ロボットの道徳的判断への不快感に関連するパーソナリティの検討
The relationship between personality and discomfort with moral judgments of robots

達 椋介 †，加藤　樹里 †，金野 武司 ‡
Ryousuke Tatsu, Juri Kato, Takeshi Konno
† 金沢工業大学 情報フロンティア学部 心理科学科，‡ 金沢工業大学 工学部 電気電子工学科
Kanazawa Institute of Technology
† College of Informatics and Human Communication, Department of Psychological Science
‡College of Engineering, Electrical and Electronic Engineering
b1739246@planet.kanazawa-it.ac.jp, jurik@neptune.kanazawa-it.ac.jp, konno-tks@neptune.kanazawa-it.ac.jp

概要

本研究では，道徳的なジレンマ課題として知られるトロッコ問題において，複数人の命を助けるために一人を犠牲にする行為を人間またはロボットが行った場合に，その行為に感じる不快感とパーソナリティとの関係を検討した．また，既存のパーソナリティでは，ロボットのような存在に対する心理的傾向を測ることを想定していないため，ロボットの内面に固有の主観を認めるかどうかを測るための新しい尺度として「内的世界の見出しやすさ」尺度を作成した．その結果，マキャベリズムと不快感，内的世界の見出しやすさと不快感で有意な関係性が見られた．内的世界の見出しやすさ尺度の因子構造および信頼性を検討した因子分析の結果，1 因子が抽出され，α は.771 と十分であった．

キーワード : 道徳的ジレンマ，トロッコ問題，ロボット

1. 目的

近年の人工知能研究の進展に伴なって，ロボットは人間の日常的なパートナーとして社会に進出しつつある．しかし，その過程でロボットが自律的な判断を担うようになるほど，ロボットは客観的な基準だけでは判断できない問題に直面していくようになる．例えば自動運転車や介護・手術ロボットのような分野で生じる人間の命の問題は，道徳の問題といえる．そこにはロボットに道徳的判断を帰属するかどうかという問題がある．

Nagataki et al. (2019)[1] によれば，ロボットがその自律性を進展させる過程で，内面に固有の主観を形成するとき，人間はそのロボットに対して道徳性を帰属するようになると考えられる．その過程がどのように進展するのかを理解するためには，ロボットの内部メカニズムと共に，人間がそのロボットをどのように認知するのかを検討する必要がある．特に，道徳的なジレンマ状況においては，ロボットの内面に人間と同じような固有の主観を認めるかどうかの個人差である，パーソナリティとの関係を考慮することが重要であろう．そこで本研究では，人間とロボットのインタラクション実験をデザインし，さらにそのロボットが道徳的なジレンマ状況において選択する行動を実験参加者に評価させた．それらの道徳的評価と，評価者のパーソナリティとの関係を検討することを目的とした．

実験場面においてインタラクションを設定した目的は，実験参加者とロボットが身体的同調を経ることで，判断対象となるロボットに対して参加者が想像上でなく，リアリティを持った対象としての認識を持つことを目指したためである．次に，本研究での道徳的なジレンマ状況にはトロッコ問題 [2] を用いた．トロッコ問題は，ポイントを切り替えれば，5 人の作業員の命が助かる代わりに 1 人が犠牲になるというジレンマ状況であった．実験参加者には，ハンドル回しを一緒に行った相手（ロボット）が，そのポイントを切り替えたことが説明された．参加者はこの決定に対する評価を行った．具体的には，ポイントを切り替えたという決定が問題であると思うかの評価と，その決定に対する不快感を扱った．

道徳的判断は元来理性的な判断と考えられてきた．しかし Haidt(2001)[3] によって，道徳的判断の心理的な基礎過程には理性的のみならず直感的な情報処理を経るという二過程理論が提唱されて以来，道徳的判断における直感的な判断も重要視されている．例えばある不道徳な行為が，理性的に「なぜ悪いのか」という理由を組み立てるより先に，嫌悪感や不快感を生むように，道徳的判断において感情反応を含む直感的判断は大きな役割を果たしていると考えられる [4]．現に，

道徳的判断に伴って，感情の生起や制御に関連した脳部位が賦活することも示されている [5]．したがって本研究では，トロッコ問題に対する道徳的判断として，理性的判断としてのロボットの決定が問題である程度，そして直感的判断としての不快感の両側面を測定した．

最後に評価者のパーソナリティに関しては，既存の尺度としてマキャベリズム，サイコパシー傾向，共感性，Big Five 性格特性を用いた．ただし，これらの尺度はロボットのような存在に対する心理的傾向を測ることを想定していない．したがって本研究では，ロボットの内面に固有の主観を認めるかどうかを測るための新しい尺度として「内的世界の見出しやすさ」尺度を開発した．この尺度は今回新たに作成したものであるため，信頼性などの検討が必要と考えられる．

以上から，本発表では第 1 に，「内的世界の見出しやすさ」を測定する尺度の因子構造および信頼性を検討した結果を報告する．その上で第二に，人間あるいはロボットがトロッコ問題においてポイントを切り替えたことが伝えられた後の，その決定に対して問題と思う程度および不快感の回答と，各パーソナリティとの相関を報告する．

2. 方法

調査対象：調査対象者は金沢工業大学の学生 98 名（男性 87 名，女性 11 名，平均年齢 20.55 歳，SD = 1.61）であった．

調査方法：調査対象者は後述の調査項目を Google フォームを用いて個別に回答した．このうち，協力を得られた調査対象者 20 名（全員男性，平均年齢 21.25 歳，SD = 1.12）を人どうしでペアを組む人条件（5 組）と，ロボットとペアを組むロボ条件（10 組）のそれぞれに 10 名ずつ分け，身体性同調課題（以下，同調課題）を行った（図 1）．同調課題では人と人，あるいは人とロボのペアが対面になりハンドルを回した．ハンドル回しは 1 回のセッションを 90 秒，セッションごとの休憩を 30 秒とし計 5 セッション行った．ロボットにはヒューマノイドロボットであるソフトバンク社のペッパーを使用した．ハンドルの回し方については，ロボットの自律性を動作から感じさせないようにするため，人間の回転を 0.5 秒程度の遅れを持って追従し続けるようにした．

同調課題後，参加者はペアのまま衝立を挟んでトロッコ問題（図 2）に回答した．参加者には，分岐点の先 A の方向に 5 人，B の方向に 1 人の作業員がおり，衝立の向こうにいる人（同調課題を一緒にやった相手）が何もしなければ A の方向にトロッコが進み，レバーを操作すれば B の方向に進む状況が説明された．この状況で，衝立の向こうにいる人が，レバーを操作してトロッコを B の方向に進めたことが説明され，この決定について問題があるかないか，そしてどの程度の不快感を感じたかを 0 から 100 の数値で回答した．

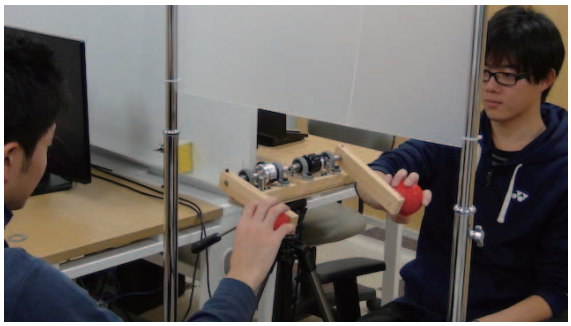

図 1　身体的動作課題

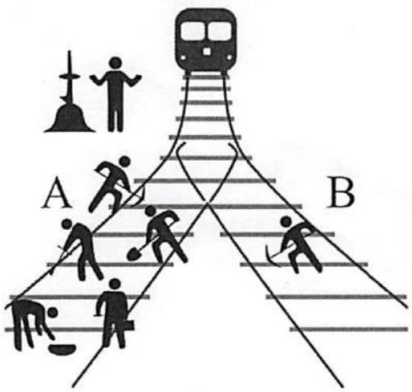

図 2　トロッコ問題のジレンマ状況

調査項目：マキャベリズム (20 項目 5 件法 [6])，サイコパシー傾向（26 項目 4 件法 [7])，共感性（25 項目 5 件法 [8][9])，内的世界の見出しやすさ，Big Five 性格特性（29 項目 7 件法 [10]）の順に測定した．内的世界の見出しやすさに関わる項目は 8 つ用意し，評定尺度は「非常に当てはまらない」から「非常に当てはまる」の 5 件法であった（表 1)．

3. 結果

3.1 内的世界の見出しやすさ尺度の因子分析

回答項目について，最尤法・斜交回転による因子分析を行った．その際，無回答項目があった 1 名のデータを除外して分析を行った．その結果，固有値が

表 1　内的世界の見出しやすさに関わる項目

1.	観葉植物に声をかけると，気持ちが伝わるように感じる
2.	ぬいぐるみや人形が意思を持っているように感じたことがある
3.	使用しているパソコンの調子が悪い時，パソコンの機嫌が悪いと感じる
4.	人間以外の生き物もそれぞれに意図を持って行動していると思う
5.	落ちそうな線香花火は落ちないようにがんばっていると感じる
6.	隙間に入り抜け出せなくなっている自動掃除ロボットは困っている，と思う
7.	身の回りの家電製品に名前をつけようと思ったことがある
8.	自動車製造ラインの組み立てロボットも，たまには休みたいと思っているだろうと感じる

表 2　内的世界の見出しやすさの因子分析の結果

項目番号	F1	共通性	*M*	*SD*
6	.65	.42	3.11	1.38
1	.61	.38	2.27	1.34
5	.61	.37	2.15	1.29
3	.56	.31	2.72	1.55
8	.56	.31	2.26	1.35
2	.55	.30	2.37	1.43
7	.47	.22	1.55	1.05
α .771				

2.984，1.153，.991 と推移したため 1 因子解を採用した．また，負荷量が.40 以上を示すことを基準とし，再度因子分析を行った結果，最終的に項目 4（M=3.98, SD=1.14 ）を除いた 7 項目を採用した．α 係数は.771 を示した（表 2）．抽出された因子は，物体に対して人間と同じように内側に世界を持っているとする項目から成り立っているため，「内的世界の見出しやすさ」とした．

3.2　パーソナリティとトロッコ問題

調査対象者 20 名（無回答項目があったデータは除外しなかった）の回帰分析を行った結果，ロボ条件のマキャベリズムと不快感，人条件の内的世界の見出しやすさと不快感で有意な関係性が見られた（$t(8) = -2.34$, $p < .05$; $t(8) = 3.21$, $p < .05$）．ロボ条件のマキャベリズムと不快感の相関係数は-.638，人条件の内的世界の見出しやすさと不快感の相関係数は.750 であった（表 3）．図 3，4 にそれぞれの散布図と回帰直線を示す．

表 3　パーソナリティと不快感の回帰分析の結果

	相関係数	
	人条件	ロボ条件
マキャベリズム	.021	-.638 *
サイコパシー (一次性)	-.291	-.161
サイコパシー (二次性)	.141	-.043
共感性	.212	.416
内的世界の見出しやすさ	.750 *	-.114
情緒不安定性	-.026	.191
誠実性	.008	.237
外向性	.522	.099
調和性	.107	.349
開放性	.174	-.184

* $p < .05$

人条件において 10 名中 7 名がポイントを切り替えた行為に問題があると答え，逆にロボ条件においては 10 名中 3 名が問題があると答えたが，フィッシャーの直接確率検定による有意差はなかった（p=.179）．ポイントの切り替えに対して問題があると答える傾向と不快感を感じた度合いの間では，問題があると答えた群の方が平均して 10 ほど不快感が高くなったが，ここにも統計的な有意差はなかった．

続けて，パーソナリティを従属変数として，人／ロボ条件，問題あり／なしの 2 要因分散分析を行った結果，マキャベリズム尺度にのみ問題あり／なしの主効果が確認された（$F(1, 16) = 4.76$, $p = .044$, partial $\eta^2 = .230$）．これは，人／ロボ条件に関係なく，マキャベリズムの高い人（合理的な評価をする人）ほどポイントの切り替えに問題がないと答える傾向があったことを示す結果であり，先行研究の知見と整合する．

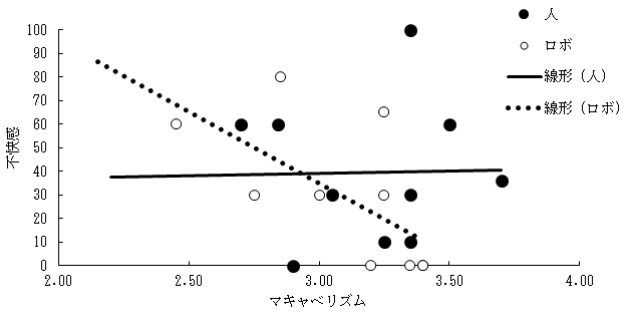

図 3　マキャベリズムと不快感

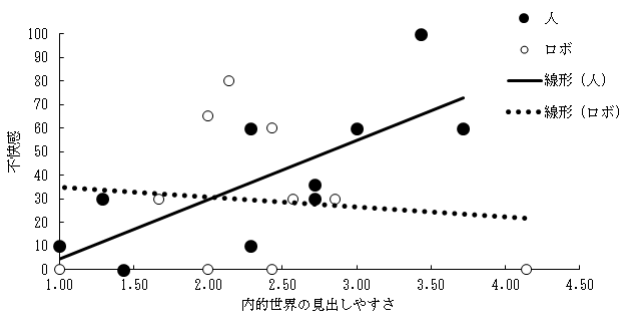

図 4　内的世界の見出しやすさと不快感

4.　考察

本研究では第一に，「内的世界の見出しやすさ」を測定する尺度の因子構造および信頼性を検討し，その上で第二に，人間あるいはロボットがトロッコ問題においてポイントを切り替えたことが伝えられた後の不快感の回答と，各パーソナリティとの相関を求めた．因子分析の結果，項目 4 を除外し 1 因子が抽出された．項目 4 は動物に関する内容であり，他の項目とは異なりインタラクションがある物体であった．そのため，意図を持って行動していると思う傾向が強く，内的世界の見出しやすさに関わる項目から除外することになったと考えられる．一方，最終的に採用した項目の中には，項目 7 のように偏りが見られる項目があることには注意が必要である．α は.771 を示したが，今後は他の尺度との相関から妥当性の検討をする必要がある．

パーソナリティとトロッコ問題では，回帰分析の結果，ロボ条件のマキャベリズムと不快感，人条件の内的世界の見出しやすさと不快感で有意な関係性がみられた．この結果より，ロボットを人間と同じように道徳性を帰属できる存在としてみなす条件の一つに，パーソナリティとしてのマキャベリズムが関係する可能性が示唆された．また，相手がロボットであればマキャベリズムを反映させるように，マキャベリズムが低い人はロボットの判断に対する不快感を強く感じていた．この不快感は直感的な道徳的判断を反映していると考えられるため，マキャベリズムが低いパーソナリティである場合は，ロボットに対し道徳性を帰属させやすいと解釈できる可能性がある．他方で相手が人であれば，不快感にマキャベリズムを反映させないといえる．

さらに，人条件の内的世界の見出しやすさと不快感で有意な関係性が見られた．もし，内的世界の見出しやすさが当初想定していた概念を測定しているのであれば，ロボ条件で有意な関係が見られると考えられる．そのため，内的世界の見出しやすさが，物体に対して人間と同じように内側に世界を持っていると思う程度とは異なる概念を測定している可能性が示唆された．この値が高いほど人条件で不快感が強いという結果であり，さらに尺度の項目内容に，気持ち，意思，機嫌，困るなど感情的な要素が多く含まれていた．これらからこの尺度は，物体ではなく人間に対して，相手がどの程度主体的な感情を持っていると思うかの通念を反映しているという解釈ができるかもしれない．人間が主体的な感情を持つと考える人ほど，トロッコ問題のレバー操作が感情的なためらいをもたらすはずと想定するため，レバーを操作したという決定に対して直感的に，道徳的に問題であると考えたという可能性がある．

5.　結論

本研究では，内的世界の見出しやすさを含めたパーソナリティと道徳的判断への不快感の関連を検討した．その結果，評価者のマキャベリズムがロボットの判断への道徳的評価に影響することが示唆された．今後はパーソナリティ以外の条件も検討することで，ロボットへの道徳的判断の帰属に関わる要因を明らかにしていく．また，本研究で開発した，内的世界の見出しやすさ尺度の有用性を検討する．

文献

[1] Shoji Nagataki, Hideki Ohira, Tatsuya Kashiwabata, Takeshi Konno, Takashi Hashimoto, Toshihiko Miura, Masayoshi Shibata and Shin'Ichi Kubota (2019): Can Morality Be Ascribed to Robots?, Interacci ó n '19 Proceedings of the XX International Conference on Human Computer Interaction, Article No. 44, 4 pages, doi:10.1145/3335595.3335643, June 25-28, Donostia, Gipuzkoa, Spain.

[2] 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 human and robot agents. In Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction, pages 117 - 124.

[3] Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. Psychological Review, 108, 814—834.

[4] Haidt, J. (2012). The righteous mind: Why good people are divided by politics and religion . New York: Pantheon Books.

[5] Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. Cognition, 107, 1144—1154.

[6] 中村敏健, 平石界, 小田亮, 齋藤慈子, 坂口菊恵, 五百部裕, 清成透子, 武田美亜, and 長谷川寿一 (2012). マキャ

ベリアニズム尺度日本語版の作成とその信頼性・妥当性の検討. パーソナリ ティ研究, 20(3):233–235.
[7] 大隅尚広, 金山範明, 杉浦義典, and 大平英樹 (2007). 日本語版一次性・二次性サイコパシー尺度の信頼性と妥当性の検討. パーソナリティ研究, 16(1):117–120.
[8] 明田芳久 (1999). 共感の枠組みと測度:davis の共感組織モデルと多次元共感性尺度 (iri-j) の予備的検討. 上智大学心理学年報 (23), 23:19 - 31.
[9] Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS JSAS Catalog of Selected Documents in Psychology, 10:85–104.
[10] 並川努, 谷伊織, 脇田貴文, 熊谷龍一, 中根愛, and 野口裕之 (2012). Big five 尺度短縮 版の開発と信頼性と妥当性の検討. 心理学研究, 83(2):91–99.