

令和元年5月21日現在

機関番号：13301

研究種目：基盤研究(B)（一般）

研究期間：2015～2018

課題番号：15H03151

研究課題名（和文）個性を持つロボットの制作による〈心と社会〉の哲学

研究課題名（英文）Philosophy of "mind and Society" through Making Robots with Personality

研究代表者

柴田 正良（Shibata, Masayoshi）

金沢大学・その他部局等・その他

研究者番号：20201543

交付決定額（研究期間全体）：（直接経費） 12,500,000円

研究成果の概要（和文）： ロボットが人間と真に共生するには、そのロボットは〈個性〉を持たねばならない、というのがわれわれの出発点である。しかし、その〈個性〉とは何であろうか？ われわれの哲学的分析では、〈個性〉をもつとは、それが〈道徳的な行為主体〉になるということである。

この哲学的テーゼを経験的なレベルで実証するために、人とロボットとの一連のインタラクション（相互交差）実験を開発し、実施したことが、本研究の全体像である。その実験の結果、われわれは、このテーゼに対してある程度の肯定的な支持を得ることができた。こうして、ロボットの〈個性〉と〈道徳性〉に関する哲学的仮説に、実験的検証の道が拓かれたのである。

研究成果の学術的意義や社会的意義

本研究の社会的意義は、ロボット時代の先を見据えた〈人間・ロボット共生社会〉の中心課題に正面から切り込み、その解決に一定の見通しを得た点にある。その中心課題とは、暗に人々が最も懸念している問題、つまり「自律した知性ロボットは人間とどう共生しうるのか？」という問いである。

この問題に対して、我々は、単なる道具ではないロボットは〈個性〉を持たねばならず、〈個性〉を持つとはそれが〈道徳的行為者〉となることなのだという哲学テーゼを以て答え、しかもそれに、ある種の実験的な確証を与えた。道徳問題というロボット時代の真の難問にこの方向性を示したことは、今後の社会に対して極めて大きな意義を持つと言える。

研究成果の概要（英文）： We started from the point that robots must have their "personality" in order for them to truly coexist with humans. But what is the "personality"? In our philosophical analysis, having a "personality" is nothing other than to become a "moral agent".

Our project consists in the process that we developed and conducted a series of interaction experiments between robots and humans so that we could ascertain our philosophical thesis on empirical level. We could attain a certain degree of affirmative answer to that thesis by those experimental results. In this manner we have opened a way to verify a philosophical hypothesis by empirical evidences.

研究分野：現代哲学

キーワード：人間とロボットの共生社会 ロボットの個性 道徳的行為者としてのロボット 人とロボットとのインタラクション実験 モラル・エージェント 身体同調動作課題 道徳的ジレンマ課題 トローリー問題

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

## 1. 研究開始当初の背景

(1) 本研究の背景は、現在、社会の各所で驚くほどの進化を遂げている AI 技術とその応用が予感させる「未来のロボット社会」の問題である。すでに常識と化しているように、このままこの技術革新が進めば、多くの人々の職業は AI とロボットに奪われ、一時的にもせよ社会には大量失業と貧富の格差が生じるであろう。しかし、われわれの大多数が本当に不安を感じているのは、この先、ロボットが知性を備え、自律したパートナーとなってわれらの生活の身近に入り込んで来た時、われわれはそのロボットとどのように共存していけるのか、ということである。この問題は、裏を返して言えば、すでに多くの SF 映画が描いたような、ロボットによる人類の支配もしくは殲滅といった極端な可能性である。

(2) このような漠然とではあるが不可避の懸念に対しては、たんに情緒的な反応をするのではなく、冷静な概念的分析和経験科学的な証拠によって対処することが肝心である。われわれが本研究を開始した背景は、まさにこのような社会的状況であった。もちろん、この研究が含む「心身問題」という存在論的テーマにはそれなりの長い哲学的背景があるが、それを述べるのはここでは割愛する。

## 2. 研究の目的

本研究「個性を持つロボットの制作による〈心と社会〉の哲学」の目的は、ロボットと人間をインタラクション(相互交流)させることによって、(1)「個性」概念の哲学的・心理学的分析の歴史を踏まえながら、来るべき「ロボットと人間の共生社会」において重要な要素となる「個性」の認知哲学的基盤を解明し、また、(2)現在、盛んに開発されつつある人型ロボットの性質と能力を「個性」の観点から評価することによって、人間のパートナーとしてのロボットの今後の開発方向に提言を行い、(3)将来、ロボットが社会に実装された段階での、人間とロボットの「個性をベースにした共生」のあり方を探求するものである。

## 3. 研究の方法

(1) われわれの研究方法は、大別して2つである。一つは、個性概念の哲学的な分析を進め、ロボットにおいて求められている「個性」の意味が何を概念的に含意しているのかを抽出し、それを基に、通俗概念とは異なる新たな「個性」概念とそれに関する哲学的テーゼを構築することである。もう一つは、この哲学的テーゼをある程度ではあっても経験的に検証できるような実験の仕組みを開発し、その実験を実施することである。

(2) この2つ目の実験的部分は、概念分析の不徹底さとも相まって、当初はなかなかその方向性と内容が定まらなかった。われわれが購入したソフトバンク社製のロボット(「かいと」と名づけた)にいきなり個性を実現できるような認知行動プログラムを開発しようと試みたが、その結果は満足のいくものではなかった。その認知行動プログラムの根幹部分は、平成27年度に作成した2つの会話プロット、「頑固・強引型」と「迎合・優柔不断型」に対応する性格に加えて、「懐疑・不安型」の会話プロットであったが、いずれも、人間とのインタラクション(相互交流)実験において、「個性」相互の重要な差異をもたらすようなものではなかったのである。

(3) 実験部分における研究方法上の手詰まりを解消できたのは、実は、「個性」概念についての哲学的分析が進んだ結果である。つまり、実験結果と往還させながら行うはずの「概念の改良 プログラムの改善」が、認知行動プログラムの開発とは別の方向を可能としてくれたのである。それは、ロボットと人間とに身体動作同調課題を課すことによって、その課題に参加した特定のロボットに対する被験者の態度の変化をデータとして計測するというものである。ここで肝心なのは、その態度変化を、ロボットの道徳性に対する評価の変化として計測する点である。

(4) なぜそれが本研究の方法上の突破口となったのかと言えば、それは、この時点までにわれわれが、新たな個性概念テーゼ、つまり「ロボットが〈個性〉をもつとは、それが〈道徳的な行為主体 moral agent〉であるということであり、道徳的行為主体であることは、他の何者も代替できない責任を引き受けるということであり、そのためにロボットは他者が経験しえない内面世界をもたねばならない」というテーゼに到達していたからである。つまりこのテーゼから、人がロボットにどの程度「個性」を感じるかは、「ロボットと行なう身体的同調動作によって、ロボットに対する道徳的判断の帰属がどのように変化するか」によって定量的に測られうる、ということが帰結する。その際、ロボットへの道徳性の帰属の測定に関しては、トローリー問題などの道徳的ジレンマ課題についての、これまでの心理学的研究成果を利用することができた。

## 4. 研究成果

(1) われわれが到達した、新たな個性概念テーゼとは、「ロボットが〈個性〉をもつとは、それが〈道徳的な行為主体 moral agent〉であるということであり、道徳的行為主体であることは、他の何者も代替できない責任を引き受けるということであり、そのためにロボットは他

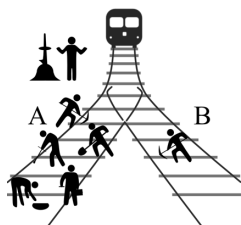
者が経験しえない内面世界をもたねばならない」というものである。個性を持つに至ったロボットは、もはや道具として扱うことはできず、互いの尊重と理解に基づく互恵的なルールの取り決めを要する存在となる。これはすなわち、ロボットを道徳的な行為主体として人間社会(道徳共同体)の一員に加えることである。以上が、ロボットの個性を検討することによって得られた哲学的仮説である。

(2) 上記仮説の妥当性を検証するためには、条件が統制できる状態で人間とロボットの身体的な相互交流(インタラクション)を可能とし、かつその相互交流の有無に応じて、ロボットに対する人間の捉え方の変化を定量的に測ることのできる枠組みが必要となる。さらに、ロボットとの身体的な相互交流においては、可能な限りシンプルなやり取りを通じて、ロボットの内的状態(個性)を押し量ることのできる方法が必要である。また、ロボットに対する捉え方においては、ロボットの行為に対する道徳的判断を取り出すことのできる方法が必要である。

(3) 前者の課題に対してわれわれは、対に配置された独立した2つの回転ハンドルを対面、もしくは背面で回すことのできる実験装置を開発した。また、後者の課題に対しては、行為の道徳性を評価するトローリー課題と、二人に与えられたお金の分配額を交渉する独裁者ゲーム課題および最後通牒ゲーム課題を組み合わせた実験を設計した。トローリー課題は、暴走したトロッキが進む先に5人の作業員がいるが、ポイントを切り替えれば1人が轢き殺されるだけで済むという状況で、一緒にハンドルを回した相手(人もしくはロボット)がポイントを切り替えたことが説明され、その行為の道徳的な問題の「あり/なし」や責任の程度が参加者に質問される。独裁者ゲーム課題では、自身の判断で独裁的に分配額を決定できる。このため、ロボットを社会の一員と捉えるならば、平等な分配が行なわれることが予想される。最後通牒ゲーム課題は、金額を提示された側が拒否権を持ち、これが発動されると、二人が受け取る金額はゼロになる。ロボットを社会の一員とみなすならば、不公平な分配が行なわれたときには、自身が一円ももらえなくとも、相手からの提案を拒否することが予想される。



(4) われわれは、人どうしの対面/背面条件、および人とロボットの対面条件での実験を実施した。人とロボットの条件では、ハンドルを回す人の動きに追従し続ける従順なロボットを作成した。その結果、人どうしの実験では、背面条件よりも対面条件の方が、回すハンドルの同調度が高くなることが確認された。そして、その対面条件の方が、トローリー課題で道徳的に問題視される行動を「より問題がある」と判断する傾向があった。また、お金の分配額に大きな違いはなかったが、不公平な提案に対しては対面条件の方が拒否する傾向が高かった。この結果は、われわれの設計した実験課題が身体的相互交流のパターンと道徳的な判断をつなぎ、その間にあるメカニズムを調べることのできる統制実験環境となっていることを示している。



(5) 人とロボットとの実験では、回すハンドルの同調度は他の条件に比べて最も高くなった。トローリー課題で道徳的に問題視される行動に対しては、人どうしでの背面条件と同程度に「問題があり、かつ責任がある」と回答する傾向が見られた。ロボットにおいて人間の背面条件と同程度の傾向となったことは、ロボットが単に合理性のみを持って判断する存在として評価されたのではないことを示唆している。他方、お金の分配においては、分配額を独裁的に決めることのできる状況においても、ほとんどの参加者が半額をロボットに与えた。また、ロボットからの不公平な分配額の提案に対しては、自身の受け取れる金額がゼロになるにも拘わらず、提案を拒否する傾向が、人どうしの対面条件より強く現れた。これらの結果は、人間がロボットを道徳的な主体として捉える萌芽と解釈できる。

(6) われわれが設計した実験枠組みによって、ロボットに実装する身体動作のメカニズムがどのようなものであれば、人どうしでの対面条件と同等の「責任に関する問いと評価」が可能となった。このように、哲学的考察を通じてわれわれが導いた仮説を具体的な実験結果によってある程度ではあるが検証できるようになったこと、これが、本プロジェクトの最も大きな成果である。今後、この方向の研究を継続してさらに加速進展させるため、われわれは、本科研費プロジェクトの分担者である中京大学の長滝祥司教授を代表者として、ほぼ同一のメンバーによる新たな科学研究費「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(A)2019-23年度)を申請し、幸いにも採択の通知を頂いた。この次期研究は、「ロボット工学や心理学などの経験的手法を取り入れつつ、ロボットのような新たな存在を道徳的行為者として受容できる社会を構想するための新たな道徳理論の主要テーゼを提言する」ことが目的である。

〔雑誌論文〕(計 63 件)

大平 英樹、社会規範の遵守と逸脱-認知神経科学的アプローチ、*生物科学*、70(3)、査読有 2019、171-177、[http://www.ruralnet.or.jp/seibutsu/070\\_03.htm](http://www.ruralnet.or.jp/seibutsu/070_03.htm)

Takeshi Konno、Mechanisms of intentional joint visual attention、*Philosophical Inquiries*、7(1)、査読有、2019、63-76、doi:10.4454/philing.v7i1.239

Shoji Nagataki and Liberati Nicola、Vulnerability under the gaze of robots: relations among humans and robots、*AI & SOCIETY*、査読有、2018、1-10、DOI:10.1007/s00146-018-0849-1

橋本 敬、金野 武司、長滝 祥司、大平 英樹、入江 諒、河上 章太郎、佐藤 拓磨、加藤 樹里、柏端 達也、三浦 俊彦、久保田 進一、柴田 正良、ロボットは道徳的な行為主体になり得るか、<個性>を持ち得るか、*日本認知科学会第 35 回大会発表論文集*、査読有、2018、958-960、[http://siva.w3.kanazawa-u.ac.jp/image/JCSS2018\\_sP2-15.pdf](http://siva.w3.kanazawa-u.ac.jp/image/JCSS2018_sP2-15.pdf)

Shoji Nagataki、Masayoshi Shibata、Takashi Hashimoto、Tatsuya Kashiwabata、Takeshi Konno、Hideki Ohira、Toshihiko Miura、and Shinichi Kubota、On the robot as a moral agent、*Interacción 2018 Proceedings of the XIX International Conference on Human Computer Interaction*、Article No.24、査読有、2018、5 pages、DOI: 10.1145/3233824.3233832

〔学会発表〕(計 79 件)

金野 武司、ロボットの個性を考えることの意味とそれを研究するための試行錯誤について、第 10 回感情と思考の科学研究会、*金沢工業大学*、2018

Takashi Hashimoto、Recognition of Agency and Morality in Robots: Effect of Reasoning and Bodily Coordination、*International Symposium for "Potentials and Perspectives of Communication among Humans and Agents Including Robots and Animals"*、2018

Shoji Nagataki、Similarity and Alterity: How Machines Become Moral Agents、*4S/EASST CONFERENCE Boston-2017*、2017

Shoji Nagataki、Masayoshi Shibata、Tatsuya Kashiwabata、Takashi Hashimoto、Takeshi Konno、Hideki Ohira、Toshihiko Miura、Shinichi Kubota、Robot As Moral Agent: A Philosophical and Empirical Approach、*CogSci 2017*、London

柴田 正良、人間がロボットと共生する日…ロボットの心から人類の道徳まで…、第 61 回 香料・テルペンおよび精油化学に関する討論会 (The 61<sup>st</sup> TEAC in Kanazawa)、*金沢工業大学*、2017

〔図書〕(計 14 件)

三浦 俊彦、改訂版：可能世界の哲学「存在」と「自己」を考える、*二見文庫*、2017、302

柏端 達也、*勁草書房*、現代形而上学入門、2017、287

柏端 達也、*コミュニケーションの哲学入門*、慶應義塾大学出版会、2016、108

森下 直貴、久保田 進一、他 9 名、*生命と科学技術の倫理学*、丸善出版、2016、262(112-131)

Nagataki, S.、Friis, et al.、*Technoscience and Postphenomenology*、*Lexington Books*、2015、259(225-237)

〔その他〕

ホームページ等

<http://siva.w3.kanazawa-u.ac.jp/>

6 . 研究組織

(1)研究分担者

研究分担者氏名：長滝 祥司  
ローマ字氏名：(NAGATAKI shoji)  
所属研究機関名：中京大学  
部局名：国際教養学部  
職名：教授  
研究者番号(8桁)：10219587

研究分担者氏名：橋本 敬  
ローマ字氏名：(HASHIMOTO takashi)  
所属研究機関名：北陸先端科学技術大学院大学  
部局名：先端科学技術研究科  
職名：教授  
研究者番号(8桁)：90221837

研究分担者氏名：金野 武司  
ローマ字氏名：(KONNO takeshi)  
所属研究機関名：金沢工業大学  
部局名：工学部  
職名：講師  
研究者番号(8桁)：40288436

研究分担者氏名：大平 英樹  
ローマ字氏名：(OHIRA hideki)  
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)：80263193

研究分担者氏名：柏端 達也  
ローマ字氏名：(KASHIWABATA tatsuya)  
所属研究機関名：慶應義塾大学  
部局名：文学部(三田)  
職名：教授  
研究者番号(8桁)：50537058

研究分担者氏名：三浦 俊彦  
ローマ字氏名：(MIURA toshihiko)  
所属研究機関名：東京大学  
部局名：大学院人文社会系研究科  
職名：教授  
研究者番号(8桁)：20201543

研究分担者氏名：久保田 進一  
ローマ字氏名：(KUBOTA shinichi)  
所属研究機関名：金沢大学  
部局名：国際基幹教育院

職名：特任助教

研究者番号（8桁）：60710267

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。