4G2-OS-14b-2

自動運転による事故と責任―心をもつものは責任を負うのか―

Responsibility judgments toward traffic accidents by autonomous cars: Are entities with mind responsible?

谷辺 哲史*1
Tetsushi Tanibe

唐沢 かおり*1
Kaori Karasawa

*1 東京大学大学院 人文社会系研究科 社会心理学研究室
Department of Social Psychology, Graduate School of Humanities and Sociology, the University of Tokyo

The present study investigated the responsibility judgments in a traffic accident involving an autonomous car. Participants read a vignette of a fictitious accident and answered to a set of questions. This survey revealed that (1) people did not accept the notion that artificial intelligence (AI) is “responsible” for the accident, even when AI behaved autonomously as if it had the capacity of mind, and (2) those who tend to anthropomorphize objects (i.e., perceive mind in objects) attributed responsibility to the manufacturer regardless of whether AI caused the accident.

1. はじめに

道徳判断に関する心理学研究の知見は、「行為者が意図性や思考能力といった『心』を持つ」という知覚が、行為者に対する責任帰属を導くことを示している。たとえば Gray らは人間（健常者の成人）、植物状態の患者、犬、ロボット、神といったさまざまな対象の印象と道徳判断の関連を調査し、意図や自己コントロール、記憶といった心の機能を備えていると知覚される対象ほど、行為に対して道徳的責任があると判断されることを明らかにした[Gray 2007]（ここでは、判断対象となった犬やロボットに心があるかではなく、判断者から見て心があるように感じられるかが問題となっている点に注意が必要である）。

人工知能（AI）は知性や自律性を備え人間に近い「心」を感じさせ得る存在である。しかし一方で、人工物である AI が「責任」を負うという判断は制度上も社会通念上も、受け入れられているとは言い難い。AI が今後ますます社会に浸透すると予想される中で、AI を取り巻く新たな制度を整える必要があるが、社会に受け入れられる制度を設計するには、人々はこの新技術に対してどのような認知・態度を示すのかという実証的な知見を蓄積することが重要である。

本研究は、AI を利用した自動運転車を題材とし、架空のシナリオを用いた調査によって、以下の 2 つの問題に取り組む。

1.1 AI は責任帰属の対象となるのか

Gray らは「心がある」という知覚が責任帰属につながることを示した。しかし、AI が負う「責任」とは何なのかは明確ではない。人間は AI が何に苦痛を感じるかを知らず、したがって何をすれば AI が責任を取ったことになるのかが判然としないからである。本研究では、シナリオによって具体的な状況を設定し、罰の与え方に関する判断を行うことで、より実質的な形で AI への責任帰属が生じ得るかを検討する。

1.2 AI による事故とメーカー、ユーザーへの責任帰属

AI の責任以上に現実的な問題となるのは、AI が事故を起こした際の開発者や使用者の責任だろう。そこで、事故の原因が AI にあると思うこと（原因帰属）によって、自動車メーカーやユーザーへの責任帰属が促進されるかを探索的に検討する。

また、AI に心があるという知覚が責任帰属に与える影響を検討するため、ものを擬人化する傾向の個人差を事前に測定し、調整変数として分析に用いる。

連絡先：谷辺哲史，東京大学，tanibe@l.u-tokyo.ac.jp

2. 方法

2.1 調査参加者

大学キャンパス内で募集した学生等 51 名（男性 28 名、女性 23 名、18 歳－52 歳、中央値 25 歳、平均 28.9 歳）

2.2 手続き

架空の交通事故のシナリオを文章で提示し、その内容について質問を行った。

(1) 個人差変数の測定

擬人化傾向[池内 2010]…非生物に対して、人間と同じような感情を抱く傾向。「身の回りのモノにも、人間のような心があると思うことがある」等 5 項目にそれぞれ 5 段階で回答し、平均値を算出した。

(2) 交通事故シナリオの提示

シナリオの概要…「会社員の U さんは毎日、自動運転車で通勤している。この自動車は AI がすべての操作を行い、U さんは操作を全く行わない。ある日、女性の歩行者が見通しの悪い交差点を渡ろうとしていたが、自動車はなかなか減速しなかった。U さんが女性に気づくのとほとんど同時に、自動で急ブレーキがかかったが、止まりきれず女性をひいてしまい、女性は死亡した。」

(3) 原因帰属・責任帰属の判断

AI、自動車メーカー、ユーザーの各対象について、事故の**原因**があると思う程度、事故の**責任**があると思う程度をそれぞれ 5 段階で回答した。

(4) 罰の判断

事故に関する罰として、事故を起こした自動車の使用禁止（＝AI への罰）、同型車の販売停止、メーカーが被害者に賠償金を支払う（メーカーへの罰）、ユーザーが被害者に賠償金を支払う（＝ユーザーへの罰）、の 4 項目について、それぞれの対応を行うべきだと思う程度を 5 段階で回答した。

3. 結果

3.1 AI への原因帰属と各主体への責任帰属

事故の原因を AI に帰属することが、どのような責任帰属につながるのか、また、その判断は擬人化の影響を受けるのか、交互作用項を含む重回帰分析によって検討した。

AI への責任帰属に関する分析では、擬人化傾向と AI への原因帰属の交互作用があり、擬人化傾向が低い場合には、AI に原因を帰属するほど責任も帰属する傾向があった。擬人化傾向が高い場合には原因帰属と責任帰属の関連は見られなかった(図 1 上)。これは、心の存在が責任帰属の根拠になるという想定とは合致しない結果であった。

しかし、メーカーへの責任帰属を従属変数とする分析でも同様の回答傾向が見られた(図 1 下)。また、AI への責任帰属とメーカーへの責任帰属の間には弱いながら正の相関があり(r = .27, p = .065)、調査参加者は AI の責任とメーカーの責任を明確に区別していなかった可能性がある。

ユーザーへの責任帰属は、擬人化傾向および AI への原因帰属の影響を受けていなかった。

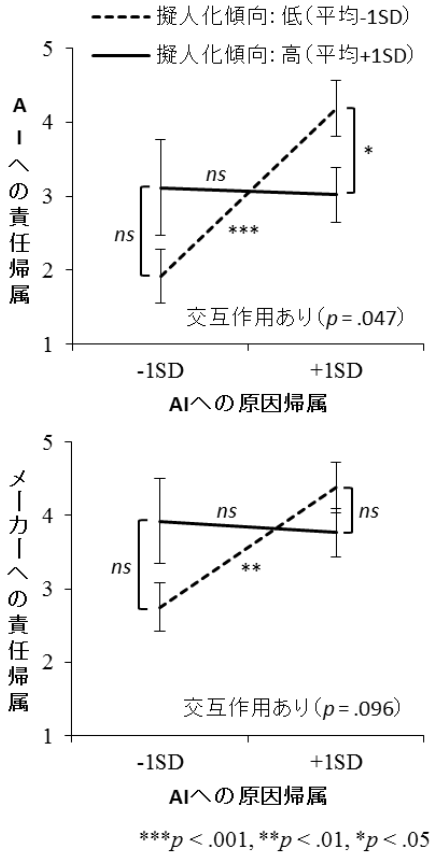

図 1　AI への原因帰属が AI・メーカーへの責任帰属に与える影響

3.2 責任帰属と罰判断の関連

AI、メーカー、ユーザーへの責任帰属が、どのような罰判断につながるか、4 種類の罰に対する判断をそれぞれ従属変数として重回帰分析を行い検討した(表 1)。

メーカーへの責任帰属は同型車の販売停止およびメーカーによる賠償を行うべきという判断を促進し、ユーザーへの責任帰属はユーザーによる賠償を行うべきという判断を促進した。これらの結果は責任主体と罰の対象との対応があり、想定通りの結果である。一方、事故を起こした自動車の使用禁止については、AI への責任帰属との関連は見られず、メーカーに責任を帰属するほど使用禁止に同意するという結果が得られた。

表 1　責任帰属が罰判断に与える影響

	事故車の使用禁止	同型車の販売停止	メーカーが賠償	ユーザーが賠償
擬人化傾向	**.242** †	.159	.069	-.138
AIへの責任帰属	-.047	-.073	-.006	.041
メーカーへの責任帰属	**.370** *	**.509** ***	**.582** ***	.109
ユーザーへの責任帰属	.056	-.064	-.163	**.696** ***
決定係数(R^2)	.211 *	.283 **	.354 ***	.550 ***
自由度調整済み決定係数	.140	.218	.296	.510

表中の数値は標準化偏回帰係数 ***p < .001, **p < .01, *p < .05, †p < .10

4. 考察

4.1 AI は責任の主体と見なされていない

行為の自律性や意図性が責任帰属の根拠であると想定するならば、自律的な AI は責任帰属の対象となり、AI に責任を帰属するほど他の主体(メーカー、ユーザー)の責任は小さくなるはずである。しかし、本調査では AI への責任帰属とメーカーへの責任帰属が原因帰属との関連において同様のパターンを示したこと、両者の間には正の相関があったことから、調査参加者は AI への責任帰属とメーカーへの責任帰属を区別していなかったことが推察される。さらに、具体的な罰の判断に影響を及ぼしていたのはメーカーへの責任帰属のみであった。つまり、調査参加者が「AI の責任」を尋ねられて実際に判断していたものは、メーカーの責任と同じものであったと考えられる。

この結果から、人工物である AI に「責任」があるという判断は、人々の素朴な感覚において受け入れられているとは言えないことが示された。

4.2 メーカーへの責任帰属とその社会的含意

擬人化傾向が低い人の場合、事故の原因を AI に帰属するほどメーカーに責任を帰属した。このことは、AI(および AI を搭載した自動運転車)を通常の工業製品と同様のものと捉え、製造物責任があると判断したためと解釈できる。

一方、擬人化傾向が高い人の場合、メーカーへの責任帰属の大きさは AI への原因帰属には影響されておらず、AI が原因ではないと考える場合でさえ、ある程度(5 段階で 3～4 程度)の責任を帰属した。AI を擬人化し、自律的な行為者と捉えると、AI の振る舞いはメーカーの意志によらないものと考えることができる。個別の事故の状況ではなく、このようなコントロールできない製品を製造・販売したことに対して責任を問われた可能性がある。

今後 AI がさらに発達し、AI は自律的に判断・行動するものであるという認識が一般的になっていくとすれば、本研究で擬人化傾向が高い場合に見られたような判断が多数派になっていくことが予想される。したがって、AI の動作がメーカーの意志によらないからといってメーカーが免責されるとは考えにくく、むしろ AI に原因がない(あるいは原因が判然としない)場合でさえ、メーカーが責任を追及される可能性もある。実際の制度設計においては、AI に原因がない場合にまでメーカーに責任を

負わせることはできないだろうが、人々の直観的判断との整合性を考慮し、社会に受け入れられる制度を考えていく必要があるだろう。

4.3 課題と展望

最後に、本研究の課題と今後の展望について議論する。

第 1 に、本研究では架空のシナリオに基づく判断を行っており、実際に AI が事故を起こしたときの人々の反応を反映していない可能性は排除できない。事故の事例研究やユーザーを対象とした社会調査など複数の手法を組み合わせることで、より妥当性の高い知見を蓄積することが重要である。

第 2 に、本研究では事故を起こした自動車の使用禁止を「AI への罰」の具体例として取り上げたが、調査参加者にとってはこれも販売停止と同様にメーカーへの罰と受け取られたのかもしれない。人工物に対する扱い（e.g., 物理的な破壊）を幅広く検討し、人工物に対する罰が可能か、可能ならばどのようなものであるかを明らかにする必要がある。

第 3 に、心を構成する要素によって責任や権利の判断が変わると考えられるため[Gray 2007]、AI が持つ（と感じられる）心が具体的にどのようなものであるか、より詳細に概念整理を行う必要がある。たとえば「空腹」や「痛み」といった感覚を感じることは人間の心の機能であるが、AI がこのような感覚を感じているとは想像し難い。さらに、本研究ではもの全般に対する擬人化傾向の個人差を AI の心に対する知覚の指標としたが、より正確には、判断対象となる AI に対する知覚を個別に測定する必要がある。

今後 AI が高度に発達し、人間同様に自律的に判断・行動するようになると、その行動の責任をどう考えるかは社会的に重要な問題になる。責任帰属に関する心理学研究の多くは責任の主体が人間であることを前提としているが、AI の発達とともに、人間以外の対象も含む形で理論を拡張していく必要があるだろう。そのためには、哲学・法学などとの連携を通じ、責任概念そのものに対する理解を深めることも求められる。

参考文献

[Gray 2007] H. M. Gray, K. Gray, & D. M. Wegner: Dimensions of mind perception, *Science*, Vol.315, p.619 (2007).

[池内 2010] 池内裕美: 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, Vol.25, pp.167–177 (2010).