

令和 6 年 6 月 12 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2019～2023

課題番号：19K03215

研究課題名(和文)人工知能をもたらす結果に対する一般の人々の責任判断：判断の心理過程と文化差

研究課題名(英文) Responsibility judgments by lay people for the outcomes caused by artificial intelligence: Psychological processes and cultural differences in judgments

研究代表者

膳場 百合子 (Zemba, Yuriko)

早稲田大学・理工学術院・教授

研究者番号：00548886

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：AIが引き起こした被害に対し一般の人々がどのように責任判断をするか、また、日本人の判断にどのような特徴があるかを日米比較web調査実験で調べた。いずれの文化の人々も、AIの心的能力(思考・感情)や道德能力は人間や組織より低いと考え、AIが被害をもたらすと、人間や組織がもたらした場合より責任を多方面に拡散させることが分かった。また両文化とも、AIに感情を知覚する人ほどAIに道德的能力を知覚していた。文化差として、日本人は米国人に比べ、「AI」と「人間や組織」の心的能力や道德能力の差を小さくとらえる傾向があり、また、AIの行為に対して製造者より使用者の責任を重視する傾向があることが分かった。

研究成果の学術的意義や社会的意義

人間や組織に対する責任判断は文化ごとの人間観や組織観によって異なることが知られていたが、人間を模した存在であるAIの場合に同様の文化差があるのかどうかは明らかではなかった。今回の調査を通じ、AIが関与する出来事への責任判断の特徴に、多くの文化共通性があることが明らかになった点に、まず意義がある。それと同時に日本人がアメリカ人ほどAIと人を差異化していない点や、AIの行為の責任を誰に多く負わせるかが文化によって違う点などいくつかの興味深い文化固有性が見いだされた点も意義がある。

研究成果の概要(英文)：The present research investigated how people judge responsibility for harm caused by AI, and whether there are culturally unique patterns in Japanese judgments. Results of Japan-U.S. comparative on-line experiments revealed both cultural commonalities and differences. In both cultures, people perceived AI to have less mental (thinking and feeling) and moral ability than a human or an organization, and when harm was caused by AI, compared to when it was caused by a human or an organization, people were more likely to extend responsibility to various targets. Also in both cultures, the more people thought AI can experience emotion, the more they perceived moral ability in AI. As cultural differences, Japanese tended to view the difference between the mental and moral capacities of "AI" and "a human or an organization" as smaller than that of Americans, and they also tended to emphasize the responsibility of the user rather than the manufacturer for the actions of AI.

研究分野：社会心理学

キーワード：AI 責任判断 原因帰属 文化

様式 C-19、F-19-1 (共通)

1. 研究開始当初の背景

AI の開発と普及が進み、AI が引き起こす被害の責任を誰に割り付け、どう処理するのが良いのか、という問題が新たに生まれてきている。本研究はそうした社会的背景と関心のもと開始した。社会心理学では伝統的に、責任は「意図」や「動機」を持ち道徳的善悪が判断できる「人間」に割り付けられるものとされてきた (e. g., Shaver, 1985; Weiner, 1995)。しかし、現代社会では人間以外の存在 (組織や AI) が人間の社会活動を分担したり主導したりすることが増え、それらが人に被害を及ぼした際の責任を誰にどう割り付けるかが問題となってきた。

人間以外に対する責任判断のうち、「組織」に対しては人々が組織に意図や責任を知覚し (日置・唐沢, 2010) 代表者を責めるなどの方法で責任追及することが、先行研究で分かっている (Zemba, 2006, Zemba, Young, & Morris, 2006)。一方、AI に関しては人々が誰に責任があると考え、どのような責任処理を良しとするかがほとんど分かっていない。AI が引き起こした被害に対し、一般の人々はどのような責任判断をするのだろうか。また、人間や組織に対する責任判断は文化によって異なることが知られているが、AI に対する判断も同様だろうか。これらの間を検討するために、本研究を行った。

2. 研究の目的

本研究の目的は、AI が引き起こした被害に対し、一般の人々がどのように責任判断をするか、また、日本人の判断にどのような特徴があるかを明らかにすることであった。具体的には、①AI のとらえ方、②AI が引き起こした結果に対する責任の判断、③AI が引き起こした結果に対する望ましい問題処理の判断、のそれぞれについて日米の文化比較を通じて文化共通の判断過程と、文化固有の判断過程 (特に日本人の特徴) を明らかにすることを目指した。

3. 研究の方法

AI のとらえ方 (どの程度心を備えた、道徳的善悪判断能力がある存在とみなすか) や、AI が引き起こした結果に対する責任、問題処理に関する判断を調べるために、実験操作付きの web アンケート調査を調査会社に委託し実施した。上記の①~③の目的の検討は日米比較 web アンケート調査 2 件を通じて行った。それとは別に、研究過程で生じた新たな問題を別途独立の web アンケート調査で調べた。

(1) 日米比較 web アンケート調査 1

1 件目の比較文化調査では、日米 1200 名の回答者 (20~69 歳、男女比均等) が医療診断とその結果が描かれた文章を読み、文章の内容についての質問に答えた。回答者に提示された文章の中で、診断行為者 (AI vs 個人 vs 組織) と出来事結果 (診断が正しく治療に成功 vs 誤診で治療に失敗) が操作されており、話は全部で 3 (行為者) x 2 (出来事結果) の 6 種類あり、そのうちのいずれか 1 つがランダムに回答者に提示されるようになっていた。質問には、診断行為者に対するとらえ方に関する質問 (心的能力※、道徳能力の知覚を尋ねる質問)、一般論として AI がもたらした被害の責任は誰が負うべきか、AI がもたらした利益の手柄をもらう資格は誰にあるか、AI をコントロールする責任は誰にあるか、を問うものが含まれていた。※心的能力の知覚は Gray ら (2007) の尺度を用い、主体的に思考することに関わる能力と、感覚や感情を経験することに関わる能力のそれぞれの知覚を調べた。

(2) 日米比較 web アンケート調査 2

2 件目の比較文化調査では、日米 1200 名の回答者 (20~69 歳、男女比均等) が医療診断の被害、または、企業の採用支援における被害が描かれた文章を読み、文章の内容に関する質問に答えた。1 件目の調査と同様、被害原因を作った行為者の種類 (AI vs 個人 vs 組織) が文中で操作されており、話は全部で 3 (行為者) x 2 (トピック: 医療診断、採用支援) の 6 種類あり、そのうちのいずれか 1 つがランダムに回答者に提示されるようになっていた。質問には、被害原因を作った行為者に対する受け止め方 (心的能力、道徳能力の知覚)、被害発生の原因と責任、望ましい問題解決の判断、および、一般論として AI をコントロールする責任が誰にあり、被害に備える制度としてどのような制度が望ましいか (誰に賠償義務を負わせるのが良いか、保険はどうかなど)、を問うものが含まれていた。

(3) 別途行った web アンケート調査

人は AI に感情を知覚するほど AI に道徳や責任を知覚する、という関係性が日米比較の調査で得られたため、AI にどのような種類の感情能力 (感情を読み取る能力か、感情を経験する能力か) が知覚されると、AI に道徳や責任が知覚されやすくなるのか、という新たな問が生まれた。その問を調べるため、web アンケート調査を実施した。別の研究費であらかじめ日本調査を行い、それと対応する米国調査を本助成金で行った。500 名のアメリカ人回答者 (20~69 歳、男女比均

等)がAIの診断ミスで被害が生じる文章を読み、文章の内容に関する質問に答えた。文章の中で登場するAIの感情読取能力(有、無)と感情経験能力(有、無)が操作されており、話は全部で2(感情読取能力:有、無)×2(感情経験能力:有、無)+1(感情能力に関する情報なし)の5種類あり、そのうちのいずれか1つがランダムに回答者に提示されるようになっていた。質問には、AIの責任と道德能力の知覚を問うものが含まれていた。

4. 研究成果

(1) AIのとらえ方

AIのとらえ方には、以下の文化共通の特徴が見られた。まず、全体的な傾向として人々はAIの心的能力(考えたり感じたりする能力)や道德能力(道德的善悪を判断しそれに従って行動する能力)を、人間や組織のそれらよりも低く見積もっていた。人々は、人間と組織には近いレベルの心的能力(考えたり感じたりする能力)や道德能力を知覚していたが、AIのそれらの能力は人や組織よりも低く評価していた。たとえば行為者に道德的能力が知覚される確率は、人を1とした場合、組織行為者は0.9程度、AIは0.5程度だった。また、「主体性次元(主体的に思考することに关わる能力次元)」と、「経験性次元(感覚や感情を経験することに关わる能力次元)」とでは、経験性次元の心的能力において、より、AIは人や組織に劣ると人々は考えていた。いずれの文化でもAIに「経験性次元」の能力を知覚する人ほど、AIに道德能力も知覚する傾向が見られた。

AIのとらえ方には文化差もあり、アメリカの方が日本人よりも、「AI」と「人・組織」の間に大きな心的能力の差を知覚していた。この文化差は主体性に关わる心的能力次元でも感情や感覚の経験性に关わる心的能力次元でも見られた。道德能力次元でも同様で、「AI」と「人・組織」の間の道德能力の差を、アメリカの方が日本人よりも大きくとらえていた。

(2) AIがもたらした結果に対する責任の判断

一般論としての判断を求められた際、日本人もアメリカ人も、AIの行為に対する責任や手柄(AIをコントロールする責任 / AIがもたらす被害に対して取るべき責任 / AIがもたらす利益に対して手柄をもらう資格)の1, 2割程度がAIにあり、残りの8, 9割程度が製造者、使用者、所有者にあると判断していた。文化差については、日本人はアメリカ人に比べ、AIの性質を作り出した製造者の責任を軽く評価し、逆に、AIを監視・監督しうる立場の使用者(ユーザー)の責任を重くとらえる傾向が2件の調査で一貫して見られた。日本人はAIの使用者に製造者と同等かそれ以上の責任を割り付けていたが、アメリカ人は使用者より製造者に大きな責任を割り付けていた。所有者は、使用者や製造者よりは責任が小さいと判断されていた。AI自身の責任を強調する傾向は1件目の調査ではアメリカの方が日本人よりも強かったが、この文化差は2件目調査では見られなかった。

具体的な出来事への責任判断では、人々は、出来事結果がAIによってもたらされていると、人間や組織によってもたらされた場合よりも責任を行為者以外に拡散する傾向が見られた。すなわち、AIが原因をつくっている場合、原因を作った行為者の責任を小さく評価し、行為者を取り巻く他者の責任を大きく評価する傾向が見られた。AIが原因行為者だと責任の拡散が生じる、という現象は文化共通で見られた。

(3) AIが引き起こした結果に対する望ましい問題処理の判断

一般論として、AIがもたらす被害に備える制度として、a. 賠償義務をAIの「持ち主」に負わせる制度、b. 賠償義務を「AIを開発した技術者」に負わせる制度、c. 賠償義務を「製造販売業者」に負わせる制度、d. AIに法的な人格と財産権を与え賠償義務を「AI自身」に負わせる制度、e. 被害をもたらすリスクがあるAI製品は保険加入を義務付け被害の補填を保険でまかなう制度、の各制度がどの程度良い制度だと思うかを尋ねたところ、日米とも「保険」が最も支持され「AIに法人格と財産権を与えて賠償義務をAI自身に負わせる」制度は最も支持されなかつ

[ここに入力]

た。「持ち主」、「製造販売業者」、「技術者」に賠償義務を負わせる各制度への支持がその中間であった。

具体的な出来事（医療診断ミス）に対する問題解決法として、a. ルールの厳格化、b. 行政の管理の強化、c. 失敗事例の情報公開、d. 失敗した行為者の活動の制限、e. 失敗原因が取り除かれるまで行為者の活動資格停止、f. 組織に再発防止措置を講じることを義務付ける、のそれぞれをどの程度良い解決法だと思うかを尋ねたところ、行為者に働きかける解決法（活動停止や資格停止）よりは再発防止の環境整備（ルール厳格化、行政管理強化、情報公開、組織に再発防止措置を講じさせる）の方が日米ともやや好まれていた。ただし、行為者がAIの場合は、人間や組織の場合より行為者に働きかける解決法が好まれる傾向があった。またアメリカの方が日本人よりも各問題解決法を良い解決法と評価していたが、直接行為者に直接働きかける解決法（直接行為者の活動停止や資格停止）で文化差はもっとも顕著で、日本人はアメリカ人に比べ、行為者に直接働きかける解決法を好まない傾向が見られた。

(4) その他

AI にどのような種類の感情能力（感情を読み取る能力か、感情を経験する能力か）が知覚されると、AI に道徳や責任が知覚されやすくなるのか実験で調べた。別の研究費で事前に日本で実施していた同じ内容の調査と、本助成金で実施した米国での調査の結果から、日本でも米国でも、AI が感情を経験する能力を備えていそうに見えるほど、人はAI に責任を知覚することが明らかになった。また、AI に法的人格を与える制度を支持する人の割合は、下の年代（20～40代）の方が上の年代（50～60代）でよりも高いことも明らかになった。

<引用文献>

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *science*, 315(5812), 619-619.

日置孝一, & 唐沢穰. (2010). 集団の実体性が集合的意図と責任の判断におよぼす影響. *心理学研究*, 81(1), 9-16.

Shaver, K. G. (2012). *The attribution of blame: Causality, responsibility, and blameworthiness*. Springer Science & Business Media.

Weiner, B. (1995). Inferences of responsibility and social motivation. In *Advances in experimental social psychology* (Vol. 27, pp. 1-47). Academic Press.

Zemba, Y. (2006). Responses to organizational harm: Mechanism of blaming managers as proxies for a culpable organization. *Asian Journal of Social Psychology*, 9(3), 184-194.

Zemba, Y., Young, M. J., & Morris, M. W. (2006). Blaming leaders for organizational accidents: Proxy logic in collective-versus individual-agency cultures. *Organizational Behavior and Human Decision Processes*, 101(1), 36-51.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計7件（うち招待講演 0件 / うち国際学会 2件）

| |
|------------------------------------|
| 1. 発表者名 膳場百合子 |
| 2. 発表標題 AIの感情能力と責任 –米国サンプルでの追試– |
| 3. 学会等名 日本社会心理学会第64回大会 |
| 4. 発表年 2023年 |

| |
|---|
| 1. 発表者名 Yuriko Zemba & Maia Young |
| 2. 発表標題 Public perceptions of AI's moral ability and responsibility: Japan-US comparison |
| 3. 学会等名 26th International Congress of the International Association of Cross-Cultural Psychology (国際学会) |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 膳場百合子 |
| 2. 発表標題 AIの感情能力と道德能力 「感情読取」や「感情経験」能力がありそうなAIは道德的に見えるか |
| 3. 学会等名 日本社会心理学会第63回大会 |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 Zemba, Y. & Young, M. |
| 2. 発表標題 Responsibility judgments for outcomes caused by AI: Japan-US comparison |
| 3. 学会等名 25th International Congress of the International Association for Cross-Cultural Psychology (国際学会) |
| 4. 発表年 2021年 |

| |
|--|
| 1. 発表者名 膳場百合子, Young, M. |
| 2. 発表標題 AIの道徳的判斷能力と責任を人々はどうかとらえているか 日本人とアメリカ人の判斷の比較 |
| 3. 学会等名 日本グループダイナミクス学会第67回大会 |
| 4. 発表年 2021年 |

| |
|-----------------------------|
| 1. 発表者名 膳場百合子, Young, M. |
| 2. 発表標題 包括的な因果認知傾向と責任判斷 |
| 3. 学会等名 日本社会心理学会第61回大会 |
| 4. 発表年 2020年 |

| |
|---|
| 1. 発表者名 膳場百合子, Young, M. |
| 2. 発表標題 AIがもたらした結果に対する責任判斷--日本人とアメリカ人の比較-- |
| 3. 学会等名 日本社会心理学会第60回大会 |
| 4. 発表年 2019年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

| | 氏名 (ローマ字氏名) (研究者番号) | 所属研究機関・部局・職 (機関番号) | 備考 |
|-------|----------------------------|-----------------------|----|
| 研究協力者 | ヤン マイヤ (Young Maia) | カリフォルニア大学アーバイン校 | |

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

| 共同研究相手国 | 相手方研究機関 | | | |
|---------|-----------------|--|--|--|
| 米国 | カリフォルニア大学アーバイン校 | | | |