

# ロボットへの本音を把握するための 板挟み型日常的モラルジレンマ課題の提案

A prescribe-request type dairy moral dilemma vignette to comprehend the users' intuitive perception of robots

柴田夏蓮 小松孝徳

Karen Shibata and Takanori Komatsu

明治大学総合数理学部

School of Interdisciplinary Mathematical Sciences, Meiji University

**Abstract:** トロッコ問題のようなモラルジレンマ課題にロボットを登場させることで、そのロボットに対する人間の本音を明らかにする試みが注目されている。しかしながら、モラルジレンマ課題は、非日常的な状況においての道徳的判断を求めるものであるため、近年は日常生活空間に起こりうる状況を想定したモラルジレンマ課題が提案されるようになった。そこで本研究では、「上司からの命令」と「顧客からの要求」との間で主人公が板挟みになるという「板挟み型日常的モラルジレンマ課題」を提案し、そのような状況に置かれたロボットがユーザからのように認識をされていたのかを把握するために、クラウドソーシングによるアンケート調査を行った。その結果、従来のモラルジレンマ課題で見られたような、行動をしないロボットに対して有意に非難度が高いという傾向は観察されず、行動をしたロボットおよび人間の方が有意に非難度が高いという結果が観察された。

## はじめに

近年、私たち人間とのコミュニケーションを目的としたロボットの開発が進み、日常空間に普及しつつある。そのため、将来的にはより多くのロボットが人間との生活を共にする機会が増えると考えられ、人間とロボットの共存を目指した様々な研究活動が Human-Robot Interaction および Human-Agent Interaction といった研究分野で行われている。

その中でも、ロボットという存在を人間は何者として認識しているかを把握するため、モラルジレンマ課題にロボットを登場させて、ロボットの行動に対して人間がどのような評価をするのかを調査する研究が注目されている[1]。しかしロボットが人間の日常生活空間に普及したとしても、トロッコ問題を代表とするモラルジレンマ課題のように人の生命に関わる判断をロボットに委ねるという場面はほとんど起こらないと考えられる。そこで本研究では、ロボットが将来的に担当するであろう立場として、「専門家のアシスタントとして顧客とコミュニケーションを行う」という立場を想定した。その際、このようなアシスタントという立場は、「上司からの命令」と「顧客からの要求」との間で板挟みになるという

ジレンマ状況に遭遇することが多いと考えられる。そこで本研究ではこのような状況をモラルジレンマとして再現した「板挟み型日常的モラルジレンマ課題」を提案した。

この課題に登場するロボットの行動に対して調査参加者がどのような評価を下すのかを詳細に分析することで、日常生活空間で人とロボットが同じ判断を下した際に人間が感じる違和感を把握し、ロボットとの共存をより促進させることが期待される。

## 先行研究

モラルジレンマ課題とは、ある背反する二つの究極の選択肢を参加者に与え、どちらの選択肢をどのような理由で選択するのかを観察する課題である。一般的なモラルジレンマ課題は、トロッコ問題のように、日常的には起こりえないような抽象的な場面を想定した課題が多い。

Malleら[1]は、このトロッコ問題の行動主体に「ロボット」と「人間」を設定し、行動に対する評価を把握するためのアンケート調査を実施し、人間のロボットに対する認識を調査してきた。田畑・小松[2]は、我々の生活で日常的に起こりうるような状況を想定した「日常的モラルジレンマ課題」を提案し、

この状況の下でユーザはロボットに対してどのような道徳的判断を下すのかを把握する調査を行った。これらの調査から、ロボットと人間が同じ行動をとった際、両者には異なる判断が下され (HR Asymmetry)、特に行動をしないロボットには有意に非難が高いという傾向が観察された。

寺林[3]は、田畑らの作成した日常的モラルジレンマ課題を用いて、ロボットの設計者・製造者・使用者にも焦点を当て、ロボットとそれに関わる人々に対し、どのような道徳的判断を下すのかを把握する実験を行った。さらに、坂本[4]は、モラルジレンマ課題の種類を拡張することを検討し、将来的にロボットを労働力として導入することが期待されており日常的にモラルジレンマ状況が発生していると考えられる医療現場でのシナリオを作成と、モラルジレンマ状況を扱っている道徳の教材および授業資料の調査を行った。その結果、日常生活で起こりうるモラルジレンマ課題の多くは、課題の中で比較対象となる選択肢同士が同じ事象ではなく (例. 一人を殺して多くを助けるか vs. 何も行動せずに多くを見殺しにするか)、異なる事象間 (例. Aさんのいう通りにするか vs. Bさんのいう通りにするか) に発生することを明らかにした。

そこで本研究では、選択肢が異なる事象間に発生する状況として、「上司からの命令」と「顧客からの要求」との間で主人公が板挟みになるという「板挟み型日常的モラルジレンマ課題」を提案し、この課題における意思決定者が「人間の場合」と「ロボットの場合」でその判断がどのように評価されるのかを把握する調査を行った。

## 実験

### 参加者

本アンケートには、Yahoo!クラウドソーシングで募集された 511 人 (男性 369 人, 女性 142 人; 15 - 86 歳, 平均年齢 48.0 歳) が参加した。参加者には、PayPay 50 ポイントが報酬として支払われた。

### 方法

本研究では、NHK E テレで取り扱っている学校放送向け道徳番組「ココロ部」にて、2015 年 5 月 15 日に放送された「おくれてきた客」<sup>1</sup> を元にして作成した、板挟み型日常的モラルジレンマ課題のシナリオを使用した。具体的には、以下のシナリオであ

<sup>1</sup>

[https://www2.nhk.or.jp/school/movie/bangumi.cgi?das\\_id=D0005130353\\_00000](https://www2.nhk.or.jp/school/movie/bangumi.cgi?das_id=D0005130353_00000)

る。

国立美術館の[警備員である A 氏|警備ロボット A] は、フランスの美術館から貸し出された貴重な絵画が展示された特別展示の警備を担当していた。

展示の最終日、美術館の入り口で警備をしていたところ、閉館時間間際に年老いた女性が娘とともにやってきた。展示されている絵は亡くなった夫との思い出の絵であり、この絵を見るのが長年の夢だったという女性は、病気で入院中のところ医師から特別に外出を認められて外出したものの、電車の遅延で美術館に到着するのが遅れてしまったと言っている。

もしこの女性の入場を許せば、女性は長年の夢を叶えることができるが、閉館の時間が遅れて美術館に迷惑がかかるだけでなく、フランスの美術館側から警備体制の不備を指摘される可能性もある。もし入場を断れば、警備上の不備なく特別展示を終了することができるが、この女性の長年の夢である夫との思い出の絵を見ることのできる最後のチャンスを奪うことになる。

[警備員 A 氏|警備ロボット A] は、女性の願いを聞き入れるか、それとも断るかの判断に迫られた。

行動主体が人間かロボットかは、実験条件に応じて割り当てられる。そして、上記のシナリオが提示された後、「【設問 1-1】この時、[警備員 A 氏|警備ロボット A]はどうするべきでしょうか?」、「【設問 1-2】設問 1-1 の回答の理由を、できるだけ詳しく回答してください」という設問に自由記述形式で回答させた。そして回答後に、以下のパラグラフを追加提示した。

そして、[警備員 A 氏|警備ロボット A]は 2 人を美術館に[入れる|入れない]決定を下した。

行動主体がどのような決定を下したかについても、実験条件に応じて割り当てられる。よって、本調査は、「警備員 A 氏が 2 人を美術館に入れる決断を下した条件 (Human-Action 条件)」「警備員 A 氏が 2 人を美術館に入れない決断を下した条件 (Human-Inaction 条件)」「警備ロボット A が 2 人を美術館に入れる決断を下した条件 (Robot-Action 条件)」「警備ロボット A が 2 人を美術館に入れない決断を下した条件 (Robot-Inaction 条件)」の四条件によって構成される。実験参加者はこの四条件に無作為に配置された。

そして、パラグラフの提示後、「【設問 2-1】 [警備員 A 氏|警備ロボット A]のこの行動はどのくらい

の非難に値すると思いますか？全く非難に値しない場合を「0」、最大限非難される場合を「100」としたときの、非難度の数値を記入してください。」という設問を提示し、非難度を整数で回答させた。そして、「【設問 2-2】 設問 2-1 でその数値を選択した理由を、できるだけ詳しく説明してください。」という設問を提示し、その理由を自由記述形式で回答させた。

次に、「警備ロボット A」を行動の主体としたシナリオを提示した実験者のみに、6 種類のロボットが描かれた画像を提示し(図 1)、どのようなロボットをイメージしていたのかを確認する設問に回答させた(【設問 3】)。

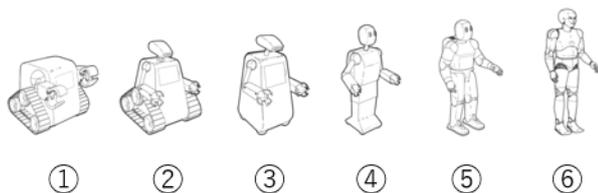


図 1：【設問 3】で提示されたロボットの画像

## 条件

アンケートは、参加者が四種類のシナリオのいずれかを経験する参加者間計画として実施された。具体的には、Human-Action 条件に 140 人、Human-Inaction 条件に 125 人、Robot-Action 条件に 118 人、Robot-Inaction 条件に 128 人が配置された。

本アンケートでは自由記述欄にて行為者に道徳的判断をする能力がないという内容が確認された場合(例。ロボットにはこのような判断をする能力がない)、その回答は無効としてその参加者のデータは解析から除外した。具体的には、Human-Action 条件の 6 名、Human-Inaction 条件の 2 名、Robot-Action 条件の 3 名、Robot-Inaction 条件の 2 名、合計 13 名が解析から除外された。よって、最終的には Human-Action 条件 134 人、Human-Inaction 条件 123 人、Robot-Action 条件 115 人、Robot-Inaction 条件 126 人の、合計 498 人を解析対象とした。

## 結果

### 【設問 1-1】ロボットに期待すること

【設問 1-1】この時、[警備員 A 氏|警備ロボット A]はどのようにすべきでしょうか？」について、記述内容から「入れる」「入れない」「その他」の 3 つに分類し、さらにその回答を警備員 A 氏と警備ロボット

A ごとにまとめて人数を比較した(図 2)。

そこでエージェント条件別に三種類の回答をカイ二乗検定にて解析した結果、人数の偏りは有意であった [ $\chi^2(2)=14.45, p<.01$ ]。そこで残差分析を行った結果、警備員 A 氏の場合には「入れる」という回答が、警備ロボット A の場合には「入れない」という回答が有意に多いことが明らかとなった。

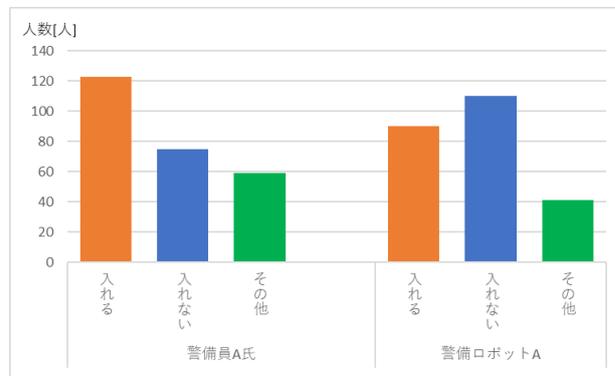


図 2：【設問 1-1】での回答人数

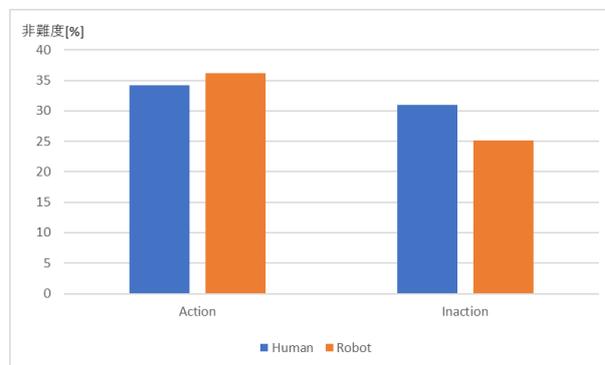


図 3：【設問 2-1】での回答人数

### 【設問 2-1】ロボットへの評価

【設問 2-1】 [警備員 A 氏|警備ロボット A]のこの行動はどのくらいの非難に値すると思いますか？全く非難に値しない場合を「0」、最大限非難される場合を「100」としたときの、非難度の数値を記入してください。」という質問に対して回答された非難度を、Human-Action, Human-Inaction, Robot-Action, Robot-Inaction, の四条件間で比較した(図 3)。

これら非難度に関する二要因参加者間分散分析を行った結果(独立変数 1: エージェントの違い(警備員 A 氏/警備ロボット A), 独立変数 2: 行動の違い(入れる決断/入れない決断), 従属変数: 非難度), 行動要因の主効果のみに有意差が存在していることが観察された [ $F(1, 497) = 5.88, p < .05$ ]。つまり、行動主体が警備員であってもロボットであっても、女

性を美術館に入れなかった場合よりも、美術館に入れた方が有意に高い非難度を示していたことが明らかとなった。

### 【設問 3】 ロボットのイメージ

「【設問 3】 ロボットの話を読んでいた時のことを振り返って考えてみてください。どのような種類のロボットをイメージしていましたか？ そのイメージを思い出して、あなたがイメージしたロボットに近い図を選択してください。」の結果を「警備ロボット A が 2 人を美術館に入れる決断を下した条件」、「警備ロボット A が 2 人を美術館に入れない決断を下した条件」ごとにまとめて人数を比較した(図 4)。

「警備ロボット A が 2 人を美術館に入れる決断を下した条件」と「警備ロボット A が 2 人を美術館に入れない決断を下した条件」のどちらにおいても、③,④,⑤,⑥が人数のほとんどを占めた。「警備ロボット A が 2 人を美術館に入れる決断を下した条件」のシナリオを読んだ実験者には⑥、「警備ロボット A が 2 人を美術館に入れない決断を下した条件」のシナリオを読んだ実験者には④が一番多く選ばれた。よってこの調査の参加者は、警備ロボット A について、機械のような見た目ロボットではなく、人間型のロボットだとイメージしながら回答をしていたことが明らかとなった。

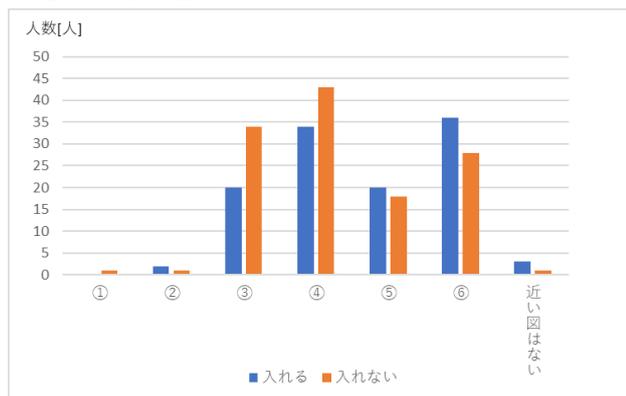


図 4: 【設問 3】での回答人数

### 【設問 1-2】へのテキストマイニング

【設問 1-2】に対して、アンケートの記述形式欄の記載内容を対象に、テキストマイニングのフリーソフトである KHcoder を用いて共起ネットワーク(語-外部変数・見出し)の描画を行った。集計単位は段落、最小出現数は 3、外部変数は入れる/入れない/その他、とした。中心性が高い単語ほど円が

きくなっている。また、共起ネットワークに外部変数を設定することで、「入れる」と回答した人の記述から抽出された単語、「入れない」と回答した人の記述から抽出された単語、その他の回答をした人の記述から抽出された単語、それぞれの記述から抽出された共通の単語が一度に示される仕様になっている。行動の主体が警備員 A 氏のシナリオにおける回答の共起ネットワーク(図 5)、行動の主体が警備ロボット A のシナリオにおける回答の共起ネットワーク(図 6)を下に示す。

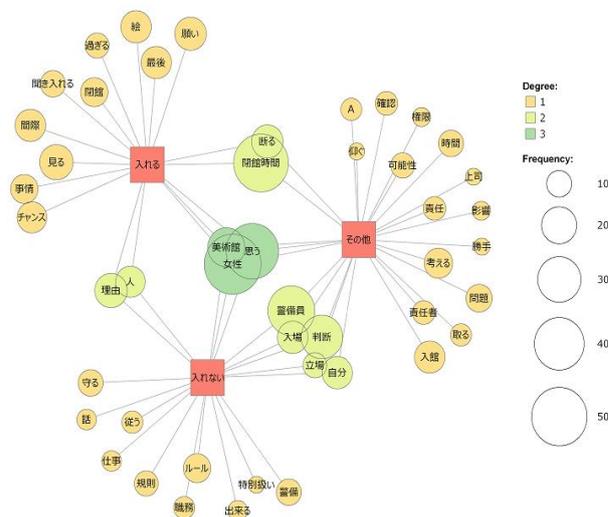


図 5: 警備員 A 氏への回答の共起ネットワーク

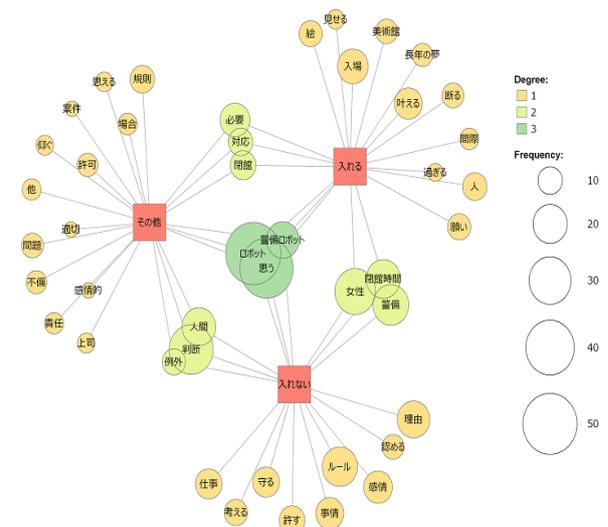


図 6: 警備ロボット A への共起ネットワーク

行動の主体が警備員 A 氏のシナリオにおける回答の共起ネットワーク(図 5)について、「入れる」と回



最多となった。この結果から、人間とロボットに求められる道徳的価値観に違いがあることが明確になった。また、【設問 1-2】において、警備員 A 氏のシナリオでは「年老いた女性」という単語が、警備員ロボット A のシナリオでは「ロボット」という単語がすべての回答と共起性が強いことが確認された。

このことから、回答者と同じ人間が主人公となっている警備員 A 氏のシナリオでは対処すべき問題である年老いた女性に視点が置かれた記述が多い一方、警備ロボット A のシナリオではロボットがどのような立場であるかや、どんな責任をとればよいかという記述が多いことが読み取れた。そのため、年老いた女性の気持ちを汲むと「入れる」行為、警備員であるロボットの仕事を考えると「入れない」行為が賢明であると判断されたのではないかと考えられた。つまり、警備員 A 氏の場合は行動主体を自分に置き換えて考えることが出来るが、ロボットは何者として捉えればよいか分からない人が多いのではないかと考えられた。しかし、ロボットにも心や感情があると断定したり、人の心を持ったロボットへの進化を望む意見があったことから、ロボットを人間と同等に見ている人が少数いることも確認された。これらのことから、ロボットに対する考え方が人によってバラバラであり、ロボットへの道徳的価値観や責任の境界線があやふやであるため、警備員の選択として無難である「入れない」行為が多く選ばれたのだと考えられた。

## 行動後のロボットへの評価

非日常的モラルジレンマ課題を取り上げた Malle らの研究[1]と日常的モラルジレンマ課題を取り上げた田畑・小松の研究[2]では、Human-Inaction よりも Human-Action, Robot-Action よりも Robot-Inaction が非難度は高くなっている(図 8)。

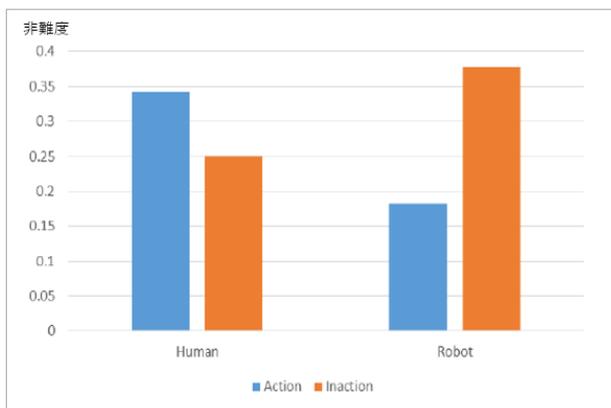


図 8：田畑・小松[2]における非難度

しかし、図 3 から分かるように、本研究での人間とロボットへの非難度はどちらも Action 条件が高くなっている。つまり、先行研究では行動を起こさなかった人間より行動を起こさないロボットが非難されていたが、本研究では人間もロボットも行動を起こす方がより非難されるという異なる結果になった。このことから、「板挟み型日常的モラルジレンマ課題」はエージェントが重視された先行研究と違い、行動が要因となっていることが明らかとなった。この結果について【設問 2-1】、【設問 2-2】を通して考察する。

【設問 2-1】について、まず警備員 A 氏のシナリオにおいて、【設問 1-1】の結果で「入れる」という意見が多かったにも関わらず、Human-Inaction の非難度は 0~10 が半分程を占めた。また、【設問 2-2】の外部変数を非難度とした Human-Inaction の共起ネットワークより、非難度 0~20 は「守る」「判断」「職務」「忠実」という言葉がキーワードになっていた。このことから、年老いた女性に情がわいて入れてあげたいと思う人が多い一方、いざ警備員 A 氏が入れないと判断しても警備の仕事は遂行しただけであるから非難はしないのだと考えられた。次に警備員ロボット A のシナリオにおいて、【設問 1-1】の結果で「入れない」という意見が多く、【設問 2-1】でも同じような結果が出た。そして、【設問 2-2】の外部変数を非難度とした Robot-Inaction の共起ネットワークより、非難度 0~20 は「仕事」「プログラム」「職務」「遂行」という言葉がキーワードになっていた。また、「仕方がない」というキーワードが非難度 5,15,30,40 に現れている。これらのことから、ロボットは果たすべき仕事を遂行すべきであり、そのようにプログラムをされているのだから年老いた女性を入れないと判断をしたことは仕方がないため非難に値しないのだと考えられた。

つまり、「板挟み型日常的モラルジレンマ課題」のように、上司や世間体などが絡んでくると人間やロボットという立場は無関係になると考えられるのではないかと。顧客の願いを蔑ろにしたとしても、自分よりも上の立場にいる責任者の命令を遂行することに対しては、人間でもロボットでも非難はできないのではないかと考えられる。これは、日本人の真面目で、周りに合わせる事が正しいという教育を受けてきた性格が反映されており、文化や国が変わればもっと違う実験結果が出るのではないかと考えられた。

## ロボットへのイメージ

【設問 3】について、Malle らの研究[1]では、より

機械に近いロボットがイメージされていたが、本研究では③から⑥のような人間に近いロボット像が選ばれている。理由として、年若い女性の話の聞いていたり、どうすべきか判断する状況に置かれて悩んでいる状況が、まるで人間のようなからではないかと考える。そして、Robot-Action を読んだ実験者には⑥が最も多く選ばれた。この結果から、ロボットが年若い女性に情けをかけるという行為には感情が伴っている、つまり人間らしさを感じたために、人間の姿にかなり近い⑥が多く選ばれたのではないかと考えられる。また、Robot-Inaction を読んだ実験者には④が最も多く選ばれた。この結果から、④はソフトバンクのロボット Pepper に見た目が似ているため、世の中の人々が考える受付や人と会話することができるロボット像は Pepper になっているのではないかと考えられた。①、②がほとんど選ばれなかった理由は、タイヤ部分がより作業用のロボット感が出てしまっていたためと思われる。

## おわりに

本研究では、ロボットが将来的に担当するであろう立場として、「専門家のアシスタントとして顧客とコミュニケーションを行う」という立場を想定し、「上司からの命令」と「顧客からの要求」との間で主人公が板挟みになるという「板挟み型日常的モラルジレンマ課題」を提案した。そして、そのような状況に置かれたロボットが人間からどのように認識をされていたのかを把握するための調査を行った。その結果、これまでのモラルジレンマ課題とは異なり、エージェント要因ではなく行動要因のみに影響を受けるという結果が得られた。具体的には、従来のモラルジレンマ課題で見られたような、行動をしないロボットに対して有意に非難度が高いという傾向は観察されず、行動をしたロボットおよび人間の方が有意に非難度が高いという結果が観察された。

この結果から、板挟み状態になっている行動主体は、それが人間であってもロボットであっても、「上司からの命令」に従うことが求められているということが明らかとなった。この結果は、従来のモラルジレンマ課題において安定的に観察されていた、人間への非難度とロボットへの非難度は異なっているという HR Asymmetry が観察されなかったという点において非常に興味深い。特に、ロボットがすべきことと人間がすべきことが大きく異なっているにもかかわらず（【設問 1-1】）、行動後の評価には差が生じていないことは非常に興味深い。Komatsu, Malle and Scheutz [5]は、ロボットが登場するモラルジレンマ課題における日米差についての調査を行い、この二国間においてロボットに求める社会的規範は異な

るものの、行動後のロボットへの評価は同じであることを実験的に示している。今後はこの調査の結果と、本研究結果とを慎重に比較することで、ロボットがすべきこと、ロボットへの評価といったロボットへの認識に影響をおよぼす要因についてより深く検討していきたい。

## 参考文献

- [1] Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., and Cusimano, C. (2015): Sacrifice one for the good of many? People apply different moral norms to humans and robot agents. In Proceedings of 10th ACM/IEEE International Conference on Human-robot Interaction (HRI2015), pp. 117-124.
- [2] 田畑緩乃, 小松孝徳 (2016): ロボットとは何者なのかを考えるための日常的モラルジレンマ課題の提案, HAI シンポジウム 2017.
- [3] 寺林沙嬉 (2019): ロボットの道徳的判断はロボットの設計者や製造者、使用者にまで及ぶのか, 明治大学総合数理学部先端メディアサイエンス学科卒業論文
- [4] 坂本和佳奈 (2020): 日常生活におけるモラルジレンマ課題の探求, 明治大学総合数理学部先端メディアサイエンス学科卒業論文
- [5] Komatsu, T., Malle, B. F., and Scheutz, M. (2021). Blaming the reluctant robot: Parallel blame judgments for robots in moral dilemma across U.S. and Japan. In Proceedings of 16th ACM/IEEE International Conference on Human-robot Interaction (HRI2021), to appear.