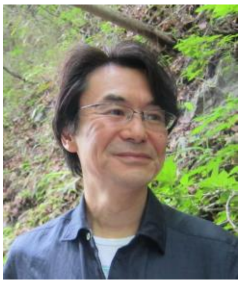
最終報告

分断を生み出すメカニズム解明と克服のための教育プログラム開発



研究代表者：**三輪和久**　名古屋大学情報学研究科　教授

共同研究者：**唐沢穣**　名古屋大学　教授

1. 研究の背景と達成目標

研究の背景

近年、世界各地で民主主義と専制主義の対立が激化し、新型コロナ感染症の拡大が引き金となって、国境を越えた政治的・経済的な緊張が高まっている。民主主義は、人権の尊重や多様な価値観の共存といった長期的なメリットをもたらす一方、時として社会の分断を深めるリスクを内包している。実際、トランプ政権期のアメリカやブレグジットの過程では、有権者の意見対立が先鋭化し、国家や地域が深刻に分裂する事態が生じた。

　こうした現象の背景には、政治制度やメディア環境だけでなく、人間の道徳心や正義感が衝突しやすい心理的メカニズムがあると指摘されている。つまり、人々が互いの「正しさ」をめぐって対立を強めるほど、民主主義の根幹である「参加と合意形成」のプロセスが機能不全に陥る恐れが高まる。本研究は、この「分断を生みやすい構造」を多角的に検証するとともに、長期的に安定した社会秩序を維持するための教育プログラムを構築し、有効となり得る示唆を提供することを目指すものである。

達成目標

本研究の最終目標は、人間の道徳心や正義感の衝突が民主主義社会における分断を生むメカニズムを解明し、実際の教育の場での介入を通じてその克服に取り組む具体的な方法論を提示することである。具体的には、道徳心理学や認知科学の理論を活用し、参加者の価値観を可視化するオンラインプラットフォームを開発・検証する。

　そこでは、対立が生じやすいテーマについて多様な立場を持つ人々が議論を重ね、相互の「正しさ」をメタ認知的に捉えられる仕組みを用意する。この体験を通じて、従来の道徳教育やリベラルアーツ教育における「善悪」や「規範」の学習を超え、多様な価値を相互補完的に理解する態度を育むことを目指す。加えて、この手法を大学や社会人研修に広く導入し、民主主義を成長させるうえで不可欠な寛容性や合意形成力を身につける教育モデルを確立する。最終的には、国内外での応用を視野に入れながら、分断を乗り越える実践的な道筋を提示し、民主主義社会の安定と発展に寄与することを目指している。

2. 主な研究成果と社会、学術へのインパクト

本研究は、道徳基盤を日本語圏向けに正確に測定する新たな尺度（j-MFQr）を開発するとともに、大規模言語モデル（LLM）を用いた議論環境を構築し、それらを通じて「分断」が顕在化する対立局面を多角的に実証研究した点に大きな独創性がある。まず、欧米で提唱された道徳基盤理論(Moral Foundations

Theory)を日本の文化・言語環境に合わせて再検証・改訂し、従来の翻訳版だけでは捉えづらかった因子構造の安定化を実現した j-MFQr は、社会や教育の現場で多様な価値観を可視化するうえで大きく貢献する。保守・リベラルといった単次元のイデオロギー軸にとどまらず、人々が重視する道徳観の違いを測定し、そこからウェルビーイングや意見対立への影響を考察可能にしたことは、従来研究にない新たな取り組みである。

さらに、LLM ベースの ChatBot と参加者がオンラインで意見を交わす実験環境を整備することで、人間同士の強い対立を直接引き起こさずに、自然な「衝突体験」を生み出すことに成功した。道徳観が異なる ChatBot とのやり取りによって被験者の意見が変容する過程を精密にログ解析できるため、対立が発生する条件や対立の緩和要因を実験室レベルで制御・検証できる。これは対面式実験やシナリオ調査では得られなかった詳細データを取得できるという点で学術的意義が高い。加えて、社会的には、大規模かつ再現性のある実験を短期間で実施できるうえ、倫理的リスクや参加者保護への配慮も行いやすい利点がある。

総合すると、道徳基盤の測定尺度とオンライン対立議論環境の両面からアプローチする本研究は、日本における分断克服の方策を探索するうえで、学術と社会を接続する独創的な枠組みを提示していると言える。民主主義が抱える「分断」のメカニズムをより深く理解し、教育や政策に活かせる新たな可能性を示した点こそが、本研究の大きなインパクトである。

3. 研究成果

3.1 日本語版道徳基盤尺度の開発

民主主義の弱点である「分断」を克服するには、人々が何を道徳的に重視しているかを正確に把握し、対立の背景にある心理的要因を解明することが不可欠である。そこで、本プロジェクトの目標達成に向けて、道徳基盤を測定できる尺度の開発は極めて重要となる。

本研究プロジェクトの初期段階では、民主主義社会における「分断」を乗り越え、多様な価値観の共存を実現する一助として、道徳心理学の視点から人々の価値観を測定する新たな尺度を作成した。具体的には、Haidt らが提唱した道徳基盤理論（Moral Foundations Theory）に基づき、日本語版の質問紙を再構築し、特に「Individualizing（個人志向)」と「Binding（集団志向)」という二つの基盤を精度高く評価できるよう工夫した。既存の英語版質問紙を単純に翻訳しただけでは因子構造が安定しにくいという問題点が指摘されていたため、本プロジェクトでは道徳基盤辞書（Moral Foundations Dictionary）の日本語版を活用しながら、項目選定や表現の微調整を行い、日本人特有の文化的・言語的背景を踏まえた新しい測定尺度（j-MFQr）の開発に取り組んだ。

本プロジェクトの成果として誕生した j-MFQr は、内集団への忠誠や権威を重んじる「Binding」基盤と、配慮・公正といった「Individualizing」基盤を区別して測定できるだけでなく、従来の「純粋性（Purity)」要因の扱いにも配慮している点が特徴である。調査の結果、日本人を対象とした場合、内集団や権威への志向が満足感やポジティブな感情にプラスに働く一方、他者の不公平や苦痛に敏感になる傾向を持つ個人志向はネガティブな感情との関連がやや高いことが示唆された。さらに、欧米でしばしば指摘される「保守層ほど幸福度が高い」という傾向は日本では統計的に明確に確認されず、代わりに「Binding」基盤そのものが幸福感を左右する有力な要因である可能性が示された。

このように、本研究プロジェクトによる新しい尺度は、日本の社会文化的文脈を踏まえつつ、道徳基盤の違いがウェルビーイングや社会的対立にどのように影響するかを捉えるうえで有用なツールとな

る。教育現場やコミュニティ活動、さらには政策立案の場など、多様な領域で本尺度を活用し、人々が互いの価値観を理解・調整するための具体的方策を探究することが期待される。このようなアプローチによって、民主主義社会が内包する「分断」を乗り越え、新たな合意形成や協働の可能性を広げていくことが本プロジェクトの狙いである。

3.2 意見対立を経験する議論環境の構築

本研究では、民主主義社会が内包する「意見の対立」や「分断」のメカニズムを実証的に把握するために、LLM（大規模言語モデル）を活用した ChatBot と被験者が議論を行うオンライン環境を開発した。

まず、本研究が構築した議論環境は、対立しやすい社会的テーマを議題として取り上げ、参加者がオンライン上で「チャット形式」のやり取りを通じて意見交換を行う仕掛けである。議論の相手は実際の人間ではなく、事前に「特定の道徳基盤（Moral Foundations Theory に基づく Individualizing または Binding)」を備えるようにプロンプト設計された ChatBot が務める。たとえば、個人の権利や公正さを重視する個人志向（Individualizing）Bot と、集団秩序や伝統を重んじる集団志向（Binding）Bot を用意し、参加者とそれぞれがオンラインで議論を交わすことで、意見対立が自然に生じるようにする。

これにより、被験者は「異なる正しさ」をもつ相手と、ある程度の対立を含んだディスカッションを疑似的に体験することができる。しかも、あたかも他の参加者と対話しているような演出を施すことで、「自分とは異なる思想・道徳観を持つ他者と対立する感覚」を得るように設計している。

本議論環境は、主に以下のコンポーネントから構成される。

フロントエンド（Web アプリ）：Python および Streamlit 等のフレームワークを用い、ブラウザ上でチャットを行う画面を構築した。参加者はこの画面上で意見を入力し、Bot からの応答を受け取る。チャットアプリのような UI を採用し、「アイコン表示」「時差応答」などを組み合わせることで、相手が“人間”らしく振る舞っているかのように見せる工夫を施している。

バックエンド（LLM＋対話制御ロジック）：ChatBot の動作には、OpenAI の GPT-4 や GPT-3.5 といった LLM を API 経由で利用し、LangChain などのライブラリで会話履歴の文脈管理を行う。Bot の「道徳基盤」は主にプロンプトの設計によって実現され、たとえば Binding（集団秩序重視）の Bot には「権威や秩序を尊重する発話」を返すような指示を与える。

シナリオに応じて Bot は「賛成／反対」「個人志向／集団志向」といった立場を一貫して保持し、被験者の発言に応じた応答文をリアルタイムで生成する。

上記のシステムを介して、異なる立場の Bot と参加者との間で対立を含んだ議論が実際に展開される。具体的には、以下のような特徴をもつ議論状況が実現される。個人志向 Bot は「個人の苦痛や不公平を重視する立場」から、集団志向 Bot は「社会的秩序や伝統を重視する立場」から、それぞれ意見を主張する。そのため、被験者の意見が Bot と食い違う際に、自然な議論の摩擦が生じやすい。

このような LLM ベースの ChatBot を社会心理学的実験に利用する意義は、以下のようにまとめられる。

精密な実験条件の設定と再現性：従来の社会心理学的実験では、実際の人間を「さくら」として配置し、特定の役割を演じてもらう手法が多用されてきた。しかし、人間同士のやり取りはバラつきが大きく、実験条件の統制が難しい。LLM ベースの ChatBot であれば、同一の道徳基盤や主張を一貫して保持しつつ、多数の被験者に対して再現性の高い議論を提供できる。

大規模実験の実施：オンライン完結型の議論環境は、被験者数を大きく増やすことも比較的容易である。

地域や時間帯の制約を受けにくく、大規模かつ多様な参加者を集めることで、対立要因や意見変容のプロセスをより広範に検証できる。
倫理的リスクの低減と対立の操作：人間同士で強い対立を引き起こすような実験は、倫理面や被験者保護の観点で困難が伴う。ChatBot であれば、感情的負荷をある程度管理した上で、刺激としての対立を生じさせることができるため、実験参加者へのリスクを相対的に抑えられる。
社会的・教育的応用：分断克服を目的とした「対立から学ぶ」教育プログラムとしても有効活用が期待される。異なる価値観や道徳観を持つ複数の Bot と安全に議論練習ができるため、異文化理解や合意形成のトレーニング環境としての展開可能性もある。

3.3 意見対立を経験する議論環境の構築

本研究の第 3 の成果として、実験を通じて、意見が対立する議論のプロセスやその特質に関する実証的知見を得られたことが挙げられる。ここで得られた総合的な成果は、従来の対面式実験やシナリオ調査では得られにくかった多面的なデータをもたらし、民主主義社会が抱える「分断」の問題を解明する上で大きな意義を持つと言える。

本研究では、大規模言語モデル（LLM）を用いて特定の道徳基盤（例：Individualizing／Binding）を与えられた ChatBot と被験者が議論する実験環境を構築した。興味深いことに、被験者の過半数が議論相手を「AI ではなく人間」だと信じ続けたことが確認された。

これは、ChatBot の応答が一定の自然性や一貫性を維持し、実験中に演出されたタイムラグや発話上限設定などの仕組みが、対面実験に近い“リアリティ”を醸成するのに寄与した結果である。実際の人間同士の議論と比較した場合、細かな表情・声の質感などが伝わらないという制約はあるものの、本研究では十分に「相手がいる環境」を再現できていたと考えられる。この点は、議論の過程を計測・分析する上で、被験者が“仮想”であることに気付かず自然に議論を展開したという意味で、実験環境の妥当性を裏付ける重要な証拠と言える。

次に、対立的な議論を行うことで、被験者の意見がどの程度変容するのかを検討した結果、ChatBot 側の道徳的立場（たとえば個人の権利を尊重する“Individualizing”か、社会秩序を重んじる“Binding”か）によって、被験者の最終的な態度が有意に変化する傾向が確認された。

具体的には、個人志向の Bot と議論を交わした被験者は、議論テーマ（例：ある政策の賛否）に関して、議論前よりも「賛成寄り」に寄る傾向が強まる一方、集団志向の Bot と対話した被験者は「反対寄り」に転じるケースが見られた。このように、相手が提示する道徳基盤や説得のロジックに被験者が引き寄せられ、議論前後で統計的に有意な態度変化が生じている点は、AI を使った議論環境が実際に意見形成・変容へ影響しうることを示すものである。さらに、被験者の道徳観や事前態度との相互作用も観測され、道徳基盤が異なる相手とぶつかることで、より強く反発する場合や逆に理解を深める場合があることも明らかになった。

加えて、本研究では意見が正面衝突する局面におけるやり取りの詳細ログや、被験者がそのとき感じた感情や認知プロセスを多角的に収集・分析できた点が大きな特徴である。対面式の実験では、被験者同士の会話を録音・録画することは可能だが、実際の発話データを定量的に扱うには限界があった。また、シナリオ調査では、あらかじめ用意された文章を読み取って回答するため、参加者自身が主導的に意見交換を行うプロセスはほとんど捉えられない。

一方、本研究の議論環境では、ターンごとに被験者と ChatBot の発話内容がすべてログ化され、発話の長さやタイミング、言及されたキーワード（例：権威、自由、公平性など）を精緻に分析することが可能となった。これにより、両者の主張がぶつかった際の感情的変化（相手を「冷たい」と感じ始める瞬間など）や、譲歩を試みる具体的な文脈（「ここは理解できるが、ここは受け入れられない」など）

が明確に抽出される。さらに、議論後のアンケートで相手の印象（温かい／冷たい、好ましい／好ましくないなど）やコミュニケーションの質（スムーズ／不快）を測定することで、対立による心理的ストレスや、逆に調停や妥協が生まれうるプロセスを捉えることができた。

4. 今後の展開

本研究で得られた二つの基盤技術――j-MFQr と LLM 議論プラットフォーム――は、民主主義社会の“分断”という抽象課題を、測定 → 対話 → 介入という具体的サイクルで扱える点に独創性がある。ただし教育的応用は、大学生対象の小規模ワークショップで初期的有効性を示した段階にとどまり、義務教育・社会人研修へ波及するモデルを十分に構築できなかった。本格展開に向けては、まず「道徳プロファイル可視化 → 仮想対立体験 → リフレクション」を 45 分×複数回の授業ユニットとして再設計する必要がある。教師用ダッシュボードには、学級全体の基盤分布をヒートマップで示し、発言ログから抽出した「対立ポイント」「共感ポイント」をリアルタイムに提示する。教員はそれを手がかりに、価値観の衝突が生じた瞬間を逃さずファシリテーションできる。さらに SEL（社会性と情動の学習）カリキュラムや情報科の SNS リテラシー単元と連結させ、“衝突のメカニズムを学び、建設的対話で解決する”という一貫した学習目標を掲げたい。教育効果を検証するため、前後比較だけでなく 1 学期間の縦断調査を実施し、相互理解志向・同調圧力・心理的安全性といった指標がどの程度改善するかを追跡することが次段階の課題である。

他分野への応用としては、j-MFQr と LLM を統合した「ウェルビーイング伴走型 ChatBot」の開発が特に有望である。本 Bot は①利用者の道徳傾向、②Big Five パーソナリティ、③直近の感情トレンドを入力に取り、応答トーン・語彙・提案行動を最適化する。たとえば、高い Binding 傾向と保守的価値をもつ高齢者には「伝統・家族・奉仕」を軸にした共感的メッセージを返し、孤立感を低減する。一方、Individualizing が強い若年層には、公正感覚や社会正義を肯定的にフィードバックしつつ、過度な怒りや疲弊を緩和するコーピング案を提示する。医療・介護領域では、長期入院患者や認知症高齢者の孤独感・不安感を軽減する非薬物的介入として効果が期待される。企業領域では、チームビルディング研修に組み込み、メンバー各自の道徳観が衝突する局面を安全にシミュレーションし、DEI（多様性・公平性・包摂）ワークショップの深化に貢献できる。自治体では、住民説明会前に Bot が住民の不満や価値観差を吸い上げ、議論設計を支援することで、対立の先鋭化を未然に防ぐ「対話型合意形成支援システム」としての導入が視野に入る。

これらの応用を実現するためには、プライバシー保護・バイアス制御・説明責任といった倫理的枠組みを備えた実装ガイドラインが不可欠である。本研究チームは、データをローカル暗号化するエッジ AI 化、応答生成過程の可視化、第三者監査プロトコルの策定に取り組む予定である。最終的には、「道徳計測 × AI 対話 × 教育／福祉／公共」をつなぐ国際共同研究コンソーシアムを立ち上げ、文化差比較や多言語版尺度の開発を進めることで、世界的規模で民主主義を再活性化する知的インフラへと発展させたい。

5. 発表実績

【論文】

Miwa, K., Yamakawa, M., Nakahara, A., Shimada, A. in press. Communication Experiment with Moral-based Chatbot: Toward Solution of Divisions in Democracy. AI & Society: Knowledge, Culture and Communication.

【学会発表】

Miwa, K. Relationship Between Moral Foundations, Political Ideology, and Life Satisfaction in Japan. In ISQOLS 2023.

Miwa, K., Onuma, Y., Yamakawa, M., Fujita, S., and Kobayashi, T. Lexical Analysis of Japanese Moral Textbooks Based on Moral Foundations Theory. In EARLI 2024.

Miwa, K. Can AI Generate Moral Narratives Like Humans? in APPE 2025.

Miwa, K., Yamakawa, M., Nakahara, A., Shimada, A. AI-Driven Moral Chatbots in Promoting Empathy and Bridging Divides in Moral Education. In AME 2025.

山川真由・島田彩乃・三輪和久（2024）．道徳基盤と社会問題に対する意見の関連，『日本認知科学会第41 回大会発表論文集』, 212-214.

【特許】

なし

【その他】

なし