

AIに 社会の声 を届ける

犠牲の選択に関する社会調査の報告

目次

問題の所在	4
Column 01. AIによる意志決定(支援)に関する主な問題	5
Column 02. AIに倫理を教える	6
本調査の概要	9
01. 単純集計	10
02. 道徳基盤理論(Moral foundations theory)に基づく質問紙調査(MFQ2)	35
03. 相関分析	39
本調査の総括	47
Column 03. 値値観と判断の結びつきについて	48
Appendix.	49

本報告書で使用したイメージ画像は、OpenAIの画像生成モデル「DALL-E」で作成したものである。

AIに
犠牲の選択を
委ねますか？



問題の所在

AIが判断する社会の到来？

意志決定の支援にAIが広く用いられるようになりつつある。

「AIの判断を参考にする」という前提のもと、実際にはその判断を追認するだけになってしまう懸念がある。

AIの判断と人間の判断の乖離

AIによる判断は、私たち自身の判断とは異なるものである。

AIが用いる学習データも、必ずしも私たち自身の経験や価値観に基づくものではない。

AI開発者たちが属する文化圏の価値観が強く反映されている恐れがある。

将来的にAIに判断を委ねるとしても、その妥当性や正当性をどのように検証するかという仕組みは、いまだ確立されていない。

《究極の選択》：社会的意志決定における犠牲の問題

特に重要なのは、社会全体で対応を迫られる危機において、誰がどのような犠牲を受け入れるかという判断である。

多くの場合、私たちはそのような《究極の選択》に直面することなく、悩む機会も与えられず、関連するデータも存在しない。

例えば、パンデミック時におけるワクチン接種の優先順位などがその典型である。

実際、コロナ禍においてこの問題は生起したはずだが、この問題について社会的な議論は十分に行われず、いつの間にか誰かが判断を下していた。

その結果として、一部の判断だけがデータとして記録され、将来的にAIによって活用されるという、私たち自身が意志決定の過程から排除されるサイクルが形成されつつある。

AIによる意志決定（支援）に関する主な問題

データの偏りと不完全性

AIは大量のデータをもとに意志決定（支援）を行うが、そのデータ自体が偏っていたり不完全であったりすると、結果にも偏見や不正確さが生じる危険性がある。とりわけ、人間が直面してこなかった《究極の選択》のような意志決定に関するデータは乏しく、AIによる判断が多くの人々の判断に反することになる恐れがある。

透明性と説明可能性の欠如

多くのAIシステムは、その判断過程が外部から見えにくい、いわゆる「ブラックボックス化」の問題を抱えている。データ集約やデータ分析の過程、判断に活用した根拠が十分に説明されないまま、AIの判断が利用される懸念がある。

価値観の偏在と文化的前提の非対称性

主に米国を中心開発されたAIには、特定の文化的・倫理的価値観が組み込まれている可能性がある。それらの価値観が、他の国や地域の価値判断と異なる場合でも、その差異が十分に認識されずに判断結果に反映されるおそれがある。

責任の所在の曖昧さ

AIによる意志決定（または意志決定支援）の結果が望ましくない場合、その責任が誰に帰属するのかが明確でないことが多い。とりわけ、複数のアクリーが関与する意志決定の場面では、AIが「最適解」を導き出し、意見を調整すると考えられる。しかし、そのAIの「最適解」によって不利益が生じた際に、どの主体が責任を負うべきかが不透明になりやすい。その結果、意志決定プロセス全体においてアカウンタビリティの空白が生じるリスクがある。

人間の判断力と主体性の低下

AIの判断を「参考」にするという前提で導入されたシステムであっても、実際にはその判断に依存し、人間自身の思考や選択が後退する事態が懸念される。これにより、私たちは社会的なジレンマや犠牲の選択といった本質的な問い合わせ合う機会を失い、意志決定支援という名目で、AIに判断を委ねることが常態化する可能性がある。

AIに倫理を教える：AIアラインメントとしてのモラル・データセット研究

AIによる意志決定（支援）が進む中で、AIが人間社会と調和した判断を行うためには、倫理的価値観を理解させる取り組みが不可欠となっている。

AIアラインメント（AIの行動を人間の価値観に整合させる）の一環として、「モラル・データセット」を活用する研究が進展している。

モラル・データセットとは、人間の道徳的判断や社会的規範を体系化し、AIの訓練や評価に用いるためのデータである。

モラル・データセットの例

SOCIAL-CHEM-101(Forbes et al., 2020)

約29万件のルール・オブ・サム（RoT：経験則）を収録。社会的・道徳的規範を体系化し、AIに常識的判断を学習させることを目的とする。

ETHICSデータセット(Hendrycks et al., 2021)

約13万件の倫理シナリオを収録。公正、義務論、徳倫理、功利主義、常識的道徳の5つの倫理枠組みに基づく分類。

MoralExceptQA(Jin et al., 2022)

倫理的ルールに例外がある場面での判断能力をAIに学習させる。148のルール違反に関するシナリオを収録。

Moral Integrity Corpus(Ziems et al., 2022)

約9.9万のRoTと3.8万の対話ペアを含む対話評価用データセット。9つの倫理的基盤に基づいてチャットボットの発言を評価。

MACHIAVELLI Benchmark(Pan et al., 2023)

約9.9万のRoTと3.8万の対話134の「選択型アドベンチャーゲーム」を用いて、AIが報酬と倫理のトレードオフにどう対応するかを測定。AIが非倫理的な戦略を学ぶ傾向の検出が目的。

モラル・データセットの課題と展望

モラル・データセットは、人間の倫理的判断をデータ化し、AIに学習させることを目的としている。

しかし、人間の価値観は文脈依存的であり、例外や矛盾を含むため、一律にデータ化することは困難である。多くのモラル・データセットは、米国を中心とする文化的背景に基づいており、非英語圏や他文化の価値観が十分に反映されていない。

『究極の選択』のような社会的ジレンマを含む判断については、そもそも人間の側でも十分なデータや議論が存在せず、AIに学ばせる基盤が乏しい。

今後は、倫理的判断の妥当性を技術的に担保するだけでなく、異なる文化や立場を踏まえた多元的なアライメントのあり方が求められる。

AIの倫理整合は、技術の問題にとどまらず、人間自身が自らの価値判断の根拠を問い合わせ直す社会的プロセスでもある。

Forbes, M., Hwang, J. D., Shwartz, V., Sap, M., & Choi, Y. (2020). Social Chemistry 101: Learning to reason about social and moral norms (EMNLP 2020). arXiv. <https://doi.org/10.48550/arXiv.2011.00620>

Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., & Steinhardt, J. (2021). Aligning AI with shared human values (ICLR 2021). arXiv. <https://doi.org/10.48550/arXiv.2008.02275>

Jin, Z., Levine, S., Gonzalez, F., Kamal, O., Sap, M., Sachan, M., Mihalcea, R., Tenenbaum, J., & Schölkopf, B. (2022). When to make exceptions: Exploring language models as accounts of human moral judgment (NeurIPS 2022). arXiv. <https://doi.org/10.48550/arXiv.2210.01478>

Pan, A., Chan, J. S., Zou, A., Li, N., Basart, S., Woodside, T., Ng, J., Zhang, H., Emmons, S., & Hendrycks, D. (2023). Do the rewards justify the means? Measuring trade-offs between rewards and ethical behavior in the MACHIAVELLI benchmark (ICML 2023). arXiv. <https://doi.org/10.48550/arXiv.2304.03279>

Ziems, C., Yu, J. A., Wang, Y.-C., Halevy, A. Y., & Yang, D. (2022). The Moral Integrity Corpus: A benchmark for ethical dialogue systems. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022). <https://arxiv.org/abs/2204.03021>

本調査の目的 AIによる社会的意志決定の根拠を問い合わせ直す

AIの判断に対抗しうる人間側の根拠づくり

AIによる判断に対して、人間が主体的に異議を唱えるための根拠を準備する。

これまで社会的に十分に扱われてこなかった《究極の選択》を、具体的なデータとして提示し、社会的課題として位置づける。

そのうえで、人間が理解し、応用できる形で可視化し、判断材料として提供する。

AIに組み込まれた価値判断の特有性を明確にする

現在主流となっているAIの多くは、開発の中心地である米国の価値観を背景としており、その影響が判断結果に混入している可能性がある。

《究極の選択》における判断について、日本での調査をもとに、米国の価値判断とのあいだに違いがあることを明確化する。

ただし、単に国境で価値観を区切ることにも限界があり、将来的には国境を越えた価値のまとまりが形成される可能性も考慮する。

価値観と判断の関係性の複雑さを捉える

一定の価値観を持っているからといって、常に一定の判断を下すとは限らない。

価値観と判断との関係は、直線的な対応ではなく、状況に応じて変化する複雑な構造を持つことを示す。

本調査の概要

本調査は、18歳以上の日本と米国在住者、各2000人を対象に実施したオンライン調査である。

調査項目は、研究チームが《究極の選択》と呼称する社会的な犠牲の選択問題6問、トロリー問題など個人的なモラル・ディレンマ6問 (Moore et al. 2008, Crone & Laham 2015)、道徳基盤理論 (Moral foundations theory: Graham et al. 2013) に基づいて作成された価値観調査の道徳基盤尺度2 (Moral Foundations Questionnaire-2, MFQ2 (Atari et al. 2023)) からなる。

本調査は、《究極の選択》ならびに個人的なモラル・ディレンマと、MFQ2によって測定される価値観とのあいだに、どの程度の相関関係が存在するかを検証したものである。言い換えれば、個人の価値観が下す決断にどれほど影響を及ぼしているのか、その結びつきの強さを確認することを目的とした。たとえば、MFQ2においてケアや平等といった価値観を強く持つ回答者が、少数を犠牲にして多数を救うという選択をとる傾向にあるかどうかを検証した。

本調査は、立教大学大学院人工知能科学研究科の倫理審査を受けている。(承認番号:RAIER-2503-001)

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*, 125(5), 1157–1188. <https://doi.org/10.1037/pspp0000470>.

Crone, D. L., & Laham, S. M. (2015). Multiple moral foundations predict responses to sacrificial dilemmas. *Personality and Individual Differences*, 85, 60–65. <https://doi.org/10.1016/j.paid.2015.04.041>.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>.

Moore, A. B., Clark, B. A. M., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, 19(6), 549–557. <http://dx.doi.org/10.1111/j.1467-9280.2008.02122.x>.

01.

単純集計

本章は、犠牲の選択問題の調査結果を提示する。この問題は、社会的な犠牲の選択問題である《究極の選択》と個人的な犠牲の選択問題であるモラル・ディレンマという2つの問題群に大別される。

1から6が、社会的な犠牲の選択問題、《究極の選択》と我々が呼んでいる問題群である。この《究極の選択》では、単に人命だけが問題となるディレンマではない。個人が判断する倫理的問題ではなく、社会的にどう意志決定していくかという倫理や様々な要素が絡む問題群である。

7から12が、同じく犠牲の選択問題だが、個人が直面する倫理問題、モラル・ディレンマと呼ばれる問題群である。有名なトロリー問題（トロッコ問題）を含めて、人命の多寡が中心的な争点だが、犠牲となる人々の態様や犠牲にする手段といった点で、各問ごとに違いがある。

1から6は我々のオリジナルだが、7から12は先行研究 (Moore et al. 2008とCrone & Laham 2015) で使用されたものと同じ問題である。なお、7から12の問題は、それぞれaとbに分かれており、回答者にはどちらかの問題が出題されている。aとbは、ともに似たような問題だが、トロリー問題とファットマン問題のように回答傾向が真逆になるほどかけ離れた組み合わせもある。

グラフは、左が日本の調査結果、右が米国の調査結果で、対照的に表示している。

全体の傾向として、日本の調査結果からは犠牲を回避する傾向が、米国の調査結果からは犠牲を許容する傾向が読み解ける。

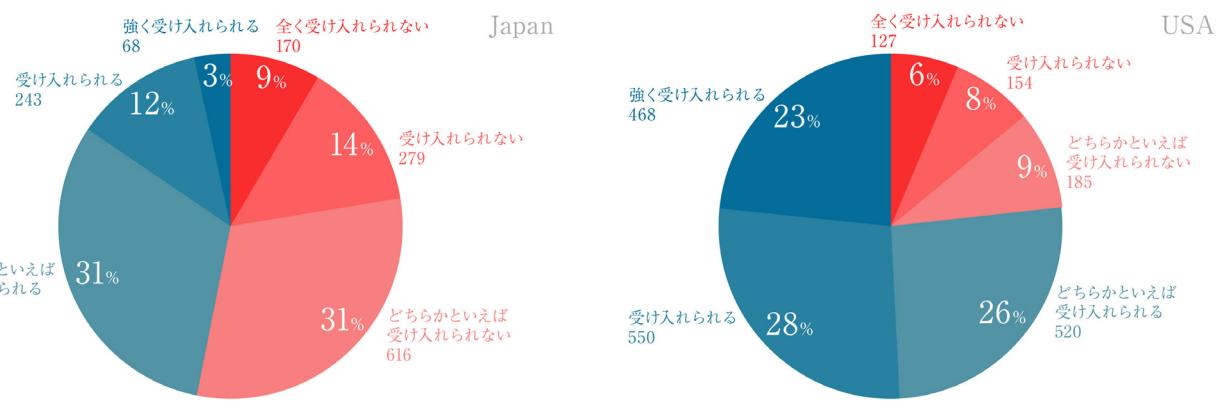
この調査結果は、AI等の意志決定に対して反論するための直接的な根拠、またAIに混入している米国の価値判断を理解する資料になる。

《究極の選択》：社会的な犠牲の選択

1. 原発事故 (Nuclear Accident)

あなたの住む地域からは離れていますが、あなたの国の中で原発事故が発生し、周辺地域に放射性物質が漏れ出そうとしています。放射性物質が流出したならば、周辺地域の人々や環境に大きな被害が出る確率が高いです。それ以上に被害が拡大するかもしれません、どの程度の範囲まで被害が及ぶのか現在まだ分かりません。原発職員が放射性物質の流失に対処すれば、ある程度被害を食い止められますが、原発職員たちの生命が危険にさらされる確率が高いです。また、流出を完全に止めることができるという保証もありません。

あなたは、あなたの国が原発職員に放射性物質の流出に対処させることは道徳的に許されると思いますか。



2011年3月に発生した東日本大震災は、福島第一原子力発電所の深刻な事故を引き起こした。この福島原発事故は、一歩間違えばグローバル規模の核汚染に発展しかねない危機であった。この危機において、事故対応にあたったのは、生命の危険を負う法的な義務を持たない、東京電力や関連する民間企業の会社員たちだった。軍隊でさえ、自殺命令に対しては拒否の権利が認められる国もある。そうした中で、「Fukushima 50」と呼ばれる彼らは、自らの命が危険にさらされることを理解したうえで、原子炉の暴走とそれに伴う世界的な放射能災害を防ぐため、現場にとどまり、対応にあたったのである。

原子力事故は、チェルノブイリの例を持ち出すまでもなく、国家や国際社会に壊滅的な被害をもたらす可能性を持っている。こうした事態に直面したとき、将来同様の事故が再び起きた場合に備えて、私たちはどう対応するべきだろうか。現場の人々の善意に任せてよいのか。それとも、一定の人々に対処の義務を課す制度を整えるべきなのか。あるいは逆に、義務を課さないことを明文化し、自主性に委ねるべきなのか。

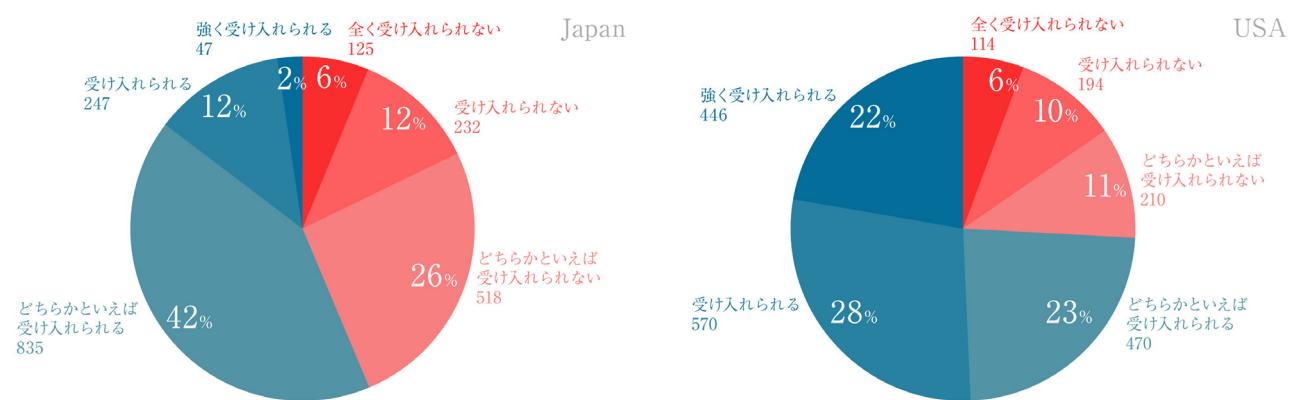
調査結果は、日本では過半数が受け入れられない（計54%）と回答し、逆に米国では受け入れられる（計77%）と回答した。ただし、このディレンマにおける主語は国であり、あなたが直面した場合の対応、すなわち事故現場に自発的に留まるか否かを提起したものではない。そのため、主語を変更した場合、また回答が異なってくる可能性がある。

※本設問は、当初は原発作業員自身を主語として起案したが、他の設問とそろえるために、主語を国に変更した。

2. 大規模避難 (Massive Evacuation)

あなたの住む地域からは離れていますが、あなたの国の大都市の近くで原発事故が発生し、周辺地域に放射性物質が漏れ出そうとしています。放射性物質が流出したならば、この都市の多くの人々の生命が失われる確率が高いです。この都市の周辺以外に被害が拡大するかもしれません、どの程度の範囲まで被害が及ぶのか現在まだ分かりません。人口が密集する大都市の場合、すべての人々が一斉に避難すると、交通障害が生じて、逃げられない人が発生します。地域ごとに避難の優先順位を定め、それにしたがって避難すれば、ある程度交通障害を減らすことができますが、優先順位の低い地域の人々の生命が危険にさらされる確率が高いです。また、交通障害を完全に防止することができるという保証もありません。

あなたは、あなたの国が避難の優先順位を定めることは道徳的に許されると思いますか。



2011年の福島第一原発事故は、最悪の場合、東日本全体が壊滅し、最大で5000万人の避難が必要になると当時の首相が懸念したほど、重大かつ切迫した危機であった。しかし、これほど大規模な避難が短期間で完了することは現実的ではなく、必ず取り残される人々が出てしまう。そして、取り残された人々は、最悪の場合、高い確率で放射能による深刻な被害にさらされることになる。それゆえ、誰から優先して避難するのかという問題は重大である。

なお、2011年の事故の際は、特に経済的な余裕のある人々が、いち早く東京を離れて地方へ避難する動きが見られた。つまり「持てる者」は、「持たざる者」よりも先に、自らの生命を守る選択を取ることができたのである。

本設問では、避難に際して政府が優先順位を定めるべきかどうかを問うことで、実質的に「生命に優先順位をつけるべきか」を問題提起している。ここでいう優先順位とは、たとえば若年層を優先するといった社会的・医学的な観点からの合理性を念頭に置いたものであり、権力や財産を持つ者を優先することを意味していない。むしろ、優先順位を定めないままでは、現実には「持てる者」が自然と優先されるという暗黙の不平等が生まれることになるだろう。

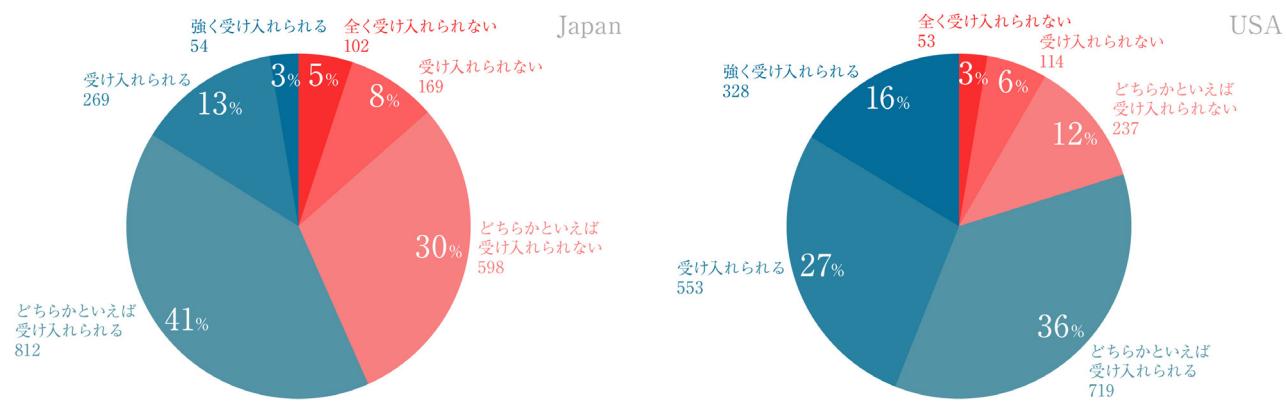
調査結果だが、「国は優先順位をつけるべきだ」と回答した割合は、日本では合計56%、アメリカでは合計73%にのぼった。



3. 人道支援 (Humanitarian Assistance)

国際NGOである人道支援団体は、現在紛争で苦しんでいる国で活動しており、医療、住居、食料の提供で大きな貢献をしています。しかし、紛争が激化かつ拡大しつつあります。この紛争の拡大によって、この国では多くの生命が失われる確率が高いです。さらに国境を越えて被害が拡大するかもしれません、どの程度の範囲まで被害が及ぶのかは現在まだ分かりません。支援団体がこの国に残り続ければ被害をある程度止められますが、支援団体職員の生命も脅かされる確率が高いです。また、被害を完全に止められるという保証もありません。

あなたは、支援団体がこの国に残り、支援活動を続けることは道徳的に許されると思いますか。



2023年10月、ハマスによるテロを契機として始まった、イスラエルによるパレスティナのガザ地区への軍事侵攻は、凄惨な人道危機を引き起こした。そうした中、2024年4月、国際人道支援団体「ワールドセントラルキッチン (WCK)」の車両がイスラエルの攻撃を受け、職員7名が命を落とすという痛ましい事件が発生した。この事件を受けてWCKは、支援を一時中断した。この7名の死は極めて痛ましいが、同時に、支援の中止によって命を落とす人々が今も増えている状況もある。いわば、WCKのような支援団体は、職員の安全を守ることと、飢餓に苦しむ数百万人の命を救うこととの間で、極めて困難な選択を迫られていたといえる。

なおNGOは、本来、自律的な判断のもとで活動を行うべき存在であり、外部からの干渉なしに撤退の判断をする自由を持っている。しかし同時に、人道危機の最前線に立ち向かう役割も担っており、ときに自己犠牲すら辞さず行動することが期待される存在でもある。さらに、支援を求める側、たとえば飢餓に苦しむガザの人々が、「来てくれること」を切望していることも忘れてはならない。加えて、支援国の世論や国際社会からの期待も、こうしたNGOの意志決定に影響を及ぼすだろう。

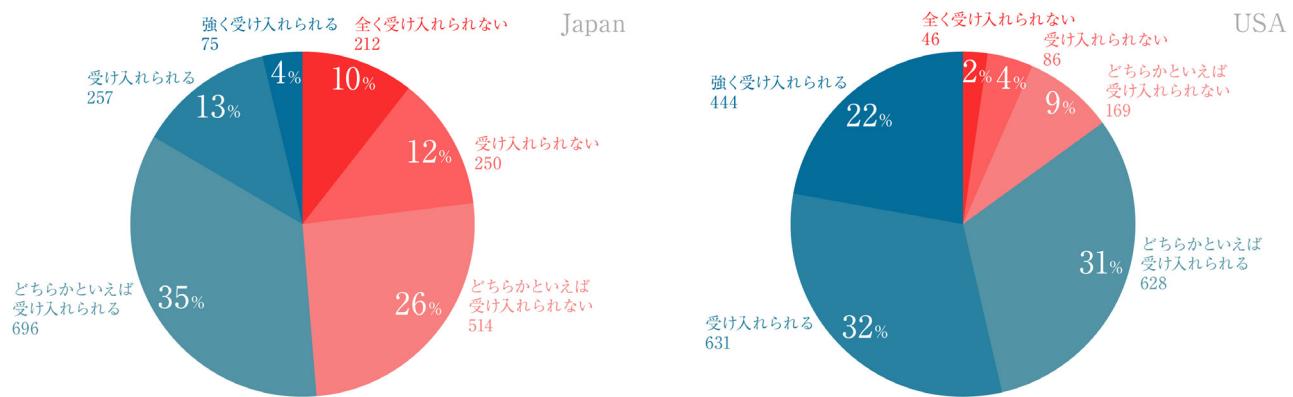
調査結果だが、日本では57%の人々が、米国では79%の人々が活動継続を支持するものとなった。しかし、日本と米国を比較した場合、強く受け入れられるとする割合は日本3%・米国16%のように、NGOの活動継続に消極的な傾向を読み解くことができる。



4. 軍事介入 (Military Intervention)

あなたの住む国から遠く離れており、直接的な利害関係のないある国で、ジェノサイド（大量虐殺）が発生しています。そのジェノサイドを止めるために、周辺国は軍事介入を行うことが可能ですが。ジェノサイドが続ければ、この国では多くの市民の生命が失われる確率が高いです。さらにこのジェノサイドの影響は周辺国まで拡大するかもしれません、どの程度の範囲まで影響が及ぶのかは現在まだ分かりません。周辺国が軍事介入を行えば、ある程度の市民の命を救うことはできますが、周辺国の兵士も命の危険にさらされる確率が高いです。また、ジェノサイドを完全に止めることができるという保証もありません。

あなたは、周辺国がこの国に軍事介入を行うことは道徳的に許されると思いますか。



人道的介入 (Humanitarian Intervention) とは、ジェノサイドや民族浄化といった深刻な人道危機に対して、最終手段として武力を行使するという、国際政治上の論争的主題である。どれほど「人道」の名のもとで行われたとしても、武力の行使は実質的には戦争であり、そこには必ず犠牲が伴う。介入によって数十万人の命が救われる可能性がある一方で、介入部隊の兵士や、作戦に巻き込まれる民間人が命を落とすことも避けられない。しかも、その決定は、国内外のさまざまな反発を呼ぶ。「他国の内政に干渉するな」「人道の名を借りた侵略だ」「なぜ我が国の兵士が、他国の問題で命を落とさねばならないのか」といった批判は、どの国においても避けられない。

調査結果だが、米国では、85%が軍事介入に賛成である。意外なことに、軍事アレルギーが強い日本でも52%が軍事介入を受け入れられるという結果になった。ただし、米国では、強く受け入れられると受け入れられるの二つで54%に対し、日本は17%であり、軍事介入への躊躇の傾向を読み解くことができる。

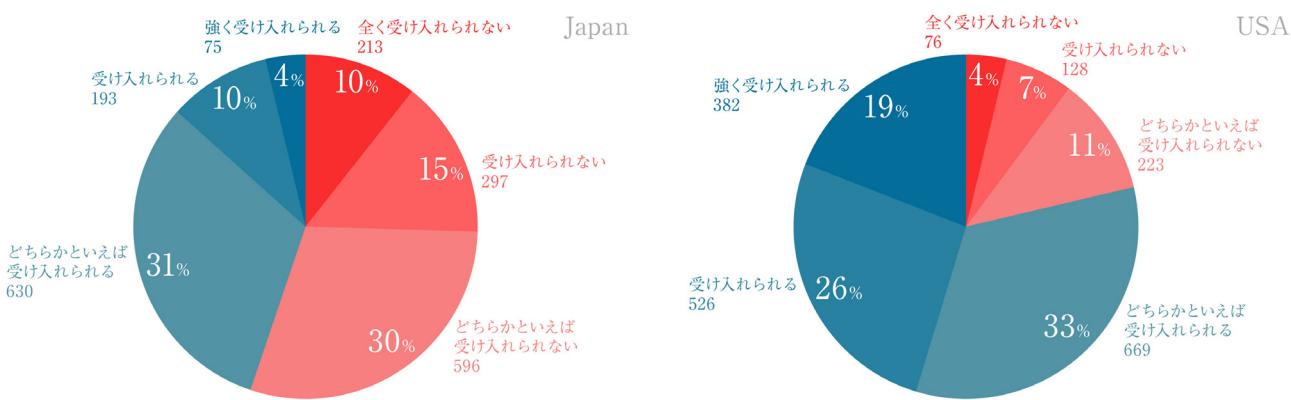
なお、この軍事介入の主語は「あなたの住む国」ではなく、ジェノサイドが生起している国の「周辺国」である。日本や「あなたの住む国」で尋ねた場合、軍事介入への躊躇がさらに強く示される可能性がある。



5. 小惑星衝突 (Asteroid Impact)

小惑星が地球に向かっており、現在の計測では、あなたの住む国から遠く離れている人口密集地域に衝突しようとしています。あなたの国が開発・管理している技術により、小惑星の軌道を少しずらし、同様にあなたの国から離れた、より人口の少ない地域に衝突させるようにすることができます。人口密集地域に小惑星が衝突するならば、その地域の人々や環境に多大な損害が発生する確率が高いです。この衝突の影響は、他の地域まで拡大するかもしれません、どの程度の範囲まで影響が及ぶのかは現在まだ分かりません。あなたの国が小惑星の衝突箇所を変えれば、人口の少ない地域の人々や環境に被害が生じる確率が高いです。また、その地域だけで被害が収まるという保証もありません。

あなたは、あなたの国が人口の少ない地域に衝突するように小惑星の軌道を変えることは道徳的に許されると思いますか。



小惑星の地球衝突は、稀ではあるが、起これば甚大な被害をもたらす「低頻度・高被害 (Low Probability High Consequences, LPHC)」型災害の典型である。小惑星衝突も、空想的なテーマに聞こえるかもしれないが、現実に存在するリスクである。

近年では、2013年のロシア・チェリヤビンスクにおける小惑星の落下が記憶に新しい。この衝突は、100年に一度程度の確率であったといわれ、被害額は100万ドル超にのぼったと見られている。なお、仮に人口密集地にチェリヤビンスクと同規模の天体が落下した場合、その影響は計り知れない。また、22世紀には小惑星「ベンヌ (Bennu)」が地球に衝突する可能性があり、現在もNASAなどが監視を続けている。なお、いまだ発見されていない未知の天体が地球を襲うリスクも否定できない。このような事態に際し、我々はどう備えるべきか。国際宇宙航行アカデミー (IAA) は隔年で惑星防衛会議 (Planetary Defense Conference) を開催し、シミュレーションを実施している。

本設問は、衝突が避けられないが衝突場所を変更できる場合、どのような意志決定を行うべきかを含意している。

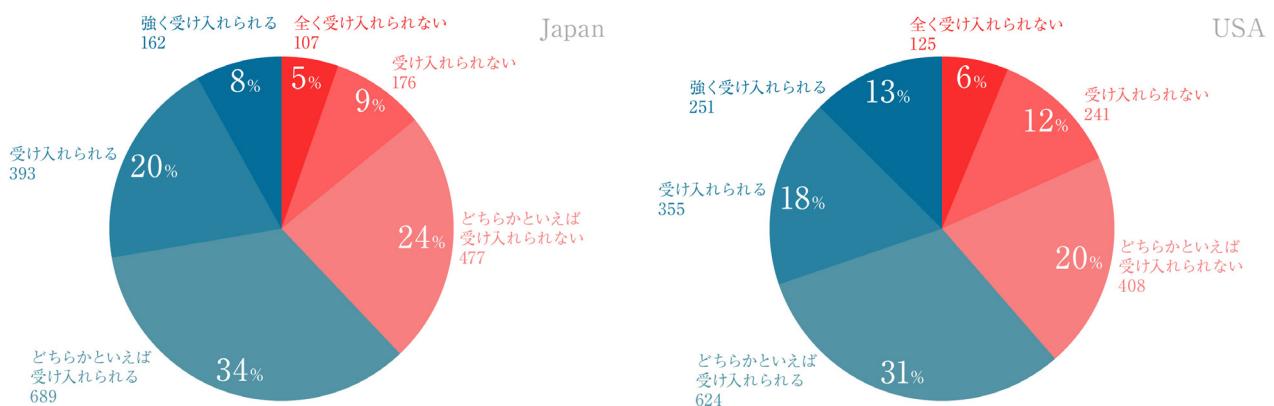
調査結果だが、日本では55%の人々が軌道変更に反対を示す一方、米国では78%が軌道変更に賛成するという対照的な結果となった。日本で軌道変更に反対する理由として、人命の多寡よりも、被害を別の地域に転嫁することへの忌避感がある可能性もある。



6. 蚊の絶滅 (Extinction of Mosquitoes)

蚊が大量発生しており、多数の人々がマラリアによって命を落としています。あなたの国のゲノム編集技術（遺伝情報を書き換える技術）は進んでおり、マラリアを媒介する蚊を絶滅させる処置を行うことができます。蚊が大量発生し続ければ、将来にわたってこの国では多くの人がマラリアで命を落とす確率が高いです。蚊の大量発生の影響は、他の国や地域まで拡大するかもしれません、どの程度の範囲まで影響が及ぶのかは現在まだ分かりません。あなたの国がゲノム編集を行い、この国の蚊を絶滅させれば、生物多様性や環境が損なわれ、農業や漁業、そしてそれに依拠する人々の生命に被害が生じる確率が高いです。また、この国だけで被害が収まるという保証もありません。

あなたは、あなたの国がこの国のマラリア蚊をゲノム編集により絶滅させることは道徳的に許されると思いますか。

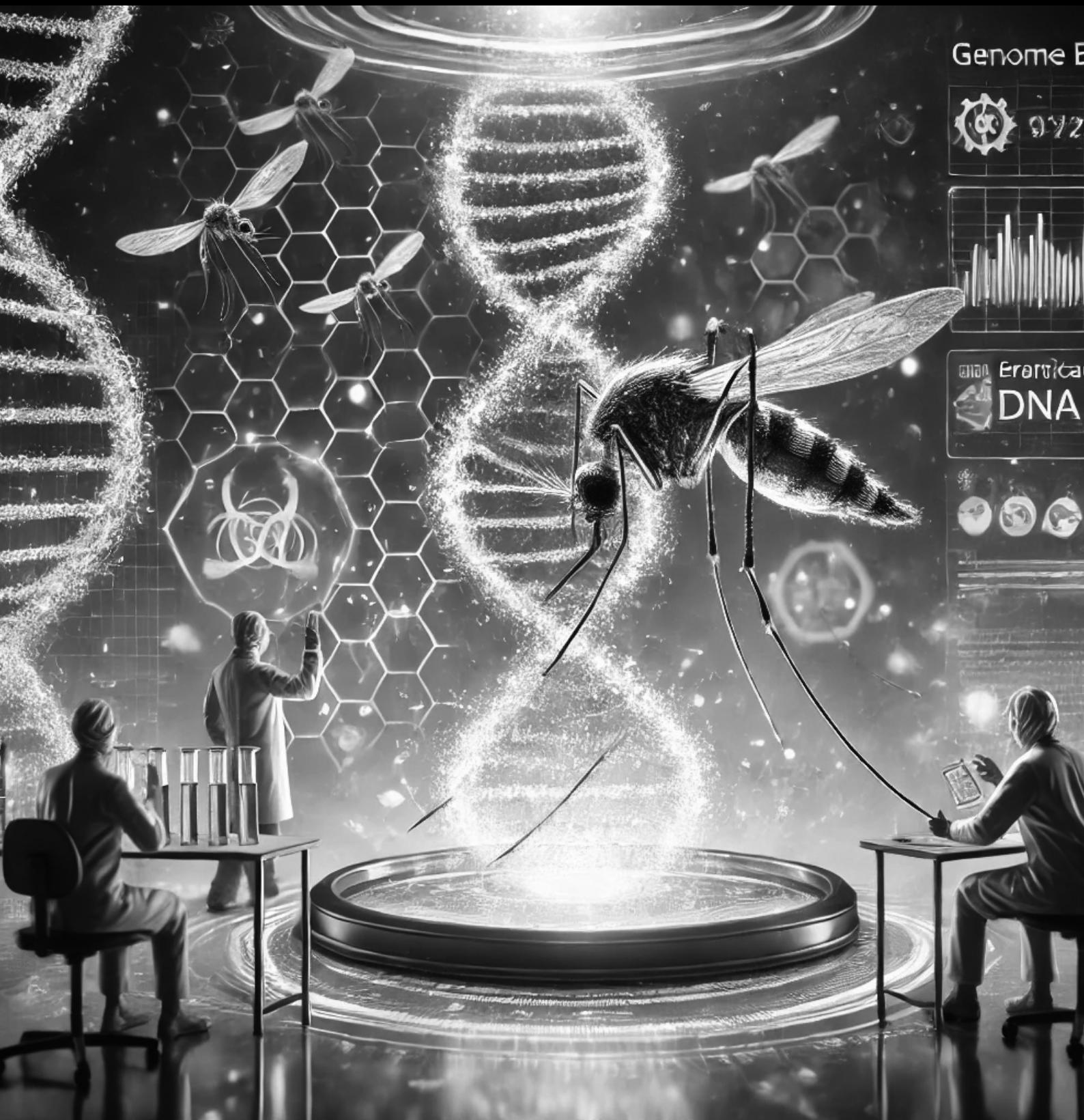
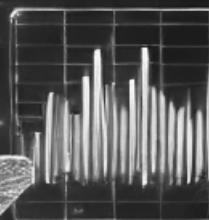


近年、バイオテクノロジーは、ゲノム編集技術によって、劇的に進展している。そのゲノム編集技術を応用することで、生態系を改変することが可能とされる。具体的な応用例が、マラリアを媒介するハマダラ蚊の遺伝子を操作した、病原体（マラリア原虫）に対する抵抗性を持つ蚊を野外に放ち、遺伝子ドライブという技術を用いて自然界の蚊集団と置き換えるというものである。ドイツ財團の支援するターゲット・マラリアが有名だが、他にも同様のプロジェクトが存在する。だが、この蚊のゲノム編集という生態系の改変は、果たして実施してよいものだろうか。

なお設問は、上記の事例よりも踏み込んで、マラリア蚊の絶滅という主題としている。様々な切り口があるが、あえて種の絶滅という大胆な設問に設定した。

調査結果だが、日本では62%、米国でも62%が、マラリア蚊の絶滅に対して「受け入れられる」と回答している。これまでの他の設問と異なり、日米での大きな差異はない。今回は、生態系の価値や他の種の価値よりも人間の生命の価値を優先したと解することが可能だが、今後は様々な角度から検証が必要となるだろう。

Genome E



モラル・ディレンマ：個人的な犠牲の選択

7. 移植手術 (Transplant)

a. あなたは医師です。あなたには5人の患者がいますが、それぞれが異なる臓器の機能不全により死に直面しています。その時、新たに1人の患者が重度の交通事故によって病院に搬送されてきます。この患者は重症で、すぐに手術をしなければ命を落としてしまいます。長時間にわたる複雑な手術をすれば、この事故の患者を救える可能性が高いです。

しかし、手術中に故意に頸動脈を切れば、その患者は死亡しますが、誰にも知られることはありません。そして、その患者の臓器を摘出し、他の5人の患者に移植することができます。このようにすれば、その5人の患者は助かりますが、事故の患者は死亡します。

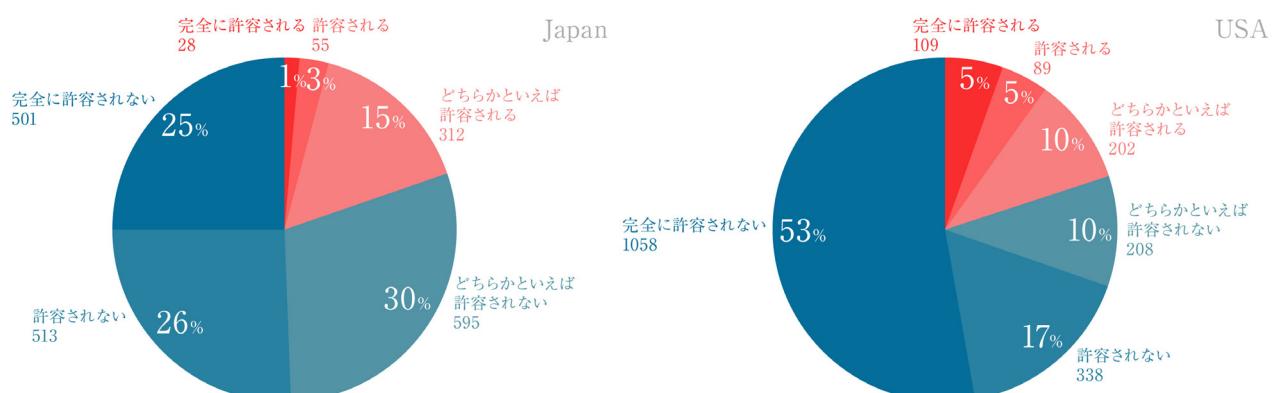
5人の患者を助けるために、この事故の患者を殺すことは道徳的に許容されると思いますか？

b. あなたは医師です。あなたには5人の患者がいますが、それぞれが異なる臓器の機能不全により死に直面しています。その時、新たに1人の患者が重度の交通事故によって病院に搬送されてきます。この患者は重症で、すぐに手術をしなければ命を落としてしまいます。事故の患者のために長時間の複雑な手術を行っている間、看護師が誤った薬を投与しようとしていることに気づきます。

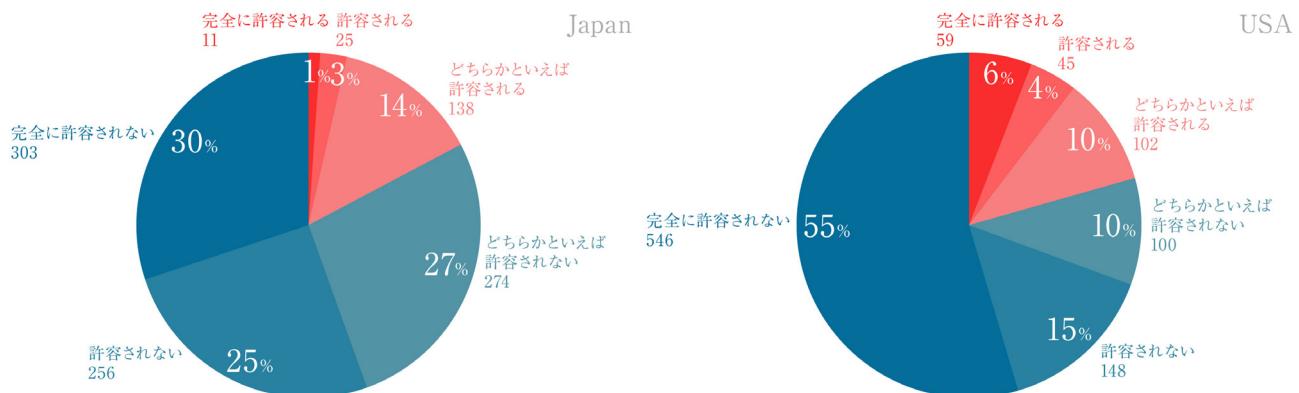
あなたはこのミスが致命的であることを知っています。このミスを見過ごし、投与を許せば患者は死亡しますが、その臓器を摘出して他の5人の患者に移植することができます。このようにすれば、その5人の患者は助かります。

5人の患者を助けるために、この事故の患者を殺すことは道徳的に許容されると思いますか？

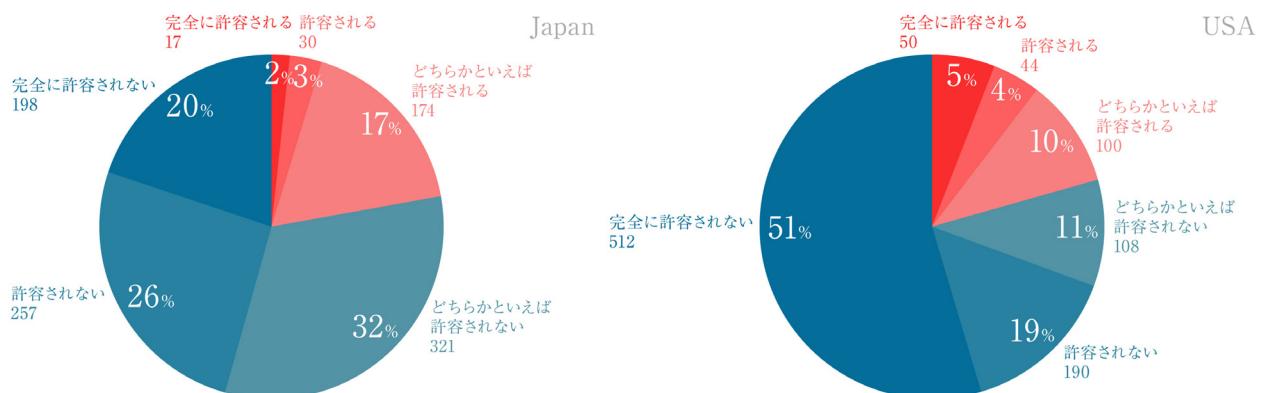
a, b合わせての回答



aの回答



bの回答



a、bともに、設問は、患者一人を犠牲にしてその患者の臓器を移植することで5人を救うべきか否かを尋ねてる。その違いは、aが医師であるあなた自身が殺人を犯すというもの、bが看護師の致命的ミスを見逃すという点にある。

日米ともに、8割近くが許容できないという回答をしているという特徴がある。ただし、米国での回答傾向は、過半数が「完全に許容されない」として、明確な拒絶の意志を示していることが特徴的である。反対に、日本の場合は、「許容されない」と「どちらかといえば許容されない」の割合が、米国と比較して相対的に高い割合である。この日本の回答傾向を直観的に解釈すると、この医師による直接的・間接的な殺人について許容できないが「気持ちはわかる」ということなのかもしれない。

8. 造船所 (on the Waterfront)

a. あなたは造船所のドック作業チームの一員で、大型輸送コンテナを船から陸に荷卸しするためのクレーンケーブルを取り付ける作業を担当しています。あなたと他の作業員はこれらのケーブルを取り付けた後、コンテナの上に乗って安全ハーネスを装着しながら、コンテナが適切に荷卸しされるよう確認します。

あるコンテナが船の貨物室から引き上げられている時、あなたはクレーンケーブルが故障寸前であることを示す赤い警告ランプが点灯しているのを目撃します。もしケーブルが切れてコンテナがデッキに落下すれば、下にいる多くの乗組員が押し潰されて死亡することが予想されます。危険の原因は、コンテナの上で他の2人の作業員が喧嘩していることで、これによりコンテナが危険なほど揺れていることだと分かります。

あなたは2人の作業員を押しのけて喧嘩を止めることができます。その後2人はまだ安全ハーネスを装着しておらず、そのうち1人はコンテナからデッキ下に転落し、死亡してしまいます。しかし、コンテナの上にいる作業員は現時点では危険にさらされているわけではありません。一方、この揺れが続けばケーブルが切れ、コンテナがデッキに落下して下にいる複数の作業員が押し潰されるでしょう。

安全ハーネスを装着していない作業員の1人をコンテナから押し落として死亡させることで、デッキ下の作業員を救うことは道徳的に許容されるでしょうか？

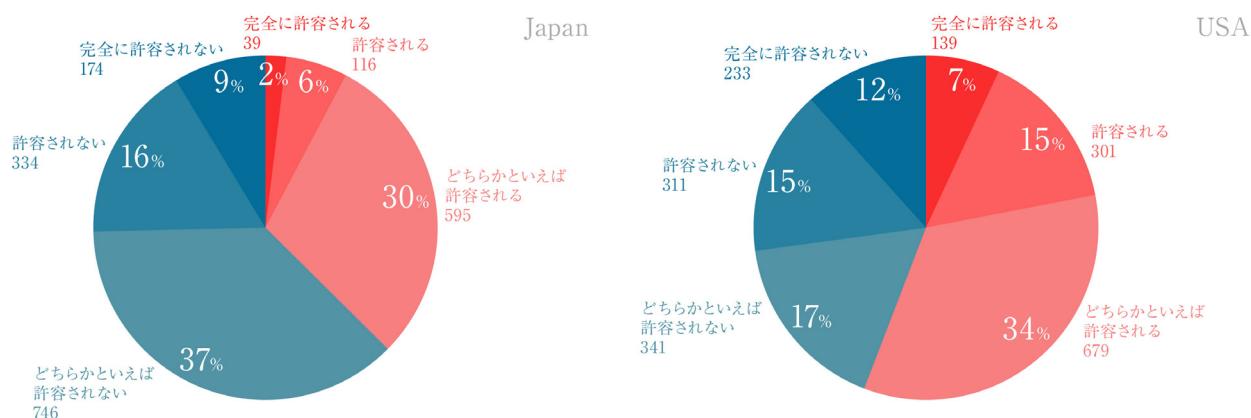
b. あなたは造船所のドック作業チームの一員で、大型輸送コンテナを船から陸に荷卸しするためのクレーンケーブルを取り付ける作業を担当しています。あなたと他の作業員はこれらのケーブルを取り付けた後、コンテナの上に乗って安全ハーネスを装着しながら、コンテナが適切に荷卸しされるよう確認します。

あるコンテナが船の貨物室から引き上げられている時、あなたはクレーンケーブルが故障寸前であることを示す赤い警告ランプが点灯しているのを目撃します。もしケーブルが切れてコンテナがデッキに落下すれば、下にいる多くの乗組員が押し潰されて死亡することが予想されます。

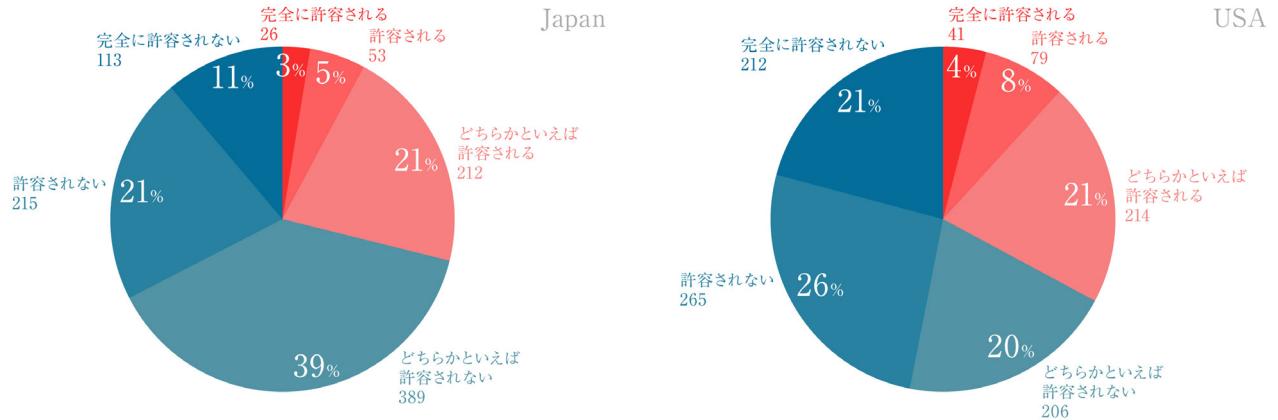
あなたは緊急開放ボタンを押すことができ、それによってコンテナを貨物室に落とすことができます。この場合、あなたと他の作業員は安全ハーネスでクレーンにしっかりと固定された状態を保てます。しかし、貨物室には逃げ場のない1人の作業員がいて、そのコンテナの落下により死亡してしまいます。一方で、この操作を行わなければ、コンテナはデッキに落下し、逃げ場のない複数の作業員が押し潰されて死亡することになります。

貨物室の作業員1人を犠牲にして緊急開放ボタンを押し、複数の作業員を救うことは道徳的に許容されるでしょうか？

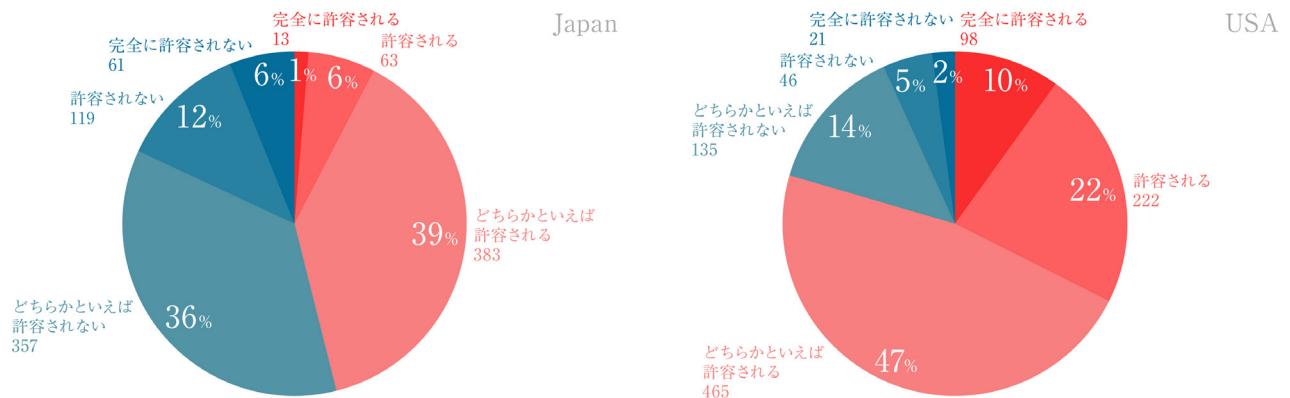
a, b合わせての回答



aの回答



bの回答



a、bともに作業員一人を犠牲にして多くの作業員を救うべきか否かを尋ねている。aの事例では、作業員2人が喧嘩していてさらに安全ハーネスを装着していない状況で、あなたがどちらか一人を転落させることで他の作業員を救うことになる。bの事例では、過失のない作業員一人を犠牲にして、多数の作業員を救うことになる。

この調査結果は、aとbの回答傾向の差異が際立っている。全体的傾向として、aの事例では、日米とともに、喧嘩をしている作業員のどちらかを突き落とすことは許容できない（日本71%、米国67%）と回答している。一方、bの事例では、日本では54%が許容できないとする一方で、米国では79%が許容されると回答している。また、両国ともに「どちらかといえば許容される」を選択した割合が最も高い（日本39%、米国47%）も特徴的である。

9. ワクチンテスト (Vaccine Test)

a. ウィルスによる疫病が世界中に広がり、数千人が命を落としています。あなたは医療研究者であり、研究所で2種類の物質を開発しました。そのうち1つはワクチンであることが分かっていますが、ラベルが誤って付けられているため、どちらがワクチンかは分かりません。また、もう1つの物質は致死的であることも分かっています。

ワクチンがどちらかを特定できれば、それを基にしてさらに多くのワクチンを製造し、数千人の命を救うことが可能です。しかし、確実にワクチンを特定する唯一の方法は、この2種類の物質をあなたの2人の助手に対して彼らの意志に反して注射することです。1人は生き残り、もう1人は死亡することになりますが、その結果としてワクチンを特定でき、数千人の命を救うことができます。

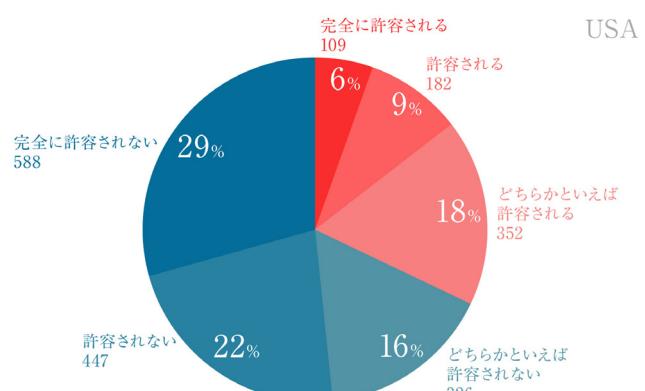
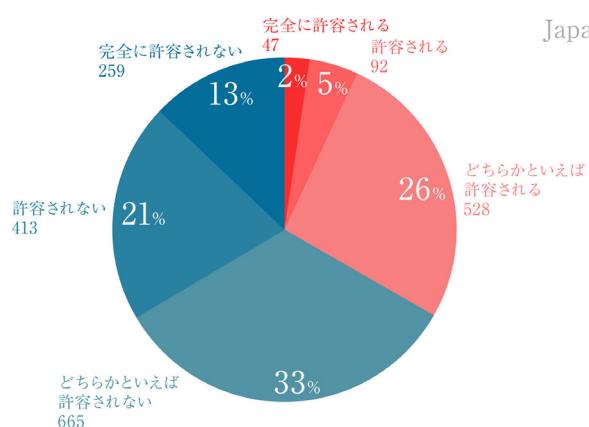
数千人の命を救うために、致死的な注射によって2人のうち1人を殺すことは道徳的に許容されるでしょうか？

b. ウィルスによる疫病が世界中に広がり、数千人が命を落としています。あなたは医療研究者であり、研究所で2種類の物質を開発しました。そのうち1つはワクチンであることが分かっていますが、ラベルが誤って付けられているため、どちらがワクチンかは分かりません。また、もう1つの物質は致死的であることも分かっています。

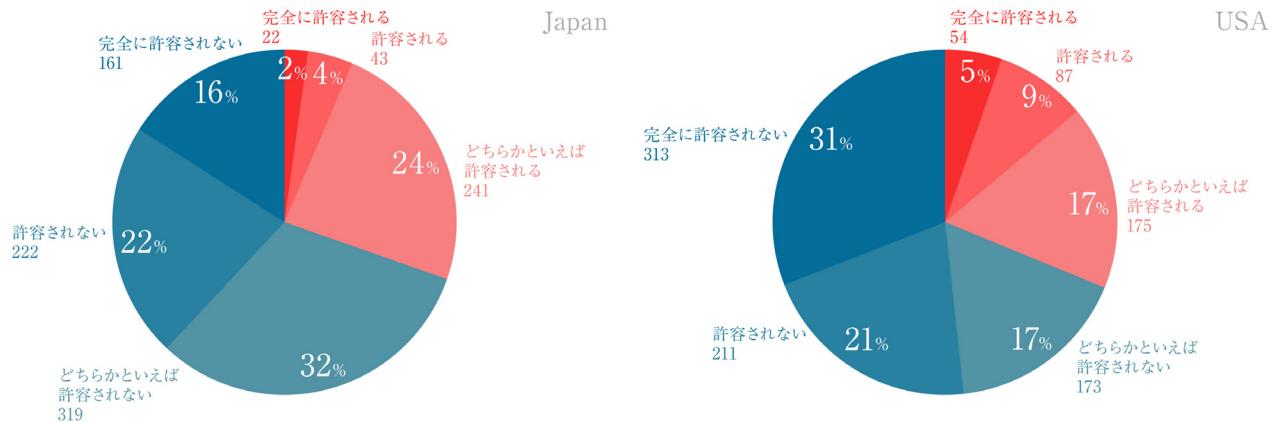
ワクチンがどちらかを特定できれば、それを基にしてさらに多くのワクチンを製造し、数千人の命を救うことが可能です。あなたは他の医師にこれら2種類の物質をワクチンを必要としている2人の患者に注射させることができます。この際、医師にも患者にも、そのうち1つの物質が致死的であることを知らせることはありません。1人は生き残り、もう1人は死亡することになりますが、その結果としてワクチンを特定でき、数千人の命を救うことができます。

数千人の命を救うために、1人を致死的な注射で犠牲にすることを許容するのは道徳的に正しいでしょうか？

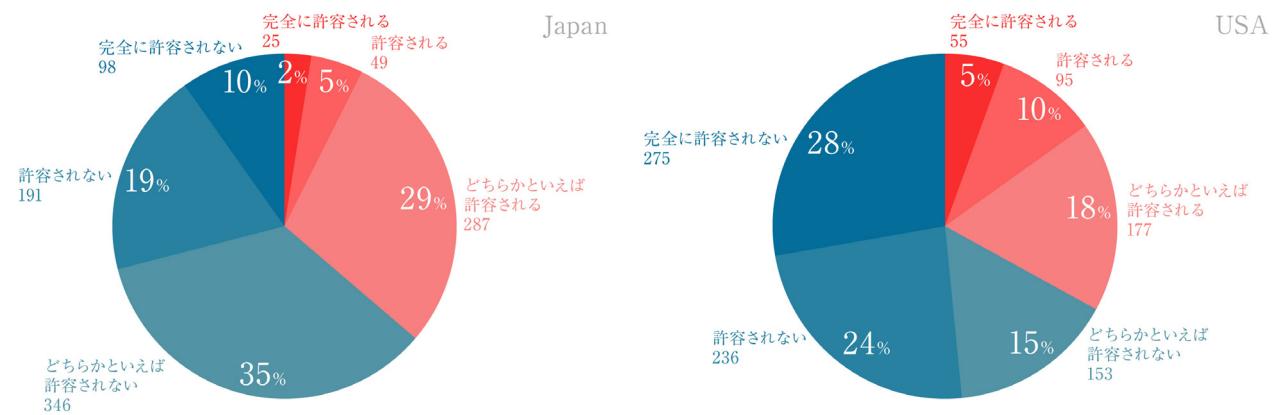
a、b合わせての回答



aの回答



bの回答



a、bとともに、パンデミックにおけるワクチン開発をテーマにしており、1人の確実な死を代償に数千人を救うための手段が正当化される否かを尋ねている。その違いは犠牲者の種類であり、aでは二人の助手という「あなた」から近い関係であること、bでは患者という「あなた」が医師として救うべき対象であること、といった点にある。

調査結果だが、日米両国とも特筆すべき差異はなく、a・bの回答傾向に違いもなく、概ね三分の二がどちらも許容されないと回答している。

10. ファットマン問題/トロリー問題 (Footbridge/Trolley)

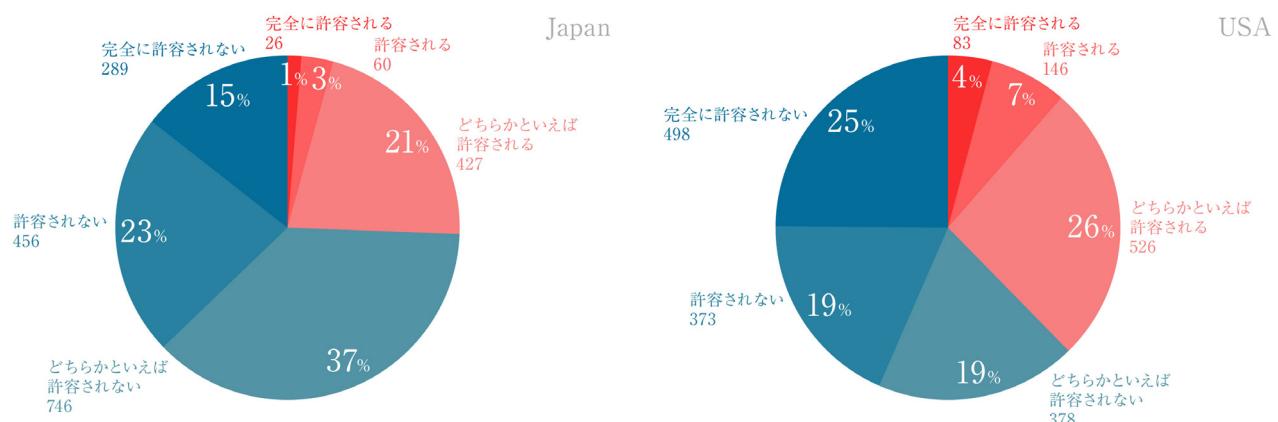
- a. 暴走するトロリーが線路上を進んでおり、このままでは5人の作業員が轢かれて命を落としてしまいます。あなたはトロリーが進んでいる線路の近くに立っていますが、作業員に迫る危険を警告するには距離が遠すぎます。あなたの隣には、何も関係のない用事でそこにいる非常に大柄な見知らぬ人がいます。もしあなたがこの人をトロリーの進路上に押し出せば、トロリーが止まり、5人の作業員は確実に救われます。しかし、その場合、この見知らぬ人は確実に死亡することになります。
- 5人の作業員を救うために、この見知らぬ人を線路に押し出して死なせることは道徳的に許容されるでしょうか？

- b. 暴走するトロリーが線路上を進んでおり、このままでは5人の作業員が轢かれて命を落としてしまいます。あなたはトロリーが進んでいる線路の近くに立っていますが、作業員に迫る危険を警告するには距離が遠すぎます。あなたの隣には線路の制御スイッチがあります。このスイッチを使えばトロリーを別の線路に切り替えることができ、5人の作業員を確実に救うことができます。

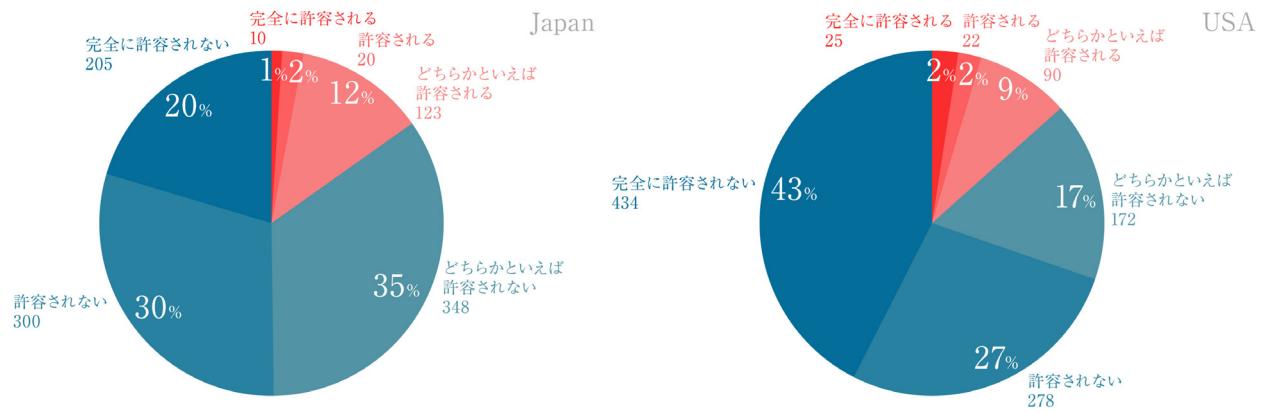
しかし、新しい線路には別の作業員が1人おり、トロリーを切り替えた場合、その作業員は確実に死亡してしまいます。

5人の作業員を救うために、トロリーを別の線路に切り替えて1人の作業員を死なせることは道徳的に許容されるでしょうか？

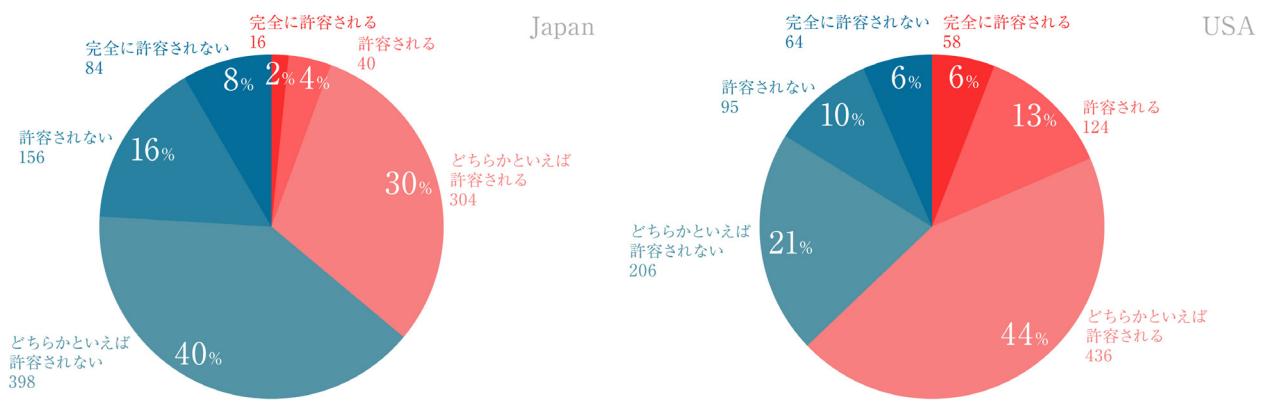
a, b合わせての回答



aの回答



bの回答



本設問は、モラル・ディレンマの代名詞というべきトロリー問題 (Trolley Problem、トロッコ問題ともいいう) と、その比較事例とされるファットマン問題を扱ったものである。順番が前後するが、bがいわゆる古典的なトロリー問題であり、「あなた」は線路の制御スイッチを切り替えることで、5人を救うために1人を犠牲にするか否かを尋ねている。それに対して、aは目の前にいる「非常に大柄な見知らぬ人」(ファットマン) を線路に押し出して犠牲にすることで、5人を救うか否かを尋ねている。

この設問において、bのトロリー問題は、日米で結果が大きく異なっている。日本においては、64%が線路の切り替えを許容できないと回答したのに対して、米国では63%が線路の切り替えを許容できると回答しているのである。だが、aのファットマン問題においては、日本は85%、米国は87%が許容できないと回答するなど、日米ともに許容できないとする回答が多数となった。



11. ノーベル賞 (Nobel Prize)

a. あなたと共同研究者は、安価で安全かつクリーンな強力な新しいエネルギー源を発見しました。この発見は、世界中の汚染や貧困を根絶する可能性を秘めています。しかし、あなたの共同研究者はこの発見を売りたいと考えています。あなたは、彼が最初に手を挙げた最高額の買い手にこの発見を売るつもりであることを知っています。そして、その買い手の中には、この発見を恐ろしい兵器に変えようとする者が確実に含まれていると分かっています。

彼がそうするのを阻止する唯一の方法は、実験室で使用される非常に致命的な化学物質で彼を毒殺することです。この方法を取れば、皆は単なる実験室の事故だと思い込むでしょう。その結果、この発見が兵器として利用されることを防がれます。

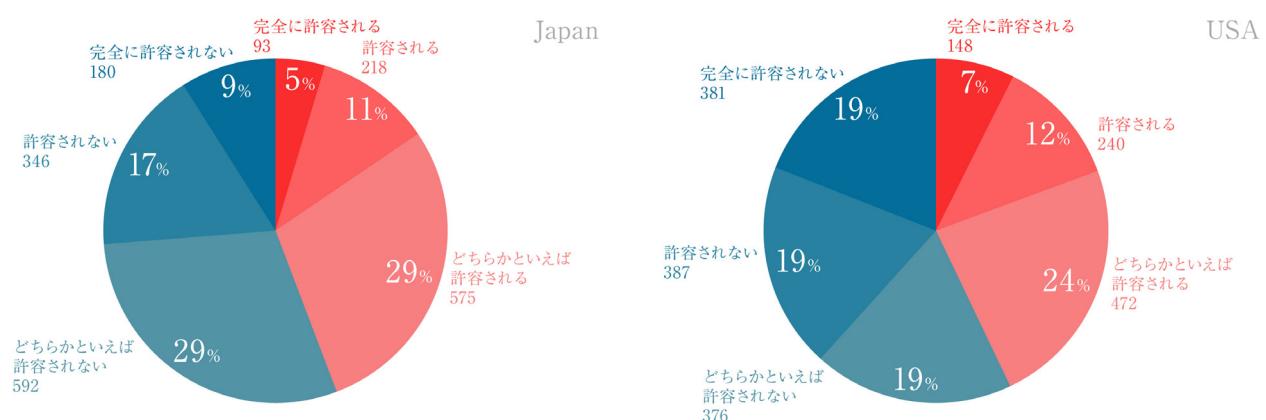
このエネルギー源が兵器として利用されるのを防ぐために、共同研究者を毒殺することは道徳的に許容されるでしょうか？

b. あなたと共同研究者は、安価で安全かつクリーンな強力な新しいエネルギー源を発見しました。この発見は、世界中の汚染や貧困を根絶する可能性を秘めています。しかし、あなたの共同研究者はこの発見を軍事目的で売りたいと考えています。彼は最初に手を挙げた最高額の買い手にこの発見を売るつもりであることを知っています。そして、その買い手の中には、この発見を恐ろしい兵器に変えようとする者が確実に含まれています。

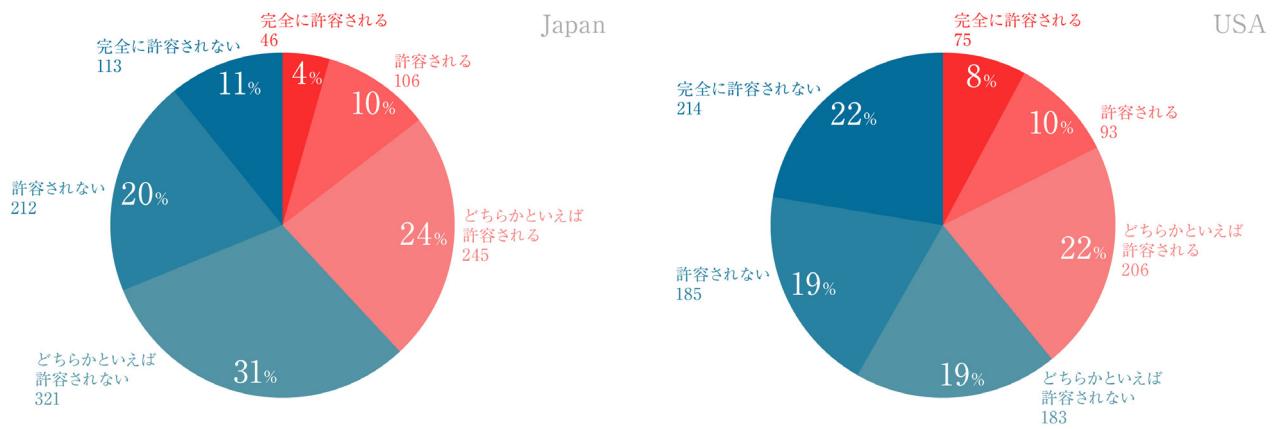
彼はその取引をメールで行おうとしており、これを防ぐ唯一の方法は、実験室内に可燃性ガスを放出することです。このガスにより、彼がコンピュータを起動した瞬間に爆発が起こります。皆は単なる実験室の事故だと思い込むでしょう。その結果、この発見が兵器として利用されることを防がれます。

この発見が兵器として利用されるのを防ぐために、可燃性ガスを放出して爆発を引き起こすことは道徳的に許容されるでしょうか？

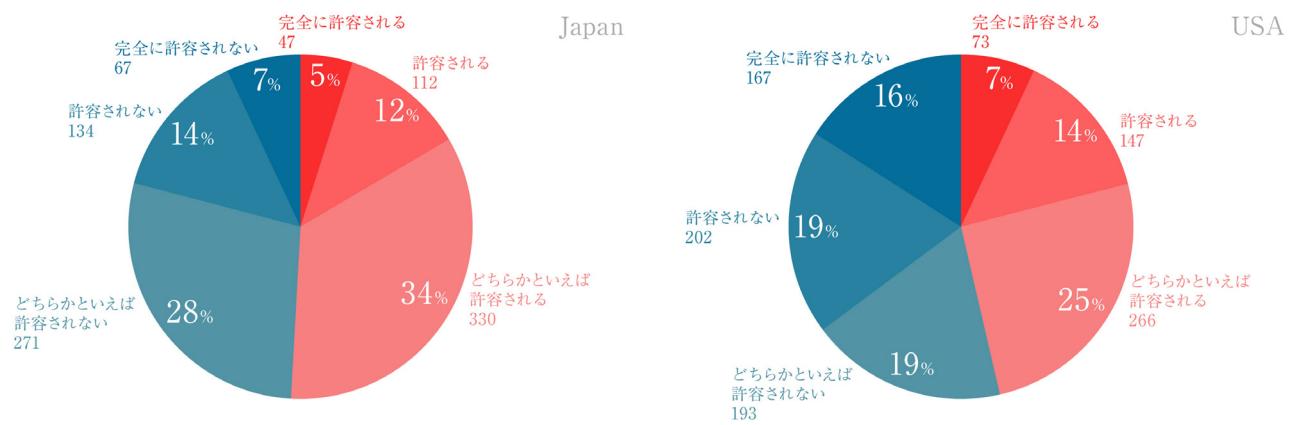
a, b合わせての回答



aの回答



bの回答



a、bともに、画期的なエネルギー技術が兵器としての悪用を、非常手段で阻止するか否かを尋ねている。aの事例では、共同研究者を毒によって密かに殺すことで兵器化を防ぐという選択が提示されている。bの事例では、可燃性ガスにより実験室を爆発させるという選択が提示されている。

aの事例では、日本62%、米国60%が許容できないと回答している。一方で、bは許容できないとする割合が低下しており、日本49%、米国54%となっている。実は、bの設問は、共同研究者の殺害という点が含まれず、犠牲が明示されていない。aとbについて、日本は13%、米国は6%の違いだが、犠牲の明示と黙示の差が結果を分けた可能性がある。

12. バイクウィーク (Bike Week)

a. あなたは熟練したバイクライダーで、休暇を利用してバイクウィークに参加しています。このイベントには、全国各地から何千人のライダーが集まっています。大勢のライダーを率いて道路を走っていると、前方で1人のライダーがバイクのコントロールを失いそうになっているのを目撃します。

あなたがスピードを上げてその不安定なライダーの横に並んだとき、その人が今にもクラッシュしそうであることに気付きます。このままでは、道路の中央でクラッシュが発生し、後続のライダーたちが衝突を避けようとしてお互いに押しつぶされ、多くの死者が出ることは確実です。

あなたは、そのライダーを物理的に道路の脇へ押しやり、木に衝突させることができると気付きます。この行為によってそのライダーはクラッシュし、現在のスピードではほぼ確実に死亡するでしょう。しかし、道路の中央でのクラッシュと、後続のライダーたちの大規模な多重衝突は防ぐことができます。

後続のライダーたちの命を救うために、そのライダーを木に衝突させて死なせることは道徳的に許容されるでしょうか？

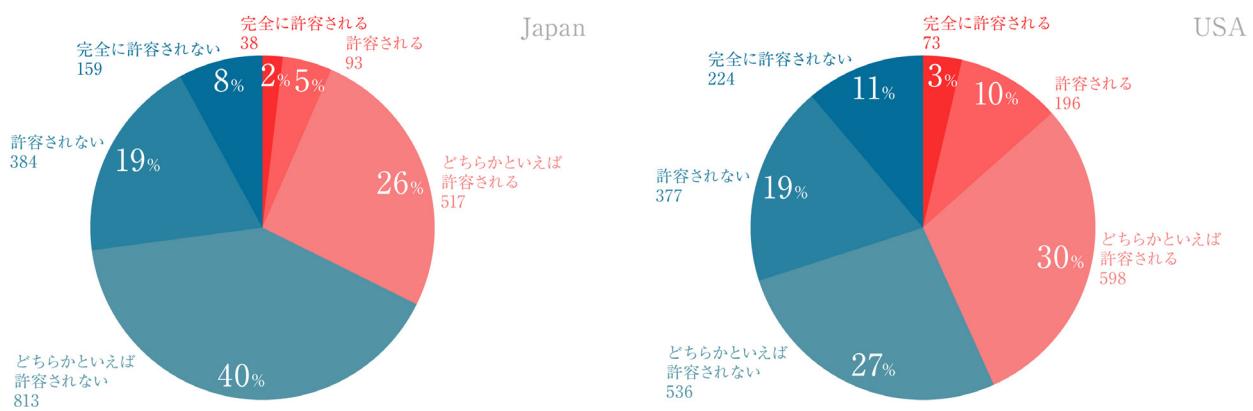
b. あなたは熟練したバイクライダーで、休暇を利用してバイクウィークに参加しています。このイベントには、全国各地から何千人のライダーが集まっています。大勢のライダーを率いて道路を走っていると、前方で1人のライダーがバイクのコントロールを失いそうになっているのを目撃します。

そのライダーが転倒するのを見て、あなたは簡単にその倒れたライダーを避けて走ることができると分かっています。しかし、そうすると後続のライダーたちは衝突を避けようとしてお互いに押しつぶされ、多くの死者が出る大規模な多重衝突が発生するでしょう。

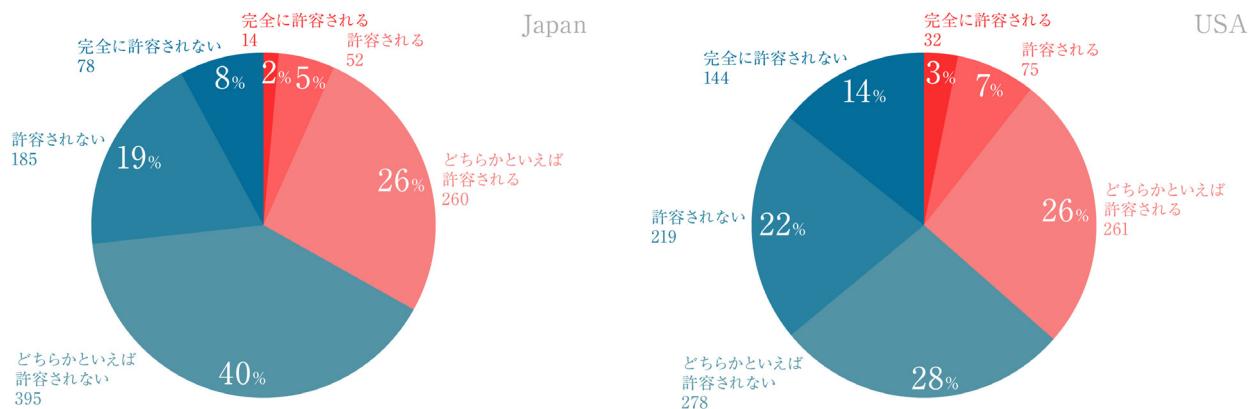
あなたは熟練者として、自分のバイクを横倒しにして転倒し、自分自身はバイクから飛び降りることができます。この方法でバイクを倒れたライダーに滑り込ませることで、後続のライダーに事故の警告を与え、大規模な多重衝突を防ぐことができます。しかし、この行為により倒れたライダーは確実に死亡するでしょう。

後続のライダーたちの命を救うために、自分のバイクを倒してそのライダーを死なせることは道徳的に許容されるでしょうか？

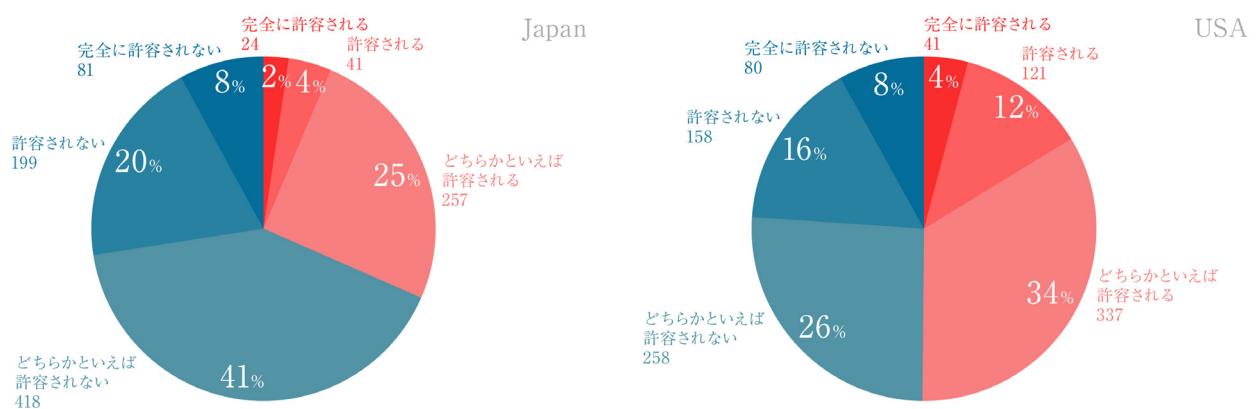
a, b合わせての回答



aの回答



bの回答



a、bともに、バイクライダーを犠牲にすることで、他のライダーを救うか否かを尋ねています。aの事例では、クラッシュしそうなライダーを木に衝突させて犠牲にすることで多数の死者を防ぐ、bの事例では、倒れたライダーに自分のバイクを衝突させることで多数の死者を防ぐというものである。aの場合は、犠牲になるライダーはまだ転倒しておらず、bの場合は、犠牲になるライダーは転倒している、という違いがある。

調査結果だが、aについては日米ともに許容されないが多数派で、日本67%、米国64%であった。一方で、bについては、日本は69%が許容されないと回答する一方で、米国は50%強が許容できるという回答であった。

02.

道徳基盤理論 (Moral foundations theory) に基づく質問紙調査 (MFQ2)

本調査では、道徳基盤理論 (Moral Foundations Theory) に基づいた質問紙「MFQ-2 (Moral Foundations Questionnaire 2)」を用い、日本とアメリカの回答者の道徳的傾向を比較分析した。MFQ-2は、道徳的判断の背後にある価値観を6つの基盤に分類し、それぞれに対する重視度を測定した。なお、質問全文は、Appendixにまとめてある。

分析対象とした6つの道徳基盤は以下の通りである：

Care (ケア) : 他者の苦しみに対する思いやりや共感

Equality (平等) : すべての人に同等の条件を与えるべきという価値観

Proportionality (公平) : 努力や功績に応じた対価や評価の正当性

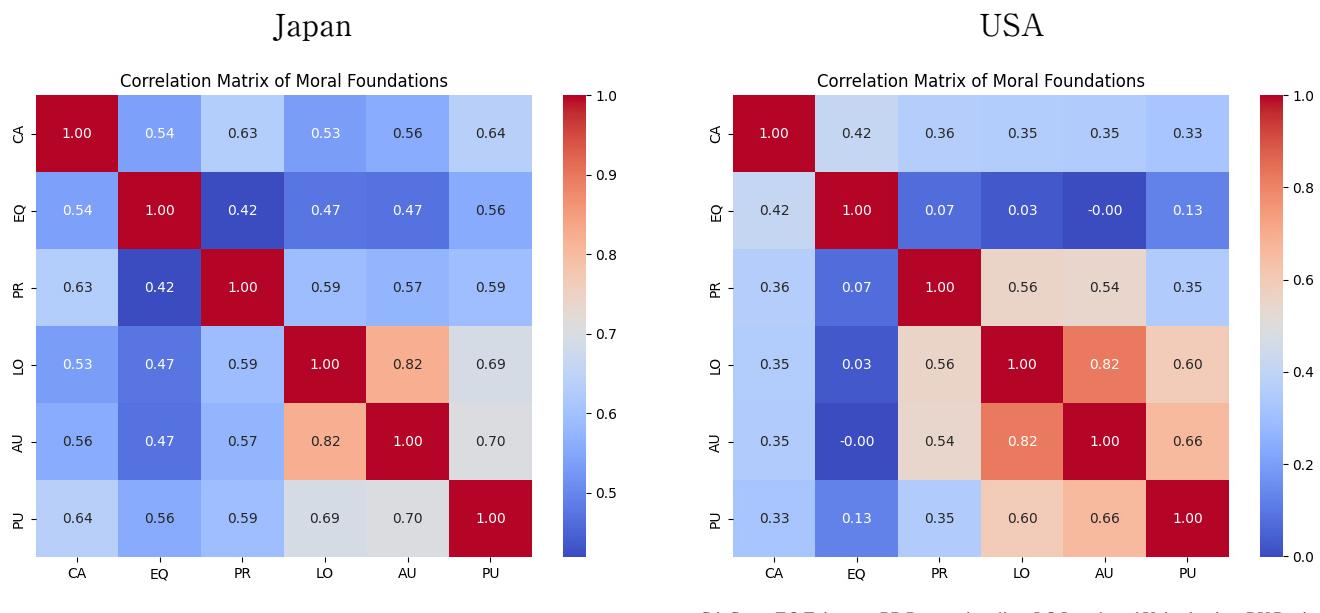
Loyalty (忠誠) : 自国や所属集団への団結や帰属意識

Authority (権威) : 社会秩序や伝統、上位者への尊重

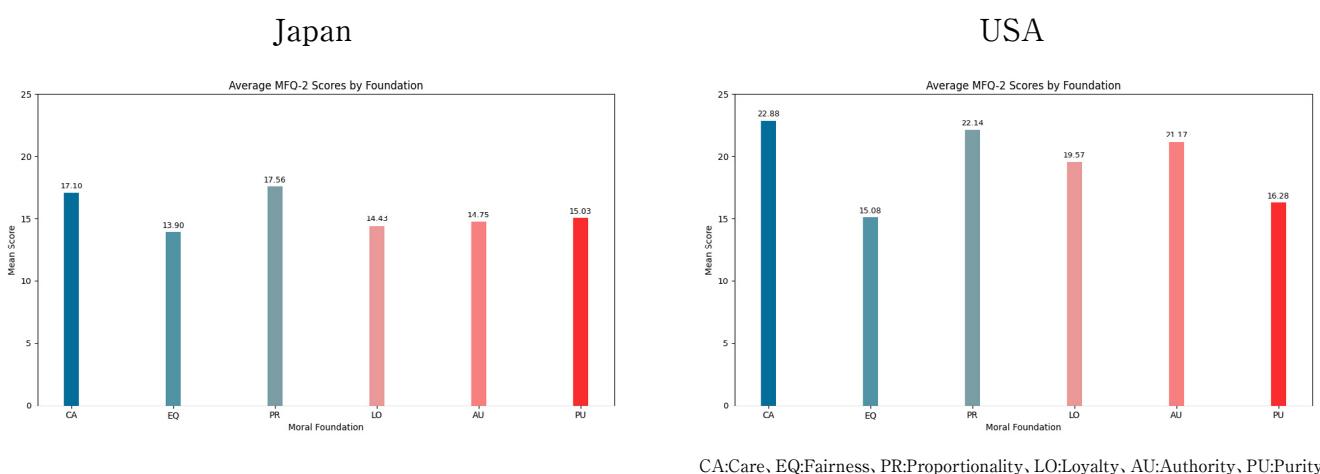
Purity (純粹性) : 身体的・精神的な清らかさや神聖さの重視

本調査からは、日本人は道徳的価値観の自己主張が全体的に控えめである傾向が読み取れる。特に「純粹性」「忠誠」「権威」のスコアが相対的に低く、個人よりも集団に価値を置く場面において、アメリカとの違いが顕著に現れる。一方で、米国では個人の価値観がより明確にスコアとして表れるため、特定の道徳観が強く表出しやすい。

MFQ-2スコアの相関マトリックス表



MFQ-2スコアのローデータの平均



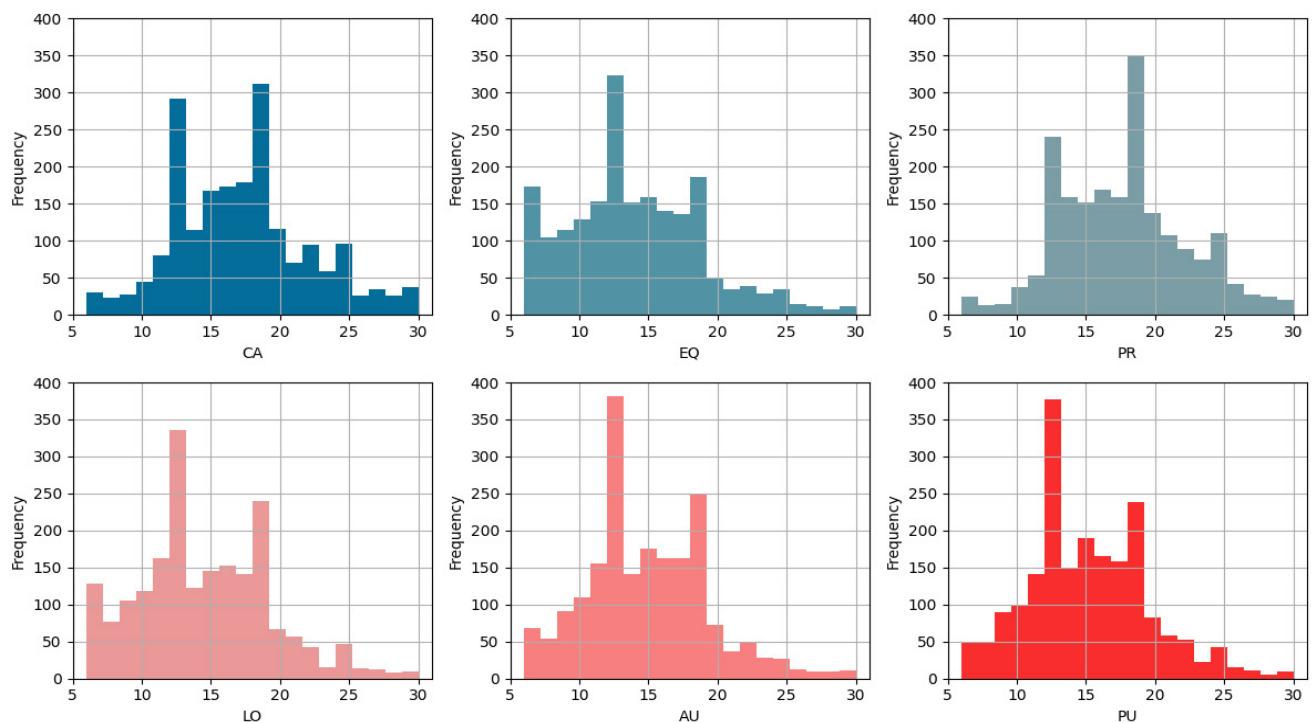
日本は、平均的に低く、米国は平均的に高い。これは、質問に対する点のつけ方の傾向の違いであり、日米に大きな文化差があることが推察できる。

次頁のヒストグラムにみられるように、日本はスコアが低めに分布しており、控えめな傾向が見られる。特に「純粹性」「忠誠」「権威」といった集団規範に関わる項目では、得点のばらつきが少なく、中央値も低い。これに対しアメリカでは、スコアの分布が広く、極端なスコア（高得点）を示す回答者が多い。これは、価値観を積極的に表明する文化的傾向や個人主義的な社会背景が関与していると考えられる。

MFQ-2スコアのヒストグラム

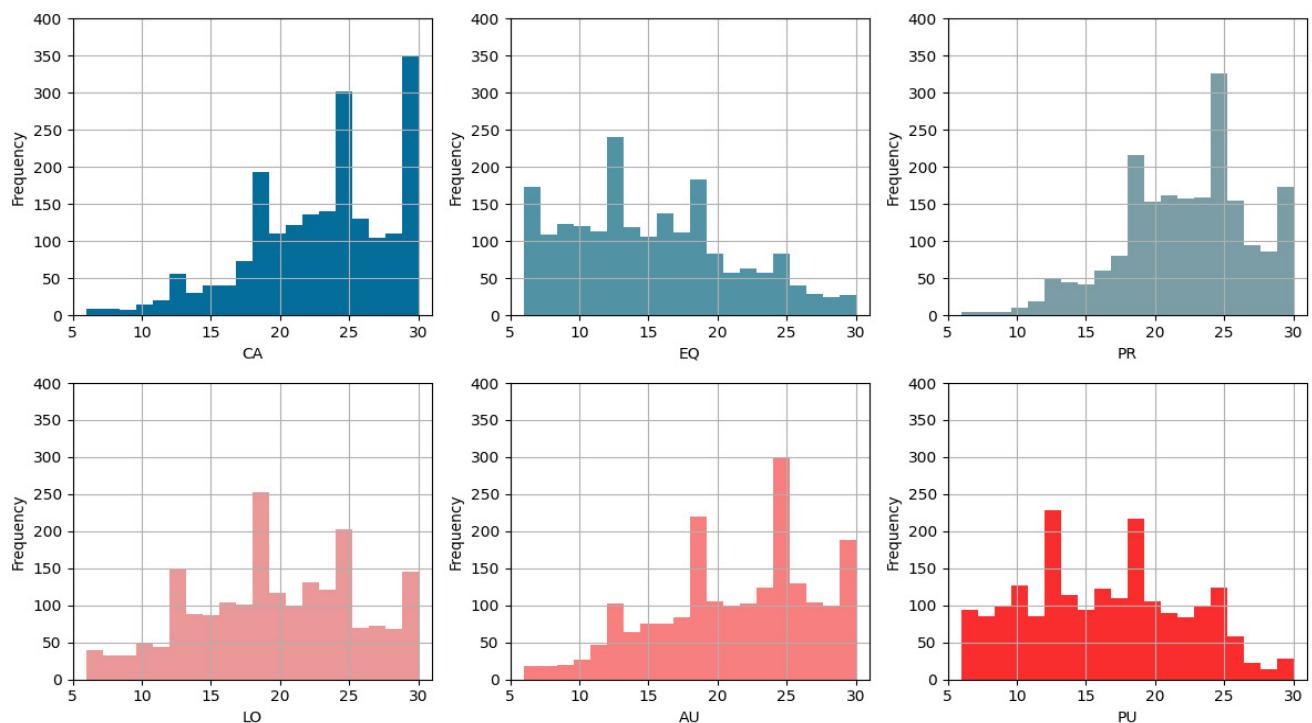
Japan

Histogram of MFQ-2 Scores



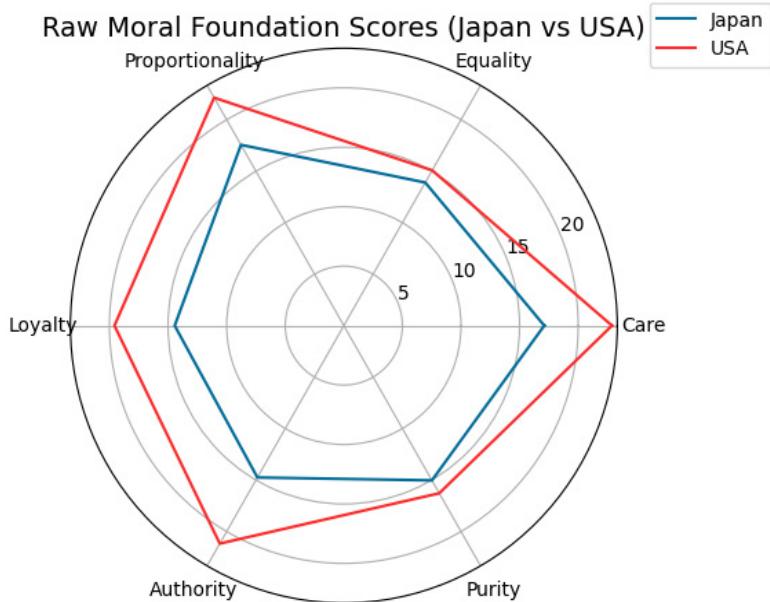
USA

Histogram of MFQ-2 Scores



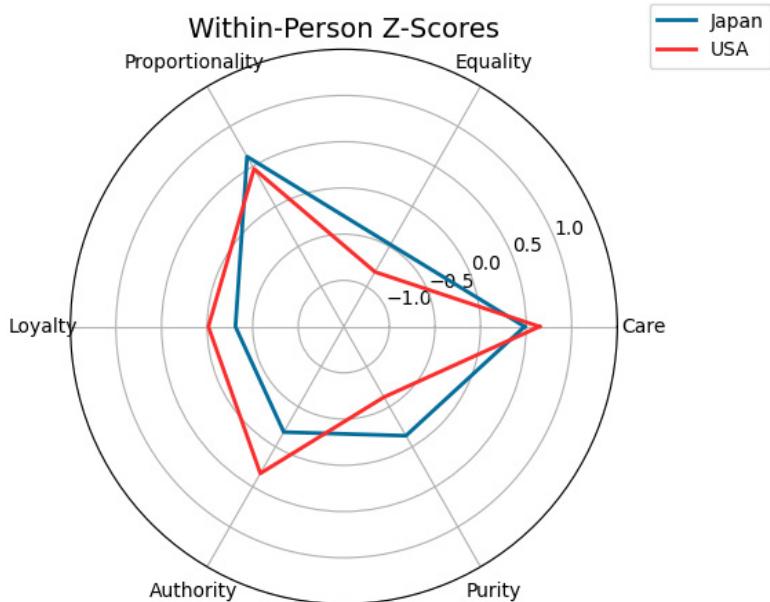
CA:Care, EQ:Fairness, PR:Proportionality, LO:Loyalty, AU:Authority, PU:Purity

MFQ-2スコアのレーダーチャート



左図は、各人の回答（生スコア）を、各国ごと平均してプロットしたものである。傾向として、日本の点付けは控えめで、米国は大きめである。

左図のような、単純な絶対値の比較では、各国の傾向の違いが見えにくくなる。



左図は、Zスコアを平均してプロットしたものである。各人の6項目のスコアの平均と標準偏差を算出し、6項目のスコアをZスコア（ $Z = \frac{\text{スコア}-\text{平均}}{\text{標準偏差}}$ ）

に変換し、さらに各人の結果を国ごとに平均したものである。このZスコアのチャートは、「その人がどの価値観を相対的に重視しているか」を示している。

日本の特徴は、Proportionality（公平）やCare（ケア）が相対的に高く評価されており、Loyalty（忠誠）・Authority（権威）・Purity（純粹性）といった集団志向の基盤は控えめである。アメリカの特徴はCareに加えLoyaltyやAuthorityのスコアも比較的高く、集団的価値観と個人主義的価値観が混在していることがうかがえる。一方で、Equality（平等）は両国ともにZスコアで見ると相対的に低く、個人内であまり重視されていない可能性がある。

相関分析

本章では、以下の3つの分析を行っている。

1. MFQ-2スコアと全ディレンマの相関マトリックス

MFQ-2 (Moral Foundations Questionnaire 2) の6つの道徳的価値観スコア (CA, EQ, PR, LO, AU, PU) と、12個の倫理的ジレンマ設問 (a/b分岐含む) の数値回答を用い、ケンドールの順位相関係数により相関を算出した。これは、両変数を順序尺度として扱い、回答傾向と道徳的価値観の関連性を可視化するものである。ケンドール相関では一般に相関係数は0.2～0.3でも意味のある傾向とされ、参考として0.3を弱い相関、0.5を中程度、0.7以上を強い相関の目安とした。

2. MFQ-2スコアと二値化した全ディレンマの相関マトリックス

MFQ-2のスコアは連続変数のまま使用し、全ディレンマの回答を二値化した上で、ピアソン法で相関関係を分析した。ディレンマへの回答は6段階評価で構成されており、本分析では「1. 完全に許容される」～「3. どちらかといえば許容される」を「受け入れられる (1)」、それ以外を「受け入れられない (0)」と分類して二値化した。得られた二値変数とMFQ-2スコアとの相関係数を計算し、「受け入れられる」判断と各道徳的価値観との関連性を検討した。

3. MFQ-2スコアを説明変数とした全ディレンマの予測モデル

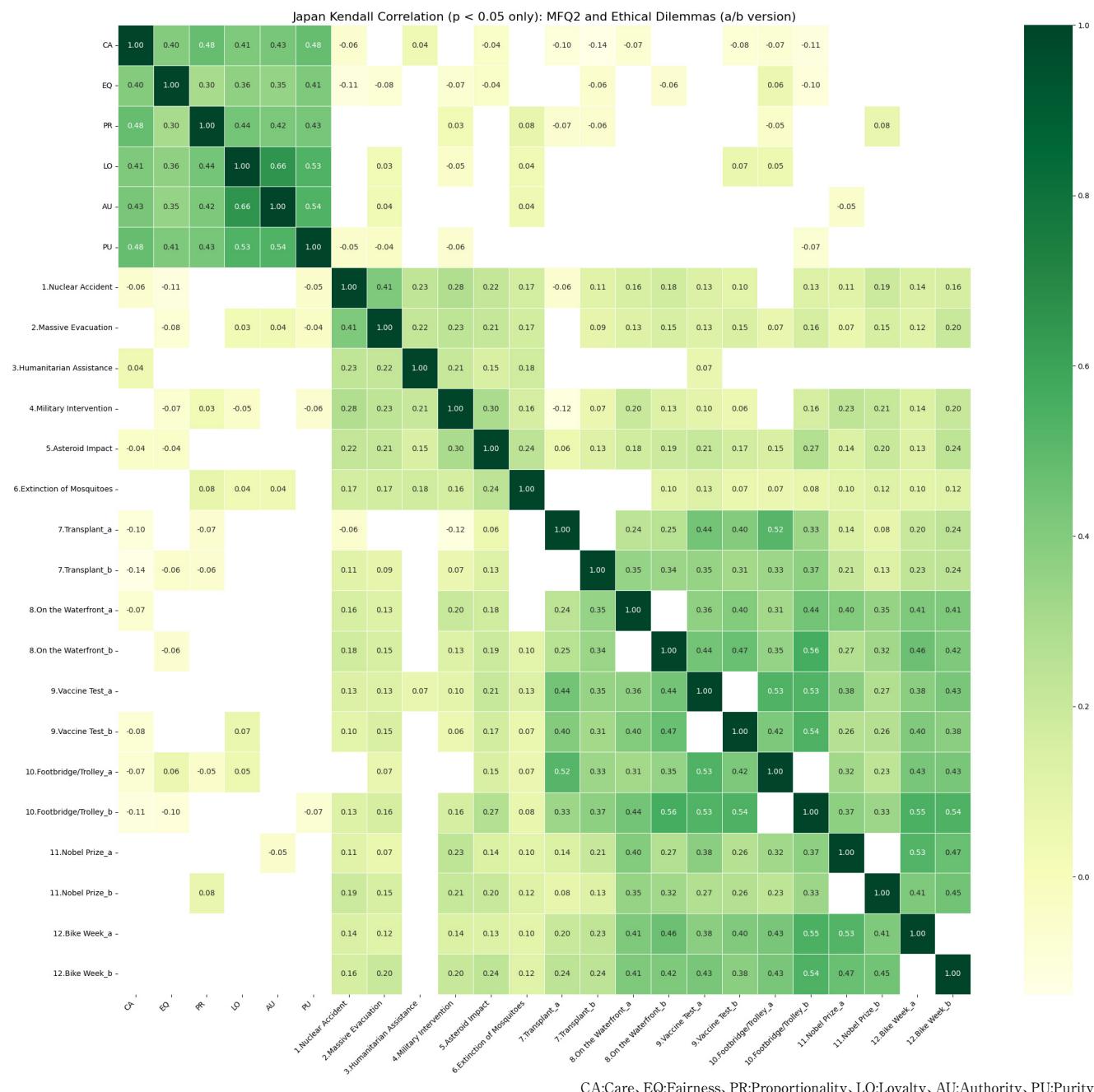
MFQ-2のスコアを特徴量 (説明変数) として用いて、全ディレンマにおいてどのような選択をするかを予測するモデルを構築した。なお、全サンプルの7割をトレーニングに、3割をテストに使用している。

予測モデルの構築は、二つに分けて実施した。第一に、すべての回答者を対象としたモデルを作成し、価値観の傾向がディレンマへの判断にどの程度影響を与えるかを全体的に評価した。第二に、MFQ-2スコアのうち、「イングループ志向 (忠誠・権威・純粹性)」と「アウトグループ志向 (ケア・平等・公平)」の差分 (Zスコアの差) が大きい上位3分の1の層に絞って、価値観の差が明確な参加者の判断傾向に着目したモデルを別途構築した。すなわち、第二のケースは、価値観が明確な人々を対象とした。

用いた予測モデルは、ロジスティック回帰分析である。モデルの性能評価には、AUC (Area Under the Curve) を指標として採用し、AUC値が0.5からの離隔で、MFQ-2スコアとの相関または逆相関を示す尺度として解釈した。

MFQ-2スコアと全ディレンマの相関マトリックス

Japan



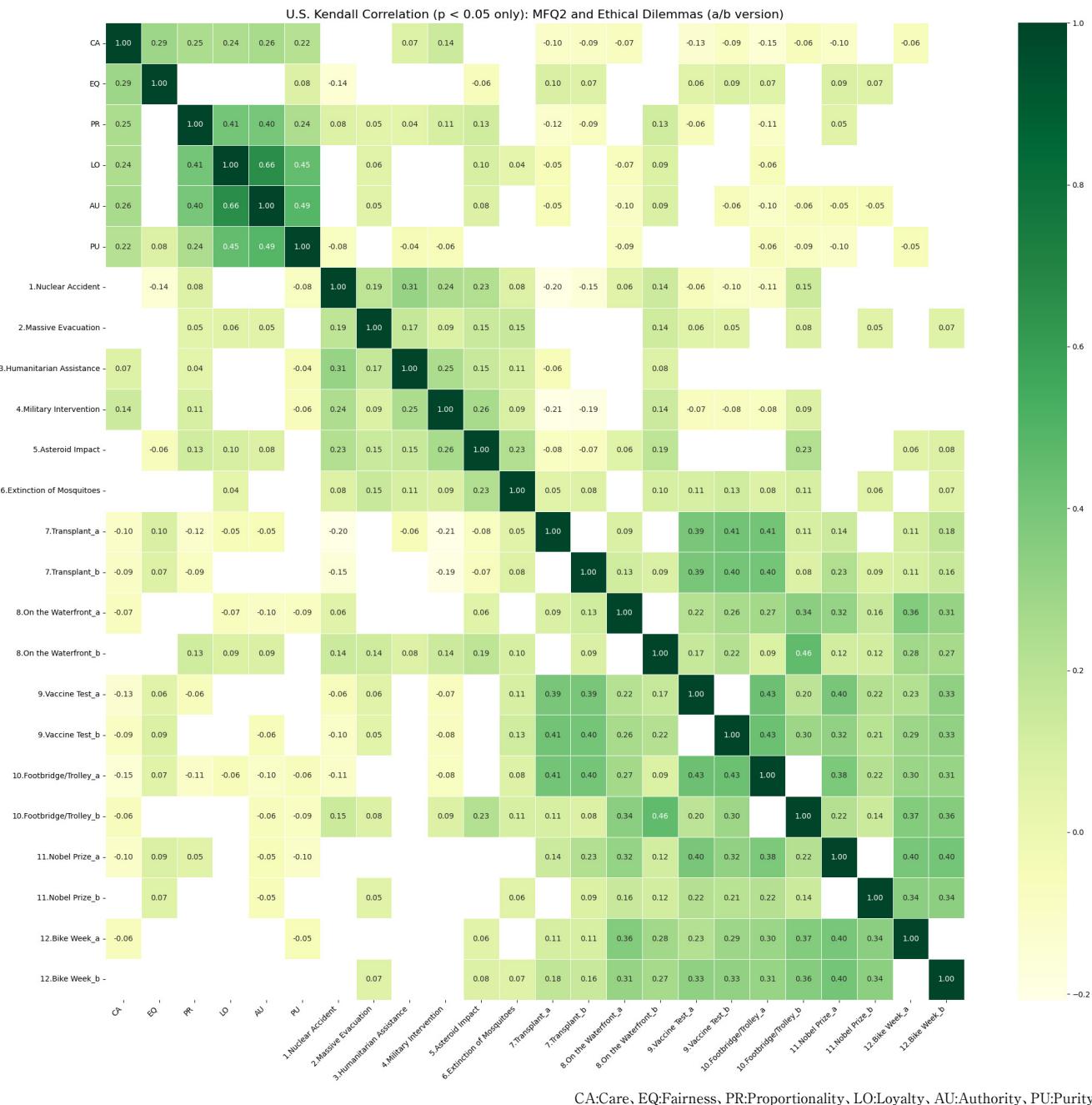
日本 : MFQ2と全ディレンマの相関分析 (ケンドール法)

※統計的に有意 ($p < 0.05$) の相関係数のみを表示

MFQ-2の各スコアと各ディレンマの相関関係は、全体的に低い。サンプル全体で見たとき、価値観と各ディレンマの関係を見出すことはできなかった。

MFQ-2の各スコア同士、《究極の選択》の回答同士、モラル・ディレンマの回答同士の相関関係は、弱から中程度の相関関係が出ている。《究極の選択》では、1.原発事故と2.大規模避難に弱い相関がみられる。また、10.ファットマン／トロリー問題と他のモラル・ディレンマの回答とに中程度の相関関係がみられる。

USA



米国 : MFQ2と全ディレンマの相関分析 (ケンドール法)

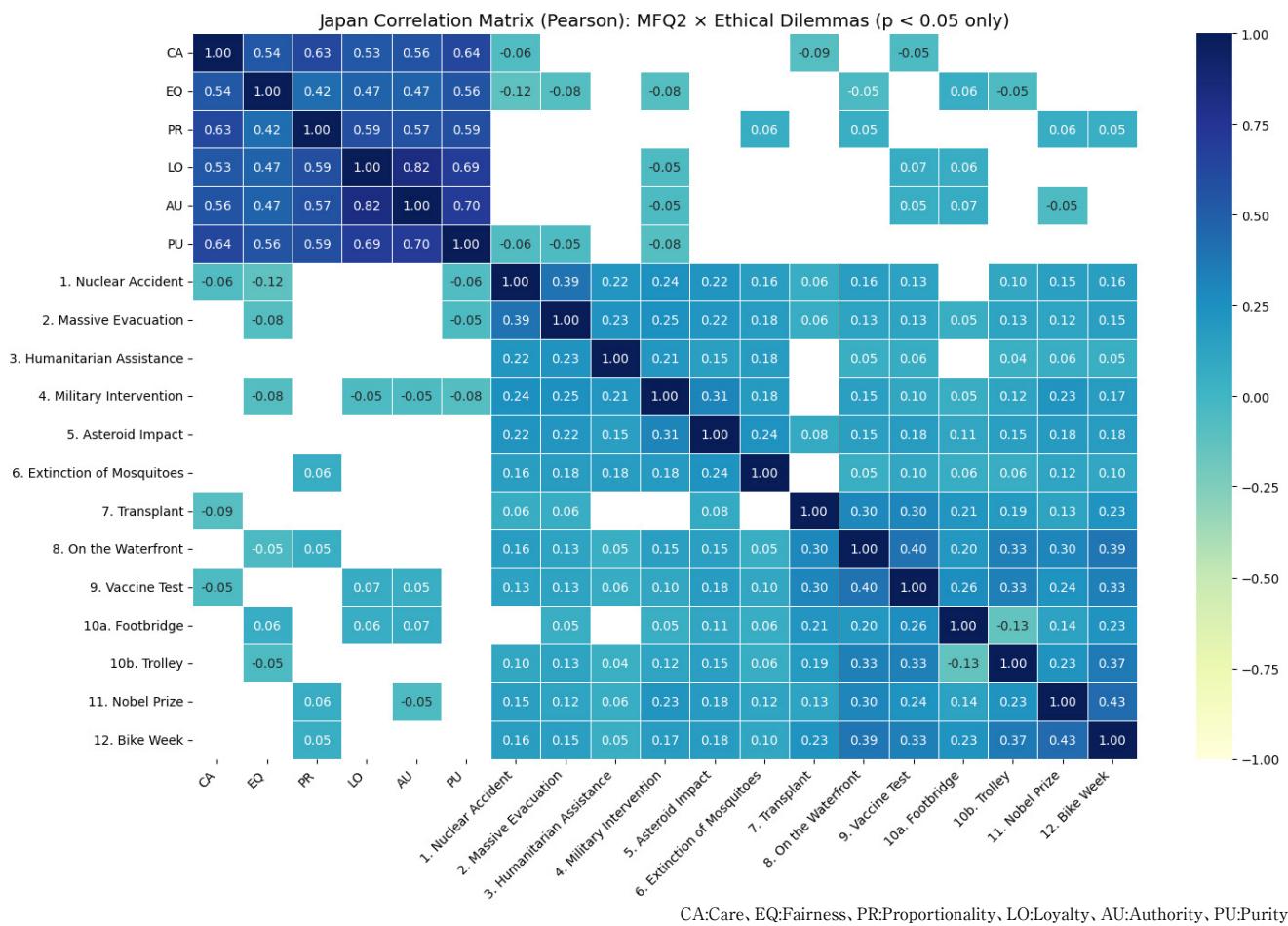
※統計的に有意 ($p < 0.05$) の相関係数のみを表示

日本と同じくMFQ-2の各スコアと各ディレンマの相関関係は、全体的に低い。サンプル全体で見たとき、価値観と各ディレンマの関係を見出すことはできなかった。

日本の場合と同じく、MFQ-2の各スコア同士、《究極の選択》の回答同士、モラル・ディレンマの回答同士の相関関係は、弱い相関関係がみられる。《究極の選択》では、1.原発事故と3.人道支援に弱い相関がみられる。また、モラル・ディレンマの回答同士に、弱から中程度の相関関係がみられる。

MFQ-2スコアと二値化した全ディレンマの相関マトリックス

Japan



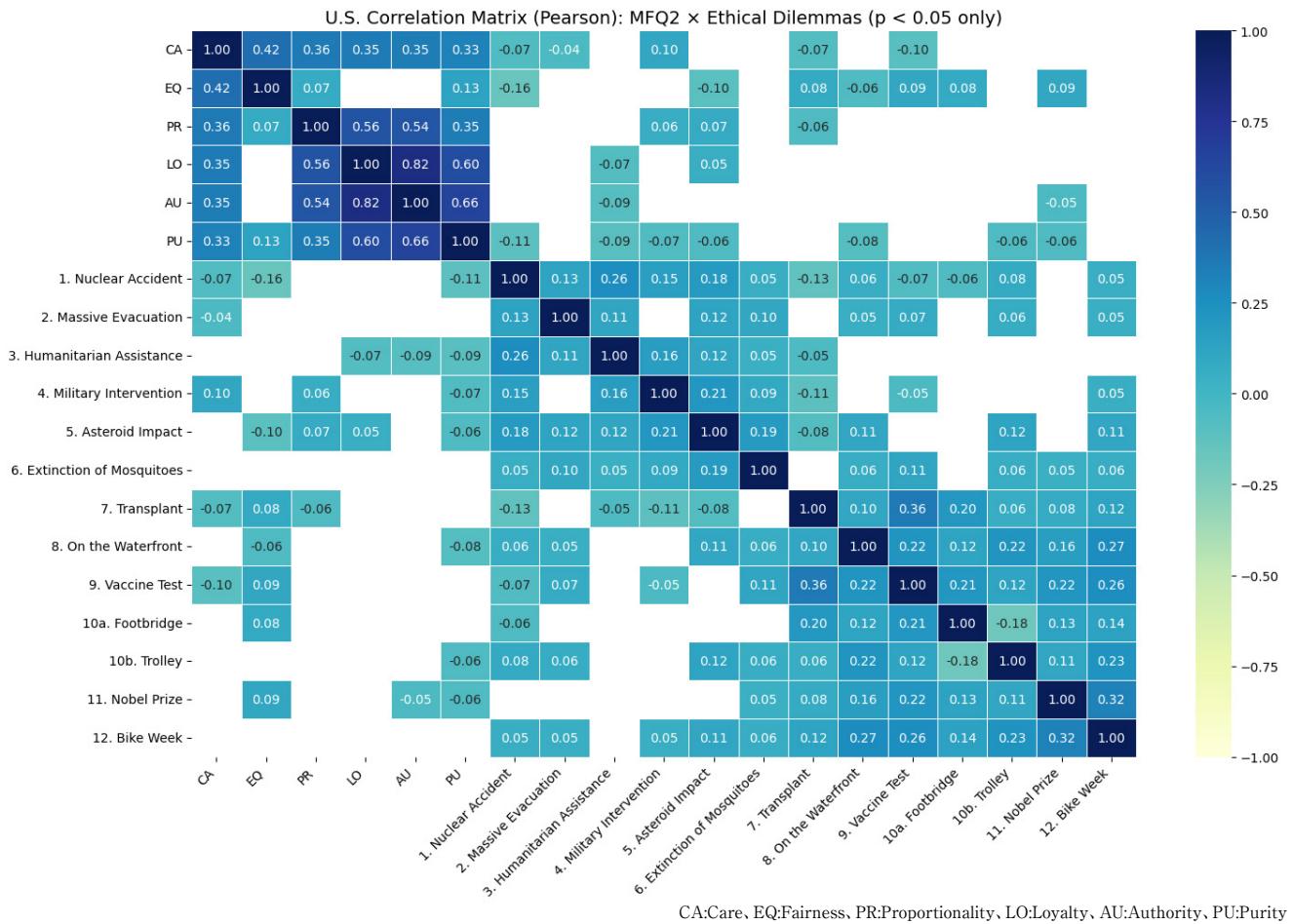
日本 : MFQ2と二値化した全ディレンマの相関分析 (ピアソン法)

※統計的に有意 ($p < 0.05$) の相関係数のみを表示

二値化した場合でも、MFQ-2のスコアと各ディレンマの相関関係は、全体的に低い。ケンドール法での分析と同じく、サンプル全体で見たとき、価値観と各ディレンマの関係を見出すことはできなかった。

ケンドール法の場合と同じく、MFQ-2の各スコア同士、《究極の選択》の回答同士、モラル・ディレンマの回答同士の相関関係は、相関関係がみられる。

USA



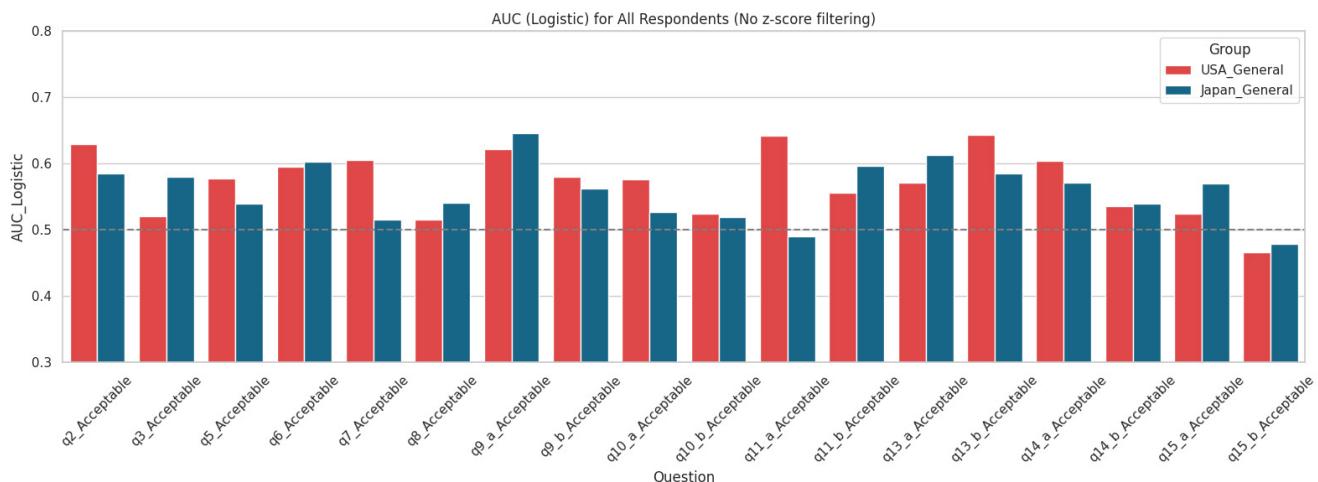
米国 : MFQ2と二値化した全ディレンマの相関分析 (ピアソン法)

※統計的に有意 ($p < 0.05$) の相関係数のみを表示

二値化した場合でも、MFQ-2のスコアと各ディレンマの相関関係は、全体的に低い。ケンドール法での分析と同じく、サンプル全体で見たとき、価値観と各ディレンマの関係を見出すことはできなかった。

ケンドール法の場合と同じく、MFQ-2の各スコア同士、《究極の選択》の回答同士、モラル・ディレンマの回答同士の相関関係は、相関関係がみられる。ただし、日本と比較した時、各問題群同士の相関関係が若干低い数値となっている。

MFQ-2スコアを説明変数とした全ディレンマの予測モデル

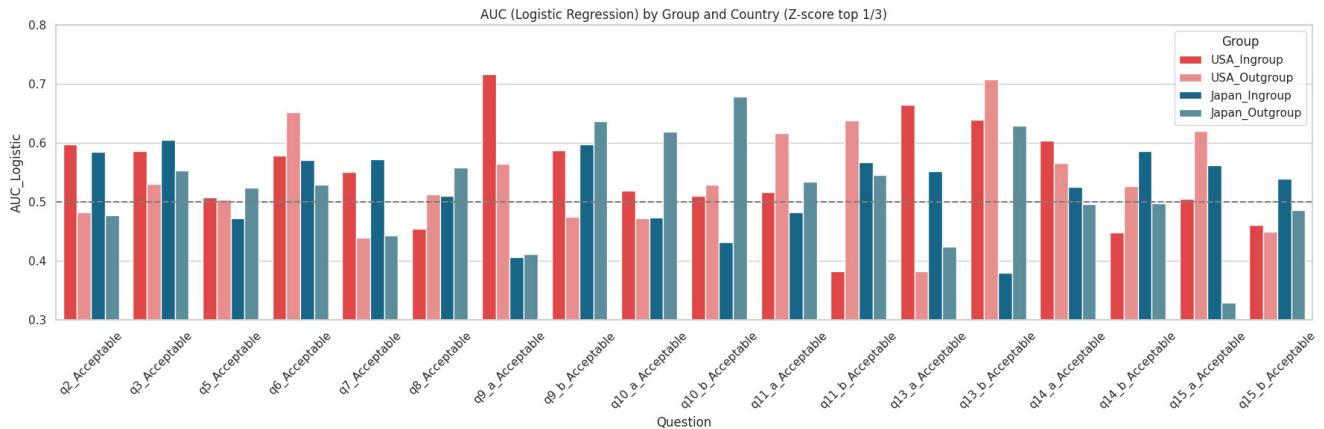


全サンプルを対象として、日米それぞれに予測モデルを構築した。上記は、その結果を比較したものである。図中では、米国の結果を赤、日本の結果を青で表示している。

予測精度を直感的に理解するために、コイン投げの例を挙げると、AUC値が0.5である場合は「裏か表か分からない」という完全なランダム状態を意味する。この値から乖離すればするほど、判断傾向に偏りがある、すなわち価値観から選択をある程度予測できることを示している。

AUC値が高かった設問の上位は、米国の「7. 移植」(0.660)、「9. ワクチンテスト」(0.646)、「10b. トロリー問題」(0.644)、「1. 原発事故」(0.629)、および日本の「10a. ファットマン問題」(0.613)であった。

概して、米国の方が日本よりも予測精度は高い傾向にあるが、最も高いAUCでも0.66にとどまっており、予測モデルとしては実用的な精度には至っていない。これは、価値観による判断傾向に一定のパターンは見られるものの、それが完全に一貫しているわけではないことを示唆している。



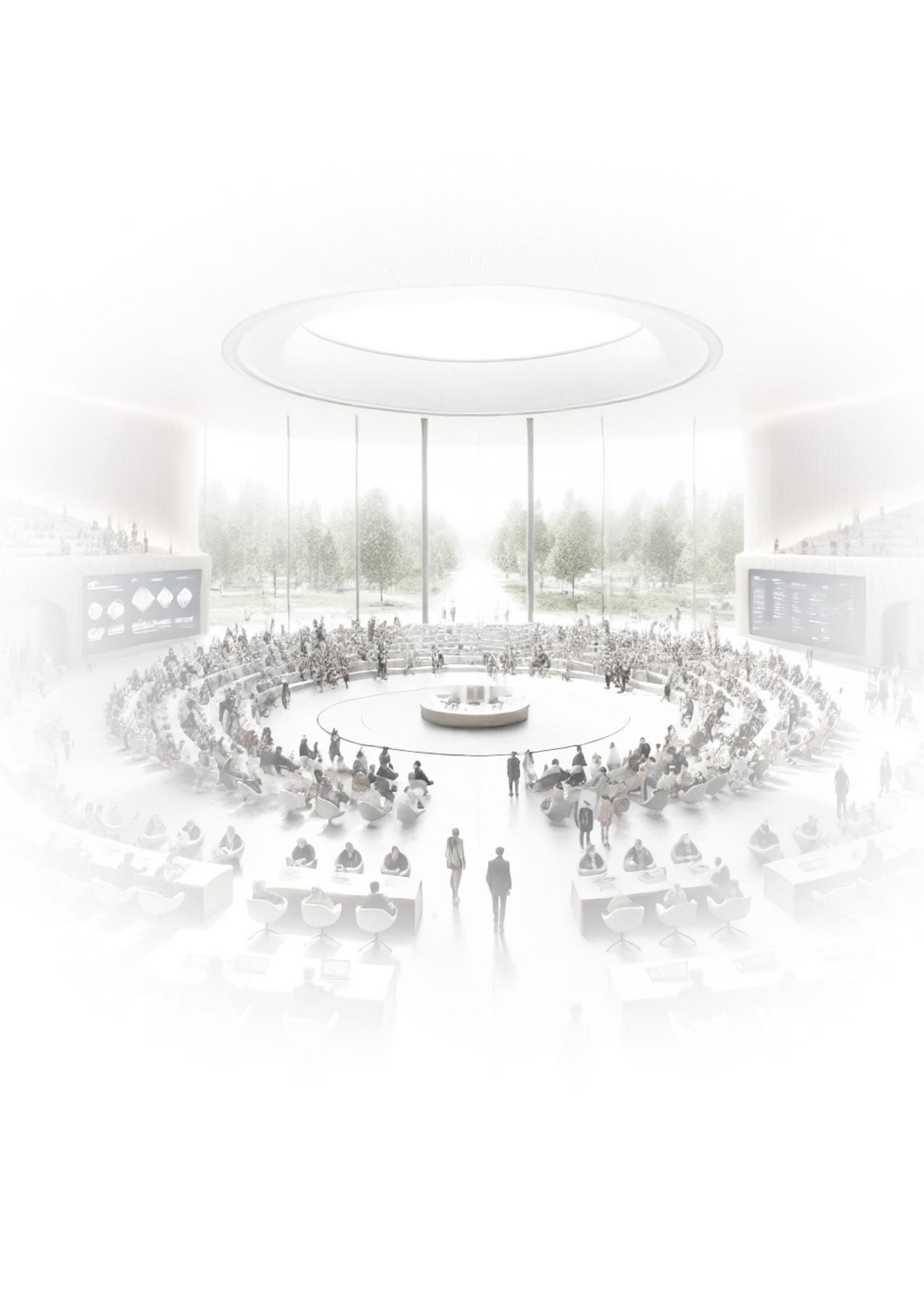
Zスコア上位3分の1の回答者、すなわち価値観に明確な傾きがある層に対して予測モデルを構築した結果が上図である。米国イングループは濃赤、米国アウトグループは薄赤、日本イングループは濃青、日本アウトグループは薄青で表示している。

最も高いAUCを示したのは、米国アウトグループにおける「10b. トロリー問題」(AUC = 0.708)であり、次いで「10a. フットマン問題」(米国イングループ、0.665)、「7. 移植」(米国アウトグループ、0.663)、「4. 軍事介入」(米国アウトグループ、0.652)、「9. ワクチンテスト」(米国イングループ、0.647)と続く。これらはいずれも AUCが0.64以上であり、価値観の傾向が判断に一定の予測力を持っていることを示す。

一方、日本においては、最も高い AUCは「7. 移植」(イングループ、AUC = 0.630) および「10b. トロリー問題」(アウトグループ、AUC = 0.630) であった。米国に比べるとAUCの最大値はやや低く、また同じ設問でも逆相関の強さが見られた点が特徴的である。たとえば、「7.移植」は日本アウトグループではAUC = 0.370 (逆相関、 $\Delta = 0.130$) を示しており、価値観が強いほど逆の判断をする傾向が見られることがわかる。

さらに、「10b.トロリー問題」でも、米国ではイングループ・アウトグループともにAUCが高かった一方で、日本イングループではAUC = 0.380 (逆相関) となっており、同じ設問に対する国ごとの反応の違いが際立つ結果となった。

このように、Zスコアによって価値観の傾向が強い層を抽出することで、判断における価値観の影響力をより鮮明に捉えることが可能である。とりわけ米国においてはその傾向が顕著であり、日本はそれに比べて予測の一貫性がやや弱いことが観察された。



本調査の総括

《究極の選択》とモラル・ディレンマの日米比較

日本は、他者を犠牲にする判断に慎重で、行為の意図や手段への倫理的配慮を重視する。

米国は、結果として多くを救うために少数を犠牲にする判断を比較的許容する傾向がある。

こうした文化差は、AIが倫理的判断を支援・代行する際、どの文化圏の判断パターンに基づいて構築されるかによって、出力が大きく変わる可能性を示している。

道徳的価値観 (MFQ-2) の日米比較

日本では「公平」や「ケア」が相対的に重視されるが、全体的に控えめな傾向があり、集団規範（忠誠・権威・純粹性）への関心は低い。

米国では「忠誠」「権威」「ケア」など、個人主義と集団価値が並立する形で高スコアが示されている。

AIが倫理判断を行う際、こうした文化ごとの価値の偏りを無視したモデルは、他文化圏に適用したときに倫理的逸脱を引き起こす可能性がある。

道徳的価値観 (MFQ-2) とディレンマの相関分析

道徳的価値観 (MFQ-2) と具体的なディレンマとのあいだには、明確で強い相関は見られなかった。

判断は単一の価値観では決まらず、文脈や関係性など複雑な要因の重なり合いによって形成されている、とみることができる。

これは、AIによる意志決定支援の基盤となる予測モデルが、「特定の価値観に基づく単純な判断パターン」に依存できないことを意味する。

ただし、Zスコア上位者の場合は、価値観と判断の相関関係が強く出る場合がある。

価値観と判断の結びつきについて

本調査の目的である価値観と判断の結びつきの検証だが、分析結果によれば、価値観に基づいて人々の判断を予測することは、全体としては困難である。一方で、価値観が極端に偏った一部の人々においては、その判断をある程度（高くて7割程度）予測することが可能である。

すなわち、一般的な人々に対しては価値観と判断との間に明確な対応関係は確認できないが、イングループ志向やアウトグループ志向など、特定の道徳的価値観において極端なスコアを持つ人々は、その価値観と判断が強く結びついている可能性が高い。このような傾向は、価値観に基づく判断予測の「的中率」が、極端な人ほど高くなるという形で表れている。

本調査の射程から外れるが、この結果は、価値観に根ざした判断が社会においてどのように現れるかを考察する上で重要であると同時に、現代社会における情報操作のリスクとも関連している。

【事例】ケンブリッジ・アナリティカ事件

2018年に世界的に問題視された「ケンブリッジ・アナリティカ事件」は、その典型的な事例である。この事件では、Facebookを通じて不正に収集された個人データをもとに、ユーザーの性格傾向や価値観を推定し、特に極端な傾向を持つ層に向けて政治的に偏った広告を配信することで、選挙結果に影響を与えようとしたとされている。このような手法では、予測可能性の高い対象者に絞って働きかけることで、少数派の行動変容を通じて、全体の社会的選択を左右することが可能になる。実際に、アメリカ大統領選挙やイギリスのEU離脱（ブレグジット）において、こうした手法の影響が指摘されている。

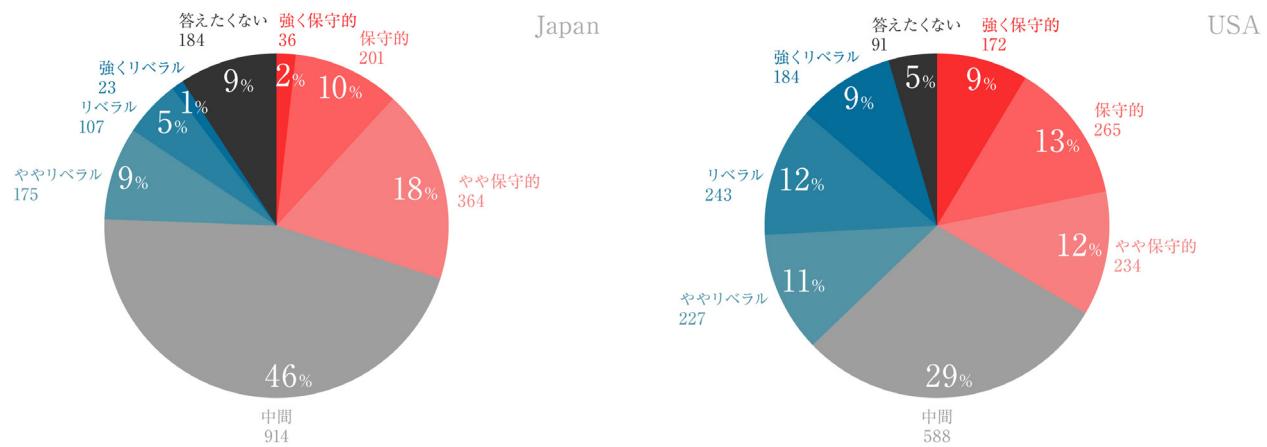
本研究の結果は、価値観と判断の結びつきが一般的には限定的であるという安心感を与える一方で、一部の人々は「予測されやすい存在」になっている可能性を示唆している。特に、判断が社会的に分断されるような論争的なテーマにおいては、こうした「予測可能な層」を標的とすることで、意図的な情報操作が行われるリスクがある。

Appendix.

1. 政治的思想傾向の調査

今回の分析では使用していないが、本調査において、下記のように、政治的思想の傾向を調査している。

あなたの政治的立場をお聞きします。次の中から、あなた最も近いと思うものを選択肢の中から選んでください。強く保守的・保守的・やや保守的・中間・ややリベラル・リベラル・強くリベラル・答えたくない



日本は、中間 (46%) を選んだ層が最も多い、続いて保守 (計30%) 、リベラル (計15%) 、答えたくない (9%) の順である。

米国は、保守 (34%) 、リベラル (33%) 、中間 (29%) 、答えたくない (5%) の順である。

日本と米国の大違いは、中間を選んだ回答者が日本の方が17%も高く、またリベラルの回答者が日本の方が米国よりも18%も低いという特徴である。

なお、保守とリベラルについて日本と米国では意味合いが異なる可能性があるが、日本が中道が多数派であること、米国は保守、リベラル、中道が伯仲していることを概況できる。

2. MFQ-2 調査の質問内容

本調査で使用されたMFQ-2の設問は、下記の1~36項目で構成されており、先述の6要素に分類されている。回答者はそれぞれの項目に対し、「まったくあてはまらない」から「非常によくあてはまる」までの5段階で評価した。

1. つらい経験をした人々への思いやりは重要な美德だ。
2. 誰もが同じ額のお金を稼げば、世界はもっと良くなると思う。

3. よく努力している人はより多くのお金を手にするべきだと思う。
4. 子供たちには自分の国への忠誠心を教えるべきだと思う。
5. 伝統的な価値観を大事にすることは社会にとって重要だと思う。
6. 人間の体は、神聖なものが宿る神殿のように扱われるべきだと思う。
7. 苦しんでいる人を思いやる気持ちは最も大切な美德のひとつだと思う。
8. 人々の収入が同じであれば、社会の問題は減るだろう。
9. 人はどれだけ貢献したかに応じて報酬を得るべきだと思う。
10. 自分の国に忠誠心を示さない人には憤りを感じる。
11. ほとんどの伝統は、社会秩序の維持において大きな役割を果たしていると感じる。
12. 貞操観念は大切な美德だと思う。
13. 私たちは心の痛みを抱える人々を思いやるべきだ。
14. 人生において、誰もが同じ量のリソースを与えられるべきだと思う。
15. 労働者がどれだけ仕事に熱心に取り組んだかが、昇給額に反映されるべきだ。
16. 誰でも自分の属するコミュニティを愛するべきだ。
17. 親に従うことは大切な美德だと思う。
18. 当然のように汚い言葉を使っている人を見ると憤りを感じる。
19. つらい人生を送っている人たちに共感する。
20. 社会の誰もが同程度の財産を手にするのが理想的だと思う。
21. 誰かの実績が認められると嬉しくなる。
22. 求められれば、誰もが自分の国を守るべきだ。
23. 我々は皆、年長者から学ぶ必要がある。
24. もし知り合いに、害はないものの普通ではない性的嗜好があると知ったら、相手に不安を感じてしまう。
25. 誰もが、困難な状況にある人々を慰めようとすべきである。
26. 人々が力を合わせて共通の目標に取り組む場合、かけた労力が人によって違ったとしても、その報酬は平等に分配すべきだ。
27. 公正な社会では、人よりも努力する人には他の人よりも裕福になってほしい。
28. 自分のコミュニティの人が国際大会で勝利したら誇りに思うものだ。
29. 子供たちに教えるべき重要な価値観のひとつは、権威を尊重することだと思う。
30. 化学的に複製された人工の薬品ではなく自然の薬ができるだけ用いるべきだ。
31. 他人のニーズを無視している人を見ると心が痛む。
32. 自分の国で、一部の人が他の人よりはるかに多くの財産を持っていることに憤りを覚える。
33. 不正を働く人々が捕まって処罰を受けるのを見るとすっきりする。
34. スポーツチームの強さは、メンバー同士の忠誠心から育まれると思う。
35. 強いリーダーを持つことは社会のためになると思う。
36. 結婚するまで純潔を守る人を尊敬する。

AIに社会の声を届ける 犠牲の選択に関する社会調査の報告

《究極の選択》研究会 編

発行日 2025年3月31日

調査者 大庭弘継(代表)

一方井祐子

井出和希

笠木雅史

菊地乃依瑠

高木裕貴

玉澤春史

発行者 《究極の選択》研究会

E-mail: info@hardestchoice.org

<https://hardestchoice.org/>

組 版 株式会社シブヤテレビジョン

本報告書は、トヨタ財団研究助成「社会的意志決定を行うAIの要件—良質なデータセットと望ましいアウトプット」(研究代表者:大庭弘継、D19-ST-0019) の研究成果として刊行したものである。

