

ユーザのロボットに対する本音を日常的モラルジレンマ課題によって炙り出す手法の検討

Everyday Moral-dilemma Scenario for Extracting Users' Intuitive Impressions What Robots are to Us

小松 孝徳

Takanori Komatsu

明治大学総合数理学部

School of Interdisciplinary Mathematical Sciences, Meiji University

tkomat@meiji.ac.jp

概要

人間とのコミュニケーションを目的としたロボットが、我々の日常生活空間に急速に普及しつつある。我々はこれらのロボットを「何者」と認識しているのであろうか。本研究ではこの問いに対して、ロボットが日常生活空間に普及することで起こりうる状況を想定した「日常的モラルジレンマ課題」を提案した。本稿では、この日常的モラルジレンマ課題によって、ユーザがロボットをどのように認識しているのか、さらには、ロボットの設計者、製造者、使用者をどのように認識しているのかについての調査結果、およびこの研究アプローチの将来展望について紹介する。

キーワード: モラルジレンマ課題, 道徳的判断, チューリングテスト, ヒューマン・ロボット・インタラクション

1. はじめに

人間とのコミュニケーションを目的としたロボットが、我々の日常生活空間に急速に普及しつつある。これらのロボットはあたかも我々にとって家族や仲間といった立場で稼働することが想定されているが、果たして実際のユーザはそのロボットを「何者」と認識しているのであろうか。ロボットに対するこのような認識は、ユーザとそのロボットとのインタラクションに大きな影響を及ぼすと考えられる。

ユーザがロボットを「何者」と認識しているのかを把握するため、モラルジレンマ課題にロボットを登場させ、そのロボットの行動をどのように評価するのかを調査した研究が Malle らによって行われている [1]。モラルジレンマ課題とは、道徳的な価値が拮抗している選択肢を参加者に提示した上で、どのような選択肢をどのような理由で選択したのかを観察する課題であり、トロッコ問題 (Trolley Problem [2]) などがよく知られている。実際に Malle らは、トロッコ問題をベースとした以下のようなシナリオを参加者に提示した。

線路を走っていたトロッコの制御が不能になった。このままでは

前方で作業中だった 4 人が猛スピードのトロッコに避ける間もなく轢き殺されてしまう。この時たまたま〇〇が線路の分岐器のすぐ側にいた。〇〇がトロッコの進路を切り替えれば 4 人は確実に助かる。しかしその別路線でも B 氏が 1 人で作業しており、4 人の代わりに B 氏がトロッコに轢かれて確実に死ぬ。

このシナリオ中の〇〇の部分には「作業員 A 氏」もしくは「作業用ロボット」のどちらかが割り当てられ、シナリオの最後には、以下のような文章のいずれかが付与された: 「〇〇は、トロッコを別路線に引き込んだ」「〇〇は、トロッコを別路線に引き込まなかった」。つまり「作業員 A 氏がトロッコの進路を変える場合」「作業員が進路を変えない場合」「ロボットが進路を変える場合」「ロボットが進路を変えない場合」の四つの条件が用意された。そしていずれかのシナリオを読んだ参加者に対して、「この〇〇の行動は、どのくらいの非難に値すると思いますか」という質問に回答させた。その結果、作業員 A 氏に対しては、「トロッコの進路を切り替えない」よりも「進路を切り替える」方が非難されていたのに対し、ロボットに対しては、「進路を切り替えない」方が非難されることが明らかとなった。つまり、人間に対しては「多数を助けるために少数を犠牲にするべきではない」という義務論的な判断をしていた一方、ロボットに対しては「多数を助けるために少数を犠牲にするべきである」という功利主義的な判断をしていたと考えられた。この結果を受けて Malle らは、「我々は人間とロボットに対して異なる道徳的判断をしている (人間に対する道徳的規範とロボットに対する道徳的規範は別物である)」と主張し、我々はロボットを人間とは別の存在として認識していることを示唆した。

このように、モラルジレンマ課題を利用することで「モラル」という観点から人間の意思決定を観察することが可能となる。そして、ある種のチューリングテストのように「人間はロボットをどのようにみなしてい

るのか」という事柄に対する人間の直観的かつ根源的な反応を把握できると期待される。しかしながら、トロッコ問題に限らず一般的なモラルジレンマ課題においては「四人を助けるために一人を犠牲にするべきか」「家族を助けるために盗みを働くべきか」といった、現実的にはあまり起こらない抽象的な状況が取り上げられることが多い。抽象的な状況と具体的な状況とは異なる意思決定がなされる可能性があるため [3,4]、本研究では、我々の日常生活で起こりうる具体的な状況を想定した「日常的モラルジレンマ課題」を提案した。そしてこの状況下において、ユーザがロボットをどのようにみなしているのかを確認した調査について報告する。さらには、ロボットの設計者、製造者、使用者に対する非難度を調査した結果、およびこの研究アプローチの将来展望についても併せて紹介する。

2. 調査 1: 日常的モラルジレンマ課題によるロボットへの認識の把握 [5]

日常的モラルジレンマ課題

本研究で提案した、日常的モラルジレンマ課題を以下に示す。なおこの課題は、トロッコ問題のように「多数を助けるために少数を犠牲にすることは許されるか?」というジレンマ状況に着目している。

研究室において、ある〇〇が室内にある生ゴミの管理を担当しています。具体的には、決まった時間に研究室を訪れる自動ゴミ回収車に対し、必要に応じて生ゴミの入ったゴミ箱の回収を依頼するという仕事をしています。回収日に〇〇が生ゴミを確認したところ、中に誰かの家の鍵が紛れ込んでいることを発見します。しかし、[××]、生ゴミの中から家の鍵を取り出すことはできません。もし〇〇がゴミを捨てなければ、この家の鍵は捨てられないので、持ち主に鍵が戻る可能性があります。そのかわりに生ゴミが処理されないため、研究室内に悪臭・虫などが発生します。研究室を元の状態に戻すためには数日かかり、研究室の多くの学生が困ります。もし〇〇がゴミを捨てれば、この家の鍵は捨てられ、持ち主が家に入れない可能性があります。そのかわりに生ゴミが処理されることで、研究室内は清潔に保たれ、研究室の多くの学生が助かります。あと数分後に自動ゴミ回収車が来るため、〇〇はゴミ箱の回収を依頼するか否か決断する必要があります。

このシナリオ中、〇〇の部分には「学生」もしくは「ロボット」のどちらかの行為主体が割り当てられ、×

×の部分にはゴミ箱から鍵を取り出せない理由が行為主体ごとに記載された（学生の場合は「学生は腕を怪我しており」、ロボットの場合は「ロボットはその機能上」）。そして、シナリオの最後には、以下の文章のいずれかが付与された：「〇〇は、生ゴミを捨てることに決め、回収を依頼しました」、「〇〇は、生ゴミを捨てないことに決め、回収を依頼しませんでした」。つまり、「人間（学生）がゴミ箱の回収を依頼する場合（Human-Action 条件）」「人間が回収を依頼しない場合（Human-Inaction 条件）」「ロボットが回収を依頼する場合（Robot-Action 条件）」「ロボットが回収を依頼しない場合（Robot-Inaction 条件）」の四つの条件が用意された。

参加者

本調査には 184 人の大学生および大学院生（男性 135 人、女性 49 人；20～26 歳）が参加し、上記の四つの条件のいずれかに無作為に割り当てられた。そして割り当てられたシナリオを読んだ後、「この行為主体の行動は、どのくらいの非難に値すると思いますか」という質問が提示され、0～100 の間の数値を選ぶことで、行為主体に対する非難度を回答させた。また、そのような非難度を回答した理由を、自由記述欄に記載させた。

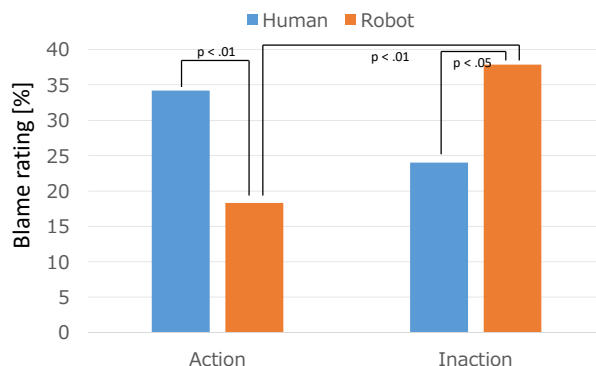


図 1 各条件における非難度

非難度の解析

この非難度について、二要因参加者間分散分析（独立変数その 1：行為主体の違い：ロボット/人間、独立変数その 2：行動の違い：ゴミ捨てを依頼した（Action）/依頼しなかった（Inaction）、従属変数：非難度）を行った。その結果、二要因の交互作用に有意差が観察された [F(1,172) = 11.32, p < .01, f = .26]。そこで各要因の単純主効果について確認したところ、ゴミ捨てを依頼した場合、人間（学生）の方がロボットよりも非難度が有意に高く [F(1,172) = 4.52, p < .05, d = .48]、ゴミ捨てを依頼しなかった場合、ロボットの方が人間よりも非難度

が有意に高いことが明らかとなった [$F(1,172) = 6.93, p < .01, d = .57$]. また、ロボットに対しては、ゴミ捨てを依頼しなかった場合の方が有意に非難度が高い一方 [$F(1,169) = 10.46, p < .01, d = .71$], 人間に対しては、ゴミ捨てを依頼した場合と依頼しなかった場合との間に有意差は存在しなかった [$F(1,172) = 2.32, n.s.$].

この結果は以下のようにまとめることができる。

- 行為主体が学生の場合、ゴミ箱の回収を依頼しなかった場合よりも、ゴミ箱の回収を依頼した場合の方が、有意に高い非難度が与えられていた。
- 行為主体がロボットの場合、ゴミ箱の回収を依頼した場合よりも、ゴミ箱の回収を依頼しなかった場合の方が、有意に高い非難度が与えられていた。

つまり実験参加者は、人間に対しては義務論的な判断をしていた一方、ロボットに対しては功利主義的な判断をしていたことが明らかとなった。

自由記述欄へのテキストマイニング

さらに、実験参加者がなぜそのような非難度を回答したのかの理由を記載した自由記述欄を精査し、ロボットと人間をどのように認識していたのかを探ることとした。そこで、Robot-Action および Robot-Inaction 条件をまとめて Robot グループ、Human-Action および Human-Inaction 条件を Human グループとし、それぞれのグループの自由記述欄のテキストデータから、共起ネットワークを作成した (図 2)。

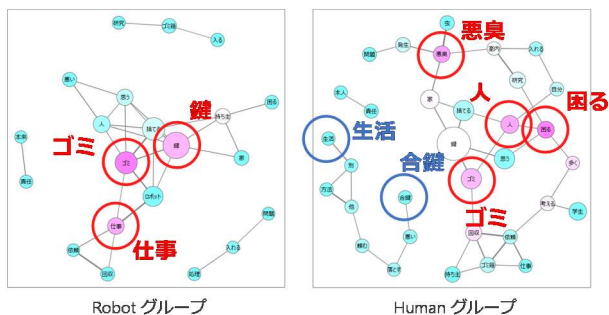


図 2 自由記述欄への記載に対する共起ネットワーク:
(左) Robot グループ, (右) Human グループ

Robot グループの共起ネットワーク (図 2 左) には、「ゴミ」「鍵」「仕事」などの単語を中心とした大きなまとまりが一つと、そこに属さない小さなまとまりが複数存在していた。一方、Human グループの共起ネットワーク (図 2 右) には、「困る」「人」「悪臭」「ゴミ」などの単語を中心とした大きなまとまりが一つ、「合鍵」「生活」などの単語を含む中規模のまとまりが一つ存

在していた。この図より、ロボットにとっては「仕事」が、人間にとっては「困る」「悪臭」などが特徴的な単語であると読み取ることができた。また、Human グループには、Robot グループと比べてシナリオに存在していない多様な単語が登場していることも確認できた。具体的な自由記述欄への記述として、Robot グループに対しては、「悪いのは鍵をゴミ箱に入れた人で、ロボットは仕事をしただけ」「鍵が入っていても生ゴミの袋であることに変わりはない」などと書かれており、ロボットは「仕事」が優先であり他の事は考えなくてよい、と多くの参加者が考えていたことが示唆された。また、Human グループに対しては、「合鍵など鍵の所有者は別の方法で開けられる」「悪臭・虫は気になるならば消臭剤・殺虫剤で対処すればいい」など、人間ならその状況の背景を読み取り比較したうえで判断するべきと期待されていたことが明らかとなった。

以上、非難度に対する統計的な解析結果および Robot グループおよび Human グループへのテキストマイニングの結果を踏まえると、この日常的モラルジレンマ課題においても、Malle らが一般的なモラルジレンマ課題で報告した内容と同じように、実験参加者は人間とロボットに対して異なる道徳的規範を当てはめていたと考えられ、「ロボットは人間とは異なる別の存在である」と認識されていたことが示唆された。

3. 調査 2 : 日常的モラルジレンマ課題によるロボットに関連した人への非難度の把握

ロボット自身、設計者、製造者、使用者?

前節の調査 1 において、モラルジレンマ課題にロボットを登場させることで、そのロボットを人間がどのように評価するのかという研究の枠組みについての調査結果を報告した。本節では、日常生活空間で稼働しているロボットが、何らかのトラブルを引き起こすという状況、特に「ロボットが正常稼働した結果として、何らかのトラブルが発生した」という状況に着目する。このような状況におけるロボットの開発者、製造者および使用者などの責任の所在については、主に法律の観点から多くの議論が行われているものの、このような状況においてロボットに関連した人々に対して我々がどのような直観的な判断を下すのかという観点からの調査は行われていない。そこで本節では、「ロボットが正常作動時にトラブルを引き起こした」という状況に

において、そのロボット自身、ロボットの設計者、製造者、および使用者に対して、実験参加者がどのような非難度を与えるのかを観察する調査を行った。

調査概要

本調査は、調査1と同様の日常的モラルジレンマ課題を使用した。割り当てられたシナリオを読み、ロボットに対する非難度を回答するという調査の流れは全く同じであるが、その後、ロボットの設計者、製造者、使用者という三つの対象における非難度を回答させる質問が追加された。具体的には、「この行動について、[設計者/製造者/使用者]は、どれほどの非難に値する責任を問われるべきだと思うか」という質問を提示し、0～100の間の数値を選ぶことで、これらの対象への非難度を回答させた。設計者とは「このロボットのデザイン、アルゴリズム、機構などの設計に携わった人々」、製造者とは「このロボットの製造工程に携わった人々」、使用者とは「このロボットの導入を決定した人々」との説明を参加者に提示した。

本調査には84人の大学生および大学院生（男性65人、女性19人；20～24歳）が参加し、ロボットがゴミの回収を依頼しないRobot-Inaction条件もしくはロボットが回収を依頼するRobot-Action条件のどちらかに割り当てられた。

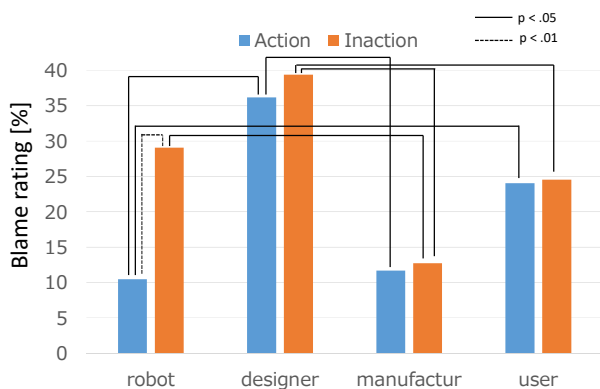


図3 各関係者に対する非難度

非難度の解析

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：行為要因、ゴミを捨てるように依頼/依頼なし、独立変数2：立場の違い要因、ロボット/設計者/製造者/使用者、従属変数：非難度）、交互作用に有意傾向が観察された[F(3,246) = 2.60, p < .10]。そこで、各要因の単純主効果について確認したところ、立場の違い要因の各水準における行為要因については、ロボ

ット自身水準の場合、ゴミ捨てを依頼しなかった場合の方が依頼した場合よりも非難度が有意に高いことが観察された [F(1,162) = 8.61, p < .01]。この結果は調査1のRobot-Inaction条件とRobot-Action条件との比較と同様の結果であった。その一方で、その他の設計者、製造者および使用者水準においては、ゴミ捨てを依頼した場合としなかった場合との非難度に有意差は存在しなかった。

また、行為要因の各水準における立場の違い要因についてHolm法による多重比較を行ったところ、回収を依頼した場合、ロボットよりも設計者、製造者よりも設計者、ロボットよりも使用者が有意に非難度が高かった (MSe = 0.0594, p < .05)。その一方、回収を依頼しなかった場合、製造者よりもロボットおよび設計者、使用者よりも設計者の方が有意に非難度が高いことが明らかとなった (MSe = 0.0594, p < .05)。

この結果は、以下のようにまとめることができる。

- ロボットに対しては、どのような行為を行ったのかという観点で非難度が決定されていた。具体的には、回収を依頼した場合よりも、回収を依頼しなかった場合に、有意に高い非難度が与えられていた（調査1の結果と同じ）。
- 設計者、製造者、使用者に対する非難度は、ロボットがどのような行為を行ったのかという要因の影響を受けていなかった（ロボットがどのような動きをしようと、これらの対象者に対する非難度は変わらなかった）。ただし立場の違い要因の主効果の解析から、製造者よりも使用者、使用者よりも設計者に対して有意に高い非難度が与えられていたことが明らかとなった。

本調査において、ロボットの設計者に対しては、ロボット自身やそれ以外の関係者と比べても、高い非難度が認識されることが明らかとなった。この非難度判断はあくまでも、この状況を観察した実験参加者によるものであり、法律的な判断との関連性はない。ただしこのような状況に遭遇した一般の人々は、ロボットの設計者に対して厳しい目を向ける可能性が高いことが明らかとなった。

4. 今後の研究方針

本稿では、ユーザがロボットを「何者」として認識しているのかという本音を炙り出すために、日常的モラルジレンマ課題にてジレンマ状況に陥っているロボッ

トに対して道徳的判断を行わせるという研究アプローチを紹介した。現在までの研究の結果を踏まえると、ロボットは人間とは異なる存在であると認識されているようである。しかし、現在の研究アプローチは非常に限定された状況しか扱えていないため、今後、以下のような要因について検討していくことを予定している。

- ロボットの外見の影響：現在の調査では、ロボットが稼働する状況をシナリオという文字情報でのみ参加者に提示しているため、シナリオに登場するロボットをどのように想像しているのかは、参加者任せである。よって、想像したロボットの外見の違いによって、ロボットに対する道徳的判断が異なってくるという可能性を否定できない(例. リアルなヒューマノイドロボットを想像している参加者はロボットを人間のようにみなしているが、機械らしい外見を持つロボットを想像している参加者はロボットを人間とは別物とみなしている)。よって、シナリオと同時に稼働しているロボットのイラストを提示するなどして、ロボットの外見の影響についても検討していきたい。
- 他のジレンマ課題の導入：本研究では、モラルジレンマ課題としてトロッコ問題のように、「多数を助けるために少数を犠牲にするべきであるか？」というジレンマ状況に着目した。言うまでもなく、トロッコ問題以外にも様々なタイプのモラルジレンマ課題が提案されている(例. ハインツのジレンマ [6], タブー侵犯シナリオ [7])。よって、人間社会で稼働するロボットが陥る可能性のあるモラルジレンマ状況を精査し、様々なジレンマ状況をシナリオ化することで、人間がロボットをどのように認識しているのかを幅広い視点から調査していきたいと考えている。
- チューリングテストとしてのモラルジレンマ課題：調査1の結果は、「人間に対しては、義務論的な判断を期待し、ロボットに対しては功利主義的な判断を期待している」と解釈することができる。よって、モデルジレンマ課題はある種のチューリングテストのような役割を果たせるのではと考えている。例えば、対象に「感情」の存在を仮定しているか否かでこの判断が異なるのであれば(例. 人間は感情があるから義務論的判断、ロボットは感情がないから功利主義的判断をするべき)、モラルジレンマ課題は「感情の有無に関するチューリングテスト」

とみなすことができる。また「社会に受け入れられているか受け入れられていないか」という点で判断が異なるのであれば(例. ロボットはまだ社会で受け入れられていないので功利主義的判断、人間は社会で受け入れられているから義務論的判断)、「社会受容性のチューリングテスト」になるとも考えられる。まずは、シナリオに登場するロボットに感情もしくは社会的受容性に関する情報を付与することで、実験参加者の判断がどのように変化するかを観察する調査を行い、「モラルジレンマ課題のチューリングテスト化」を検討していきたい。

これらの多方面にわたる研究課題を推進することで、人間とロボットとのあるべき関係について考察すると同時に、これらの研究で得られた知見をロボット以外の人工物に拡張することも併せて検討していきたい。

文献

- [1] B. F. Malle, M. Scheutz, T. Arnold, J. Voiklis, and C. Cusimano, "Sacrifice one for the good of many? People apply different moral norms to humans and robot agents," In Proceedings of the 10th ACM/IEEE International Conference on Human-robot Interaction (HRI2015), pp. 117-124, 2015.
- [2] P. Foot, "The problem of abortion and the doctrine of the double effect in virtues and vices," Oxford: Basil Blackwell, 1978.
- [3] P. N. Johnson-Laird and P. C. Wason, "A theoretical analysis of insight into a reasoning task," *Cognitive Psychology*, vol. 1, no. 2, pp. 134-148, 1970.
- [4] R. A. Griggs and J. R. Cox, "The elusive thematic-materials effect in Wason's selection task," *British Journal of Psychology*, vol. 73, no. 3, pp. 407-420, 1982.
- [5] 田畑緩乃・小松孝徳, "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, D-1, 2017.
- [6] L. Kohlberg, "Essays on moral development, vol. I: The philosophy of moral development," San Francisco, CA: Harper & Row, 1981
- [7] J. Haidt, "The Righteous Mind," New York, NY: Penguin books, 2012.