氏名：Hiroataka Osawa

研究協力者氏名：川口潤
ローマ字氏名：Jun Kawaguchi

研究協力者氏名：平理一郎
ローマ字氏名：Riichiro Hira

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。