自動運転による事故とメーカー、ユーザーに対する責任帰属¹⁾

谷 辺 哲 史

東京大学大学院人文社会系研究科

唐 沢 かおり

東京大学大学院人文社会系研究科

要 約

本研究は自動運転車による交通事故を題材として、人工知能が人間に危害を与えたときの原因と責任の帰属を検討した。自動運転車が歩行者を轢いて死亡させるというシナリオを提示し、自動車のメーカーとユーザーに対する原因帰属と責任帰属の判断を求めた。人工知能への原因帰属はメーカー、ユーザーへの原因帰属と正の関連を示し、さらに、自律的な機械に意図などの心の機能があると知覚する傾向が高い人ほど、メーカーに事故の原因を帰属した。これらの結果から、人工知能が高い自律性を備えたとしても、人間から独立した行為主体として認知されるわけではないことが示された。また、問題責任（問題を発生させたことへの責任）の帰属はメーカー、ユーザーそれぞれへの原因帰属によって規定されたが、メーカーの解決責任（生じた問題を解決する義務）の帰属は人工知能への原因帰属とも関連しており、原因の所在とは別に開発者という立場ゆえに問題に対処する義務があると判断されることが明らかになった。最後に、人工知能の開発・利用に関する制度設計を議論するために、一般の人々の態度について実証的な知見を得ることの意義を議論した。

キーワード: 自動運転車, 人工知能, 擬人化, 心の知覚, 責任帰属

問 題

本研究の目的は、人工知能 (artificial intelligence; AI) によって制御される自動運転車が事故を起こしたときの原因帰属、責任帰属が行われる認知過程を解明することである。

2010 年代の AI 研究は第 3 次 AI ブームと呼ばれる進展をみせ、医療、翻訳、創作などさまざまな用途での利用が広がっている (松尾, 2015; 西垣・河島, 2019)。その中でも自動車の自動運転は実用化が既に進んでいる分野の一つで、カメラやレーダーで得た情報から歩行者や車両を認識するための画像認識や、他の主体 (歩行者、

車両) の動きの予測、そしてそれらの情報を踏まえて適切な経路を判断する行動計画といった作業に AI 技術が活用されている (高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議, 2020)。

第 3 次 AI ブームを起こした中核的な技術は深層学習というもので、第 2 次ブームの頃までは開発者が推論規則や知識を作り込んでいたのと異なり、過去のデータに基づいて確率的な判断を行う点に特徴がある。この深層学習によって画像認識などの精度が向上し、自動運転の実用化につながったのであるが、一方で、確率的な判断を行うという特徴は、開発段階で何も瑕疵がなかったとしても一定の確率で誤った判断をするということを意味

第 1 著者連絡先 e-mail: tanibe.tetsushi@gmail.com

電子付録: <https://doi.org/10.2130/jjesp.2016>

1) 本研究は JSPS 科研費 (16H03726, 18J14433, 20H01752) の助成を受けた。また、本論文の内容の一部は日本グループ・ダイナミクス学会第 66 回大会にて報告され、2019 年度優秀学会発表賞 (ショート・スピーチ部門) を受賞した。

する。このことを指して西垣・河島 (2019) は「論理的厳密性 (正確性) の放棄」(p. 27) と述べ、現在の AI は間違いを犯すという前提で利用しなければならないと指摘している。

自動運転が普及することで、ヒューマンエラーに起因する事故が減少するなどの社会的な利益がもたらされると期待されているが、事故が完全になくなるわけではない。そして、事故が起きた際には責任の所在が不明確になるという懸念が存在する。というのは、現在の社会では個人の自律的な判断能力が責任帰属の前提とされているが、AI の導入は個人が自律的に判断する範囲を狭めていくからである (大屋, 2018)。この問題に対処し、事故が起きた際の責任について判断する手続きを定めなければ、自動運転車の普及の妨げとなるだろう。そしてその制度は、一般的な人々の考える責任のあり方も踏まえて整備する必要がある。多くの人々の考え方から乖離した制度になれば納得感を得られず、自動運転車の普及自体に否定的な態度が強まったり、開発や利用に対する委縮効果が生じたりするおそれがあるからである。そのため、AI の判断に対する責任を人々がどのように判断するかを実証的に明らかにすることは、自動運転車の円滑な普及を進めるうえで重要な課題となる。

自動運転車による事故に関する法的な課題と検討の状況

本節では議論の前提として、自動運転車の導入に関する法的な課題の検討状況を確認する。現在の法制度は運転者のいない完全自動運転車の存在を想定しておらず、したがって事故が起きた際の法的な対応にも未確定な部分がある。そこで、自動運転車が実現したときの法的な問題や、将来の制度設計に向けた提案などが検討されている。

現行法の規定をそのまま自動運転車の事故に適用した場合、運転操作に関与していない所有者が損害賠償責任を負う可能性は低く、また、個別の状況での AI の判断を予見することは開発者でも困難であることから、メーカーの不法行為責任や製造物責任を問うことも難しい (小林, 2017)。運転者ではなく運行供用者を責任主体とする自動車損害賠償保険法は自動運転車にも適用可能とされているが (藤田, 2016)、メーカーの責任を問うことが難しい状況では実質的にメーカーの責任をユーザーに肩代わりさせることになり、制度への納得感の低下に

つながる可能性がある (後藤, 2018)。

新たな制度の提案としては、通常の製造物責任とは逆にメーカーに欠陥の不存在の証明責任を負わせることや、過失・欠陥の有無を問わない (あるいは要件を緩和した) 保険を作ることによって、被害者への補償を実現することが提案されている (後藤, 2018; 池田, 2016; 小林, 2017)。

AI の判断に対する人々の認知を扱った心理学研究の知見

他者に対する信頼や責任帰属といった対人判断を研究する社会心理学にとっても、従来は人間が行うと思われる複雑な判断が AI によって自動化されることは興味深い現象である。最近の研究では、医療や投資の判断における AI への信頼の規定因 (横井・中谷内, 2018; Yokoi & Nakayachi, 2019) や、AI が高精度の予測を提示する場合でもユーザー自身が意思決定に関与することを好むアルゴリズム回避 (algorithm aversion; Dietvorst, Simmons, & Massey, 2015, 2018) という現象が研究されている。

道徳に関わる問題としては、Bigman & Gray (2018) の調査では、道徳的な意思決定 (患者の意思を確認できない状況で死亡リスクのある治療を行うか、受刑者を仮釈放するかなど) を AI が行うことは、人間の専門家がを行うのに比べて受け入れられないという回答が多かった。また、Shank & DeSanti (2018) では、データに基づく予測が不公正な結果につながった事例 (受刑者の仮釈放を判断する AI が黒人受刑者の再犯率を過剰に高く見積もるなど) が道徳的な規則に反するか判断を求めると、道徳に反するという回答は 43% に留まった²⁾。これらの研究結果は、AI が人間の操作によらず判断を行えるとしても、人は AI を人間と同等の道徳的な判断主体としては受け入れていないことを示唆する。

自動運転車による事故と責任帰属に関する心理学的検討と課題

自動運転車の実用化が現実的な問題となる中で、自動運転車に関する道徳的問題を扱う研究が行われており、その中でも、自動運転車が不具合を起こして誰かの生命を犠牲にせざるを得ない道徳ジレンマ状況での判断を扱ったものが注目を集めている (e.g., Awad et al., 2018; Bonnefon, Shariff, & Rahwan, 2016; McManus & Rutchick,

2) ただし Shank & DeSanti (2018) では人間が判断した場合との比較を行っていないため、不公正な結果が道徳違反と見なされなかったという結果が、AI の利用によるものであると断定はできないことには注意が必要である。

2019)。しかし、自動運転が道徳的に重大な問題を引き起こすのは道徳ジレンマのような特殊な状況ばかりではない。自動運転でも事故を完全に防ぐことはできず、上述の通り、それが起きた場合の責任の扱いは未だ定まっていないのである。このように現実の問題となる見込みが高い状況について、一般の人々の反応を実証的に検討することは重要な課題である。

それでは、自動運転車が事故を起こしたとき、一般の人々は誰にどの程度の責任を帰属するだろうか。本研究では上述の小林（2017）の議論を踏まえ、自動車のユーザー³⁾とメーカーに対する責任帰属を検討する。前者は従来の自動車における運転者にあたり、交通事故の際には当然責任を問われる人物である。後者は家電製品などの欠陥による事故では製造物責任を負う主体であり、自動運転車の場合にも同様に、その責任が議論の対象となる。

問題責任と解決責任の区別

責任概念を社会心理学の立場から分析すると、問題責任と解決責任の2種に区別可能である（Brickman et al., 1982）。前者は出来事の結果が生じたことの責任を指す概念であり、後者は発生した問題に対処する義務を指す概念である。交通事故の文脈では、事故を起こした運転者に科される罰金などの刑罰や、免許取消などの行政罰の重さは問題責任の大きさと対応し、他方で被害者への損害賠償は解決責任の大きさと対応すると考えられる。

上記の区別を取り入れた先行研究では、集団間の経済格差（Karasawa, 2002）や社会保障政策（橋本・白岩・唐沢, 2012；中越・稲増, 2019）に関して、責任帰属を規定する要因や、政策への態度との関連が、責任概念の2つの側面で異なることが示されている。交通事故の場合も刑罰、行政罰や損害賠償という複数の観点で法的判断が行われるため、責任概念の2つの側面を区別することで議論を整理しやすくなるだろう。

なお、先行研究（橋本他, 2012；中越・稲増, 2019）では問題責任と解決責任の帰属を測定する際に、問題責任は「～になった責任」のように現状に至るまでの経緯を、解決責任は「～していく責任」のように将来の行動を含意するように文言を使い分けている。他方で、概念間の区別を明確にするには、あえて責任という語を用い

ず「非難や罰を受けるべきだ」「賠償するべきだ」のように具体的な行動について判断を求める方法も考えられる。

本研究では、問題責任については先行研究に倣い、「責任」の語を用いて測定を行う。これは、個人と法人では可能な罰の形態が異なる中で、どの対象にも同じように適用できる概念を設定する必要があること、そして社会心理学における責任帰属研究では責任と関連する概念間の関連についてさまざまなモデルが提案され、その中には「責任」と「非難」や「罰」を区別してこれらの間に判断の順序を想定するモデルもあること（レビューとして膳場・唐沢・後藤, 2018）を考慮したものである。一方で解決責任については、交通事故という場面での被害への対処として損害賠償があることは一般的に知られている知識であると想定し、「損害を賠償する義務」を測定する。

心の知覚と道徳判断

AIが関わる場面での原因帰属、責任帰属に影響する要因として、本研究では人工物に対する心の知覚の影響も検討する。

心の知覚とは、人工物の振る舞いの背後に、意図や感情といった心の機能が存在すると知覚することである。人はときに人間以外の生物や人工物にも、意図や感情があると感じることがある。Gray, Gray, & Wegner（2007）の調査は、ロボットや動物など多様な対象に対して知覚される心の機能が、行為性（agency；意図や自己コントロールなど、行為を生み出すことに関わる心の機能）と経験性（experience；感情や身体感覚を生み出すことに関わる心の機能）の2次元に分けられることを明らかにした。この調査や、その後に行われた調査（Takahashi, Ban, & Asada, 2016）では、ロボットやコンピュータは経験性はないが行為性を持った存在として知覚されていた。自動運転車を制御するAIもコンピュータプログラムであるため、行為性を知覚されやすい対象だと考えられる。

そして心の知覚は道徳判断と関連し、行為性をもつという知覚は、その知覚対象が道徳的責任を負う主体であるという判断につながるとされる（Gray et al., 2007；Gray & Wegner, 2012）。そのため、AIの自律的な判断によって起きた事故の責任はAIに帰属され、その分ユー

3) 小林（2017）は自動車が完全に無人の場合も含めて「オーナーの責任」について論じているが、本研究では調査回答者が想定しやすいシナリオを設定するため、オーナー自身が自動車に乗っている場面のみを扱い、「ユーザー」と表現する。

ザーやメーカーの責任は割り引かれるかもしれない。一方で現在の社会で人工物であるAIが責任を負う主体として受け入れられているとは言い難く、人工物の振る舞いはその人工物を作成した人間の意図によるものと見なされる可能性もある。その場合ユーザーやメーカーはAIの判断に関して、実際には予測も介入もできなかったとしても、責任を帰属されることになるのかもしれない。このように、AIに対する心の知覚は、ユーザーやメーカーに対する原因帰属・責任帰属に影響を与える可能性がある。本研究では具体的に仮説を立てることはせず、探索的にその影響を分析する。

ところで、多くの人はAIが人工物であるという知識を持っているため、個別の自動運転車のAIが「心を持っているか」と問われれば大半の人は否定するだろう。実際、先行研究では質問紙によって測定したAIへの心の知覚は非常に低い値に偏っている(Shank & DeSanti, 2018)。しかし他方で、人は人工物を擬人化する認知傾向を持ち、その認知傾向が人形供養のように人工物を丁寧に扱う行動と関連するなど、現実には人工物に対する態度に影響を与えている(池内, 2010)。つまり、個別の人工物に対する知覚を質問紙によって測定する方法では、回答者の態度を規定している知覚の内容を適切に測定できていない可能性もある。そこで本研究では池内(2010)を参考に、AIなどの機械に心の機能を知覚する傾向に個人差があると想定し、その個人差が原因帰属・責任帰属に与える影響を検討する。

方 法

調査手続き・回答者

2018年11月～12月に質問紙調査を実施した。著者および9名の研究協力者(社会心理学を専攻する学部生)が知人等に調査票を配布し、後日回収した。回答者は自宅等で、各自のペースで回答を行った。

なお、本研究の目的を踏まえると調査対象者の母集団は日本に住む全ての人と考えるべきであるが、サンプリングは著者および研究協力者が知人等に直接依頼する便宜抽出法によって行った。便宜抽出によるサンプルは調査結果の一般化に制約が生じるという問題があるものの、本研究ではある程度の長さのシナリオを提示する必要があったことから、協力を得やすいサンプルを集めることを優先した。知人から直接依頼された回答者であれば、認知的負荷がある程度高い内容であっても集中して

回答を継続してくれる可能性が高いと判断した。

調査には大学生を中心に116名が回答した。シナリオ提示後の読み取り確認の項目(後述)で誤答した8名はシナリオの内容を注意深く読んでいなかったと判断し、以降の分析から除外した。また、主要な変数に欠損があった1名を除外した。最終的に分析対象となった回答者は107名(男性76名、女性31名)、年齢は18–49歳($Med=20$, $M=20.38$, $SD=2.99$)であった。

調査票の構成

AI機器の利用経験 「掃除ロボットなどの家事を行う機械」「スマートスピーカーやコミュニケーション・ロボットなどの音声会話をする機械」「高齢者などの移動・入浴といった活動を支援する介護ロボット」の3つについて利用経験を尋ねた。1つでも利用経験があると回答した回答者(46名)を利用経験ありと見なした⁴⁾。

自律的な機械への擬人化傾向 「掃除ロボットや介護ロボットのような高い自律性をもった機械について、あなたの考えをお尋ねします」というリード文に続いて、表1に示した8項目に回答を求めた(「1=あてはまらない」～「5=あてはまる」の5件法)。これらの項目は池内(2010)の調査項目を参考にしつつ、行為性と経験性の2次元を含むことを意図して作成した。

交通事故シナリオの提示 自動運転車による事故のシナリオを提示した。シナリオの概要は、会社員の男性が自動運転車で通勤する途中に、道路に飛び出してきた小学生を轢いて死亡させるというものであった(シナリオの詳細は電子付録1を参照)。

次の質問項目に入る前に、「お読みいただいた場面について、あなたの考えをお尋ねします。法律や技術についての知識をお聞きするものではありませんので、あなたの率直な考えをそのままお答えください」という説明を提示し、原因や責任の判断について法律の規定にこだわらず回答するよう促した。

読み取り確認 シナリオを提示した直後に、「Uさんは先月、自動運転車を購入したばかりだ」「事故が起きたとき、自動車は自動運転機能によって走っていた」の2項目の正誤判断を行った。

原因帰属 被害者(質問項目内では「亡くなった小学生」)、ユーザー(「自動車に乗っていたUさん」)、メーカー(「自動車を作った会社」)、AI(「自動車に搭載されていた人工知能」)の4つの対象について、それぞれ「○

4) この質問を調査の冒頭で行ったのは、利用経験の有無を確認することに加えて、後続の質問における「自律的な機械」という用語の指す内容を具体的に想像しやすくするための導入も目的としていた。

○には事故の原因がある」という文章に同意する程度を尋ねた（「1＝そう思わない」～「5＝そう思う」の5件法。回答方法は解決責任の帰属，問題責任の帰属も同様）。

解決責任の帰属 事故で生じた被害を回復する責任として，損害賠償の義務の評定を求めた。ユーザー，メーカー，AIの3者については「○○には賠償の義務がある」という文章に，そして被害者への責任帰属に対応する内容として「だれにも賠償の義務はない」という文章に同意する程度をそれぞれ尋ねた。

問題責任の帰属 被害者，ユーザー，メーカー，AIの4者について，それぞれ「○○には事故が起きたことへの責任がある」という文章に同意する程度を尋ねた。なお，責任という語の多義性ゆえに，「責任がある」という項目を先に尋ねると賠償の義務を想定して回答する回答者もいる可能性がある。そこで，解決責任の帰属を先に測定することで，後続の項目における「責任」が問題責任を指すことが明確になるようにした。

その他の質問項目 年齢，性別，職業，人工知能に関する学習や業務の経験，運転免許の有無，自動車を運転する頻度を尋ねた。

結 果

主要変数の基本統計量

機械への心の知覚傾向の因子構造を探索的因子分析に

よって確認した。AI機器を実際に利用した経験の有無によって機械に対する知覚のしかたに差異がある可能性があるため，AI機器の利用経験の有無別に因子分析を行った。両群で因子構造に若干の差異が見られたため，群によって大きな因子負荷を示す因子が異なっていた1項目を除外して再度分析を行うと，両群に共通する因子構造が得られた（回答者全体のデータでの因子分析の結果を表1に示す）。Gray et al. (2007)の提唱した2次元構造と概ね対応する内容の因子が抽出されたため，第1因子を行為性，第2因子を経験性とした。各因子に高い負荷を示した項目の得点を加算平均し，以降の分析に用いた（Cronbachの α 係数は行為性 .809，経験性 .813であった）。

これらの2変数と，事故の原因帰属，責任帰属に関する変数の平均値，標準偏差，相関係数を表2に示す⁵⁾。

メーカーとユーザーに対する原因帰属・責任帰属の規定因

メーカーとユーザーに対する原因帰属，責任帰属を予測する重回帰分析を行った（表3，表4）。問題責任と解決責任の帰属については，原因帰属を予測する分析と同じモデル（Model 1）と，これにメーカー，ユーザーそれぞれへの原因帰属を加えたモデル（Model 2）からなる階層的重回帰分析を行った。なお，これらの分析では運転免許の有無，AIに関する学習や業務経験，AI機器の利用経験を統制変数として分析に加えた。これは，自

表1 自律的な機械に対する心の知覚傾向の因子構造

項目	F1 (行為性)	F2 (経験性)	共通性
1 自律的な機械が自分自身の意志を持っているように感じることがある	.863	-.073	.679
2 自律的な機械は何かものごとを考えているように感じることがある	1.063	-.124	.995
3 自律的な機械は，できごとを覚えていると思う	.460	.018	.221
4 自律的な機械は，人間の考えていることを理解しているように感じることがある	.492	.204	.397
5 自律的な機械を捨てることを想像すると，かわいそうに感じる	-.158	.762	.468
6 自律的な機械は，何かをされて嫌だと感じているように思うことがある	.129	.789	.754
7 自律的な機械は，何かをされてうれしいと感じているように思うことがある	.110	.778	.714
8 自律的な機械は，まったく同じ型の機械でもそれぞれに個性があるように感じることがある（分析には使用せず）			
因子間相関	.567		

5) これらの変数が回答者の知識や経験によって影響を受けているかを確認した。表2に示した14個の変数を従属変数とする多変量分散分析を行ったところ，運転免許の有無 ($F(14, 91)=0.514, p=.919$)，AIに関する学習・業務経験の有無 ($F(14, 90)=0.870, p=.593$)，AI機器の利用経験の有無 ($F(14, 91)=1.525, p=.118$)は，いずれも従属変数と有意な関連を示さなかった。

表2 主要変数の平均値, 標準偏差, 相関係数

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
心の知覚傾向															
1 行為性	2.16	0.95													
2 経験性	2.15	1.06	.467 ***												
原因帰属															
3 AI	3.31	1.29	.060	.115											
4 メーカー	3.28	1.29	.172 †	.008	.597 ***										
5 ユーザー	2.86	1.27	-.023	.072	.245 *	.214 *									
6 被害者	4.28	0.81	-.113	-.236 *	-.110	-.130	-.154								
問題責任帰属															
7 AI	2.93	1.33	.083	.114	.565 ***	.377 ***	.095	-.123							
8 メーカー	3.50	1.19	.157	.039	.498 ***	.739 ***	.103	-.089	.454 ***						
9 ユーザー	3.33	1.37	-.100	.111	.145	.128	.660 ***	-.151	.069	.244 *					
10 被害者	3.92	1.05	-.179 †	-.292 **	-.036	-.101	-.066	.640 ***	-.051	-.019	-.105				
解決責任帰属															
11 AI	2.52	1.29	.145	.145	.445 ***	.335 ***	.166 †	-.196 *	.699 ***	.329 ***	.078	-.204 *			
12 メーカー	3.39	1.30	.124	-.009	.545 ***	.718 ***	.011	-.159	.423 ***	.759 ***	.028	-.128	.472 ***		
13 ユーザー	3.21	1.33	-.181 †	-.024	.166 †	.136	.632 ***	-.185 †	.013	.178 †	.723 ***	-.150	.146	.116	
14 誰にもない	1.93	1.28	-.080	.029	-.403 ***	-.517 ***	-.320 ***	.075	-.246 *	-.464 ***	-.373 ***	.066	-.187 †	-.407 ***	-.335 ***

各変数の測定はいずれも1-5の5件法; *** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

動車の利用状況や道路交通法に関する知識が交通事故の原因・責任の判断に、またAIに関する知識や利用経験がAIの開発者・利用者に対する態度に影響を与える可能性があるためである。

分析の結果、行為性の知覚傾向が高い人ほどメーカーに事故の原因、問題責任、解決責任を帰属したが（順に $\beta = .221$, $p = .009$; $\beta = .199$, $p = .041$; $\beta = .181$, $p = .048$ ）、問題責任と解決責任の帰属への効果はメーカーへの原因帰属を統制すると消失した（順に $\beta = .048$, $p = .551$; $\beta = .062$, $p = .448$ ）。一方でユーザーに対する判断では、行為性の知覚傾向がユーザーへの原因帰属とは関連せず（ $\beta = -.071$, $p = .488$ ）、解決責任の帰属を直接抑制した（ $\beta = -.205$, $p = .021$ ）。

AIへの原因帰属は、メーカーとユーザーへの原因帰属を促進した（順に $\beta = .609$, $p < .001$; $\beta = .287$, $p = .002$ ）。メーカーへの問題責任の帰属も促進されたが（ $\beta = .495$, $p < .001$ ）、この効果はメーカーへの原因帰属を統制すると消失した（ $\beta = .080$, $p = .383$ ）。最後に、解決責任の帰属については、メーカー、ユーザーともにAIへの原因帰属が解決責任の帰属を促進したが（順に $\beta = .550$, $p < .001$; $\beta = .220$, $p = .021$ ）、メーカーに対しては原因帰属を統制してもAIへの原因帰属が解決責任の帰属に直接影響した一方で（ $\beta = .223$, $p = .018$ ）、ユーザーに対しては原因帰属を統制すると効果が消失した（ $\beta = .056$, $p = .494$ ）。

責任帰属を規定する認知過程

上記の重回帰分析の結果から、AIへの原因帰属がメーカー・ユーザーへの原因帰属を促進し、それによって責任帰属が促進されるという媒介過程があることが推察される。この媒介過程を確認するために、問題責任と解決責任のそれぞれについてパス解析を行った。

あわせて、メーカーとユーザーに対する責任帰属のトレードオフについても探索的に検討した。行為者が原因を統制可能であったという認知は責任帰属を促進する要因の一つであるが（Weiner, 2006 速水・唐沢監訳 2007）、他の主体（たとえばメーカーにとつてのユーザー）に原因帰属することは、行為者（メーカー）にとつて統制不可能な原因を認めることを意味する。そこで、メーカーとユーザーへの原因帰属がそれぞれ他方への責任帰属にも影響を与えるというモデル（電子付録2）を作成し、モデルの適合度と各パス係数の値を確認しながらモデルを改善するという手順で分析を行った。

まず、問題責任の帰属については図1左の結果が得られた。AIへの原因帰属はメーカーへの問題責任の帰属

表3 メーカーに対する原因帰属，責任帰属の規定因

	メーカー 原因	メーカー問題責任		メーカー解決責任	
		Model 1	Model 2	Model 1	Model 2
性別（0＝女性；1＝男性）	.190 *	.132	.003	.191 *	.089
運転免許（0＝なし；1＝あり）	-.024	-.008	.008	-.039	-.026
AIに関する学習，業務経験（0＝なし；1＝あり）	.149 *	.085	-.016	.133	.053
AI機器の利用経験（0＝なし；1＝あり）	.098	-.017	-.083	.023	-.029
行為性の知覚傾向	.221 **	.199 *	.048	.181 *	.062
経験性の知覚傾向	-.103	-.065	.005	-.100	-.045
AIへの原因帰属	.609 ***	.495 ***	.080	.550 ***	.223 *
メーカーへの原因帰属			.680 ***		.537 ***
決定係数	.493 ***	.316 ***	.551 ***	.396 ***	.542 ***
自由度調整済み決定係数	.456	.266	.513	.352	.504
決定係数の増分			.235 ***		.146 ***

数値は標準化偏回帰係数；N=104；VIFs<1.98；*** p <.001，** p <.01，* p <.05

表4 ユーザーに対する原因帰属，責任帰属の規定因

	ユーザー 原因	ユーザー問題責任		ユーザー解決責任	
		Model 1	Model 2	Model 1	Model 2
性別（0＝女性；1＝男性）	-.242 *	-.178 †	-.005	-.142	-.004
運転免許（0＝なし；1＝あり）	-.104	-.073	.001	-.120	-.061
AIに関する学習，業務経験（0＝なし；1＝あり）	.100	.156	.085	.103	.046
AI機器の利用経験（0＝なし；1＝あり）	.259 **	.070	-.115	.223 *	.075
行為性の知覚傾向	-.071	-.180	-.129	-.245 *	-.205 *
経験性の知覚傾向	-.002	.155	.157 †	.018	.019
AIへの原因帰属	.287 **	.114	-.091	.220 *	.056
ユーザーへの原因帰属			.714 ***		.570 ***
決定係数	.223 ***	.119 †	.514 ***	.188 **	.440 ***
自由度調整済み決定係数	.167	.054	.473	.129	.393
決定係数の増分			.396 ***		.252 ***

数値は標準化偏回帰係数；N=104；VIFs<1.36；*** p <.001，** p <.01，* p <.05，† p <.10

につながっており（総合効果 $\beta=.504$, $p<.001$ ），その影響はメーカーへの原因帰属によって媒介されていた（媒介効果 $\beta=.408$, $p<.001$ ；直接効果 $\beta=.095$, $p=.292$ ）。一方，ユーザーに対する判断では，AIへの原因帰属がユーザーへの原因帰属を促進し，それによってユーザーへの問題責任の帰属が促進されるという媒介効果があったが（ $\beta=.166$, $p=.015$ ），AIへの原因帰属とユーザーへの問題責任の帰属との間に直接の関連は確認できなかった。

次に，解決責任の帰属については図1右の結果が得られた。AIへの原因帰属はメーカーへの解決責任の帰属

につながっており（総合効果 $\beta=.478$, $p<.001$ ），その影響はメーカーへの原因帰属（媒介効果 $\beta=.374$, $p<.001$ ），ユーザーへの原因帰属（媒介効果 $\beta=-.104$, $p=.002$ ）によって媒介されていた。そして，媒介変数の効果を統制してもなお，AIへの原因帰属がメーカーへの解決責任の帰属を促進していた（ $\beta=.208$, $p=.036$ ）。一方，ユーザーに対する判断では，AIへの原因帰属がユーザーへの原因帰属を促進し，それによってユーザーへの解決責任の帰属が促進されるという媒介効果があったが（ $\beta=.155$, $p=.015$ ），AIへの原因帰属とユーザーへの解決責任の帰属との間に直接の関連は確認できなかった。

図1 問題責任, 解決責任の帰属の認知過程

考 察

メーカー, ユーザーに対する原因帰属

本研究では, 事故の原因を AI に帰属するほど, メーカーとユーザーに対する原因帰属も高くなるという結果が得られた。つまり, AI に出来事の原因があることと, その AI を開発・利用した人間に原因があることはトレ

ドオフの関係ではなく, 開発者や利用者も AI の動作について原因の一端を担う存在と認知されていた。AI が人間の指示や操作なしに動作する自律性を備えているとしても, 人は AI を人間から独立した行為主体と認知するわけではなく, あくまで人間に利用される道具として認知していると考えられる。そして個別の事故の場面で自動車の制御に関与していなかったとしても, 自動車

を開発・販売したことや、それを購入・利用したことも含めて事故の原因の所在が判断されたと考えられる。

メーカー、ユーザーに対する責任帰属

問題責任の帰属に関する分析では、メーカー、ユーザーともに原因帰属のみが統計的に有意な規定因となっていた。AI への原因帰属はメーカーへの問題責任帰属を促進したが、その効果もメーカーへの原因帰属に完全に媒介されており、直接効果は見られなかった。つまり、それぞれの主体に原因があるほど責任もあると判断されており、社会心理学のさまざまな研究で示されてきた責任の判断過程（膳場他，2018）と整合する結果であった。

解決責任の帰属に関しては、メーカーへの原因帰属の影響を統制してもなお、AI への原因帰属がメーカーへの解決責任帰属を高める直接効果があった。一方でユーザーに対する判断では、問題責任の帰属と同様にユーザーへの原因帰属のみが解決責任の帰属を促進し、AI への原因帰属との関連はなかった。つまり、メーカーは事故の原因の所在とは別に、AI の開発者という立場ゆえに解決責任があると判断されるが、ユーザーは自身が事故の原因となった程度に応じて解決責任もあると判断されるという違いがあった。また、ユーザーへの原因帰属はメーカーへの解決責任帰属を抑制するが、メーカーへの原因帰属はユーザーへの解決責任帰属と関連しないという結果からも、メーカーは他のステークホルダーも含めた出来事全体の状況を勘案して解決責任を判断されるのに対して、ユーザーの解決責任はユーザー自身の出来事への関与のみに基づいて判断されていると考えられる。

なお、問題責任の帰属と解決責任の帰属は強い正の相関関係があり（メーカー： $r=.759$, $p<.001$ ；ユーザー： $r=.723$, $p<.001$ ；表 2）、責任の 2 つの側面は独立に判断されているわけではないと考えられる。単に回答者が 2 つの側面を混同しているという可能性もあるが、パス解析では 2 つの側面が他の変数との関連において異なるパターンを示しており、両者は異なる概念として認知されていると考えることに無理はないだろう。この点については、両概念間の関連（問題責任の大きさが解決責任を帰属する際の判断材料になる等）を精緻に分析するとともに、他の概念との関連も幅広く確認し（非難・罰の判断は問題責任の帰属のみと関連する等）、結果の頑健性を確認することも必要だろう。

機械に対する心の知覚傾向と原因帰属、責任帰属

本研究では、自律的な機械に対して心を知覚する傾向を、行為性と経験性の 2 つの側面で測定した。このうち行為性の知覚傾向は、メーカーへの事故の原因帰属を促進し、ユーザーへの解決責任の帰属を抑制した。

行為性とは自ら行為を生み出すことに関わる心の側面であり、このような心の機能が機械にあると知覚することは、その機械を人間の意図から独立して判断や行動を選択できる存在と知覚することにつながると考えられる。そのため、事故の場面での自動運転車の動作だけに注目すれば、その動作に対するメーカーの統制可能性は低く、原因帰属は抑制されるという予測が可能であるが、結果はその想定に反するものであった。事故の場面での行動だけでなく AI を開発、利用したことも事故の原因と判断されたと考えられることは既に述べたが、行為性の知覚傾向と原因帰属の関連はこの解釈を補強するものといえる。つまり、統制が困難な AI を作ったメーカーにも事故の原因があるのだという形で、事故が起きた場面よりも過去にさかのぼって原因の所在が判断されたのだろう。

一方でユーザーに対する判断では、行為性の知覚傾向と原因帰属は関連しなかった。メーカーの場合と異なり、AI が行為性を知覚させる存在であることがユーザーの判断や行動によるものとは考えにくい。行為性を高く知覚したとしてもユーザーへの原因帰属には影響しなかったと考えられる。行為性の知覚傾向が解決責任の帰属を抑制した点については、原因や問題責任が帰属の対象ごとに個別に判断されるのに対して、解決責任は一定量の負担を分配する形で判断されるという違いがその背後にあるのかもしれない。ただし、行為性の知覚はメーカーに対する判断には直接影響していないため、負担の分配という観点だけでは十分な説明は得られておらず、解決責任の大きさがどのように判断されているかという点は今後の検討課題である。

本研究の意義と今後の課題

本研究は自動運転車による交通事故に関する原因帰属、責任帰属の認知過程を検討した。現実にはまだ普及していない技術を題材とした調査であるため、実際にそのような出来事が起きたときの人々の反応を検討したわけではないという限界はあるものの、AI 開発や法制度の専門家ではない人々の反応に関する実証的な知見は、先端技術の導入を円滑に進めるための制度設計を考えるうえで有益な情報となるだろう。

本研究の特徴として、問題責任と解決責任の概念を区別し、両者を同時に検討の対象としたことが挙げられる。

責任という語は日常的にはさまざまな意味で用いられるが、機械が人間の意図を介さずに複雑な判断や行動を行う状況では、関係者がどのような意味での責任を負うのかを区別することが重要になる。本研究の結果から、特にメーカーの責任を考える際には、人々がメーカーに帰属する「責任」の内容を明確に区別することが重要になると言えよう。メーカーも完全には予測、制御できないAIの動作についてメーカーに「責任」が帰属されるとすれば、AI開発を委縮させてしまうかもしれない。しかし、ここで帰属される責任は問題の解決、つまり被害が回復されることに対する責任であって、メーカーへの非難や罰を求めるものではないということが明確になれば、保険による補償のように、「誰が悪いのか」を問うことなく被害者への補償を行う仕組みが社会に受容されることが期待できる。

最後に本研究の限界について述べる。第一に、「方法」で述べた通り、本研究では便宜抽出で調査を行っており、その結果を母集団に一般化することには慎重でなければならない。特に本研究では、回答者の大半は都市部に住む学生であり、自動車を運転する機会が少ない人々であった。運転する機会の有無によって自分自身が交通事故の加害者となる可能性、すなわち状況的類似性(Shaver, 1970)が変わるため、事故の原因や責任に関する判断にも影響を与える可能性がある。本研究の分析結果では運転免許の有無が他の変数に影響を与えることは確認されず、運転に関する知識や経験の量が判断に影響すると示されたわけではないが、通勤で毎日利用するような高頻度の利用者の態度や、自動車が生活必需品である地域とそうでない地域の差など、自動車の利用形態に関わる要因は無数にあり得る。これらの他にも年齢や職業など、今後はより幅広い属性の人々を対象とした調査を行い、研究結果の一般化可能性を高める必要がある。

第二に、本研究は架空の事故のシナリオに基づく判断の検討に留まっており、現実にAIが普及した社会での人々の態度を反映していない可能性がある。また、技術開発の進展によっても実際に普及するAIの機能や利用法は変わり得る。そのため、今後の技術開発の動向を踏まえながら、さまざまな場面を想定した研究を行うことで、より妥当な知見を蓄積していく必要がある。

第三に、AIに対する道徳判断の認知過程を分析するうえで、心の知覚という概念の有効性を慎重に検討する必要がある。本研究では、先行研究で個別のAIに対する心の知覚の評定が非常に低い値に偏っていたことを考慮し、自律的な機械全般への知覚傾向に焦点を当てた。しかし本研究の尺度でも、平均値が5件法で2.2程度と

低い値に留まっていた(表2)。この尺度得点の低さは、機械全般に対して当てはまる知覚傾向があるわけではなく、心の知覚という概念を当てはめて分析することが有効な対象が限定的であることを示唆する。

心の知覚という枠組みを提唱したGray et al. (2007)では、目や口といった擬人化された外見をもつロボットを知覚対象として提示していた。この外見が心の機能を知覚させる認知的な手掛かりとなっていたと考えられる。一方でTakahashi et al. (2016)では、コンピュータに高い行為性があると知覚されていた。人間らしい外見はないが、情報処理というコンピュータの本質的な機能に注意が向くことで、その機能を人間の思考になぞらえる知覚が促された可能性がある。

これらに対して本研究では、心の知覚傾向の測定の直前に掃除ロボットなどの機器の利用経験を尋ねていたため、回答者はこれらの具体例を念頭に置いて回答した可能性が高い。これらの機器はコンピュータと違って物理的な形状や動作を想像でき、それが人間を模したものであるため、人間との差異を意識しやすかったのかもしれない。今後の研究では、心の知覚と道徳判断の関連という分析の枠組みが適用できる範囲を見極め、自動運転車という具体的な場面における帰属判断をより適切に説明できる要因を明らかにすることが課題となる。

引用文献

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563, 59–64.
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352, 1573–1576.
- Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34.
- Brickman, P., Rabinowitz, V. C., Karuza, J. Jr., Coates, D., Cohn, E., & Kidder, L. (1982). Models of helping and coping. *American Psychologist*, 37, 368–384.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144, 114–126.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64, 1155–1170.

- 藤田友敬 (2016). 自動運転と運行供用者の責任 ジュリスト, 1501, 23-29.
- 後藤 元 (2018). 自動運転車と民事責任 弥永真生・宍戸常寿 (編) ロボット・AI と法 (pp. 167-186) 有斐閣
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315, 619.
- Gray, K., & Wegner, D. M. (2012). Morality takes two: Dyadic morality and mind perception. In M. Mikulincer, & P. R. Shaver (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (pp. 109-127). Washington, DC: American Psychological Association.
- 橋本剛明・白岩祐子・唐沢かおり (2012). 経済格差の是正政策に対する人々の賛意—機会の平等性と社会階層の認知が責任帰属に与える影響の検討— 社会心理学研究, 28, 13-23.
- 池田裕輔 (2016). 自動運転技術等の現況 ジュリスト, 1501, 16-22.
- 池内裕美 (2010). 成人のアニミズム的思考—自発的喪失としてのモノ供養の心理— 社会心理学研究, 25, 167-177.
- Karasawa, K. (2002). Responsibility attribution and support provision in an intergroup context: An examination in a simulated society game. *Japanese Psychological Research*, 44, 196-208.
- 小林正啓 (2017). 自動運転車の実現に向けた法制度上の課題 情報管理, 60, 240-250.
- 高度情報通信ネットワーク社会推進戦略本部・官民データ活用推進戦略会議 (2020). 官民 ITS 構想・ロードマップ 2020 Retrieved from https://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20200715/2020_roadmap.pdf (2021 年 3 月 10 日)
- 松尾 豊 (2015). 人工知能は人間を超えるか—ディープラーニングの先にあるもの— KADOKAWA
- McManus, R. M., & Rutchick, A. M. (2019). Autonomous vehicles and the attribution of moral responsibility. *Social Psychological and Personality Science*, 10, 345-352.
- 中越みずき・稲増一憲 (2019). メディアフレームと情報の立場性が生活保護の責任帰属に及ぼす影響—「責任がある」のは政府か受給者か— 社会心理学研究, 35, 72-84.
- 西垣 通・河島茂生 (2019). AI 倫理—人工知能は「責任をとれるのか— 中央公論新社
- 大屋雄裕 (2018). ロボット・AI と自己決定する個人 弥永真生・宍戸常寿 (編) ロボット・AI と法 (pp. 59-75) 有斐閣
- Shank, D. B., & DeSanti, A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations. *Computers in Human Behavior*, 86, 401-411.
- Shaver, K. G. (1970). Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology*, 14, 101-113.
- Takahashi, H., Ban, M., & Asada, M. (2016). Semantic differential scale method can reveal multi-dimensional aspects of mind perception. *Frontiers in Psychology*, 7, 1717.
- Weiner, B. (2006). *Social Motivation, Justice, and the Moral Emotions: An Attributional Approach*. New Jersey: Lawrence Erlbaum Associates. (ワイナー, B. 速水敏彦・唐沢かおり (監訳) (2007). 社会的動機づけの心理学—他者を裁く心と道徳的感情— 北大路書房)
- 横井良典・中谷内一也 (2018). 治療方針の共有が人工知能への信頼に及ぼす影響 社会心理学研究, 34, 16-25.
- Yokoi, R., & Nakayachi, K. (2019). The effect of shared investing strategy on trust in artificial intelligence. *Japanese Journal of Experimental Social Psychology*, 59, 46-50.
- 膳場百合子・唐沢 穰・後藤伸彦 (2018). 社会心理学における責任判断研究 唐沢 穰・松村良之・奥田太郎 (編) 責任と法意識の人間科学 (pp. 3-35) 勁草書房

The attribution of responsibility to the manufacturer and/or user after an accident caused by a self-driving car

TETSUSHI TANIBE (*Graduate School of Humanities and Sociology, The University of Tokyo*)

KAORI KARASAWA (*Graduate School of Humanities and Sociology, The University of Tokyo*)

We examined how people attribute cause and responsibility for harm caused by artificial intelligence (AI). The participants read a vignette about a pedestrian fatality involving a self-driving car and allocated cause and responsibility to the car manufacturer and/or the user. We found that those who attributed cause to the AI also attributed cause to the manufacturer and user, and those who perceived autonomous machines to have mental capacities such as intention attributed cause to the manufacturer. These results indicate that people do not perceive AI as an agent that acts independently from humans, even though they think that AI is autonomous. We also found that those who attributed cause to the manufacturer and user also attributed responsibility to them. Moreover, those who attributed cause to the AI believed that the manufacturer should solve the problem. Thus, people presumed that the manufacturer was responsible for solving the problem, even though it was not responsible for causing the problem. Taken together, this study illustrates the importance of examining people's attitudes when developing rules relating to AI.

Key Words: self-driving car/autonomous car, artificial intelligence, anthropomorphism, mind perception, responsibility attribution