資料３－１

中央大学

AI-NW研究開発8原則
と
ロボット工学3原則

開発原則分科会（第一回）

Nov. 8, 2016

中央大学

総合政策学部教授・大学院総合政策研究科委員長

平野 晋

AI-NW研究開発8原則

① **透明性の原則**

ＡＩネットワークシステムの動作の検証可能性及び説明可能性を確保すること。

② **利用者支援の原則**

ＡＩネットワークシステムが利用者を支援し、利用者に選択の機会を適切に提供するよう配慮すること。

③ **制御可能性の原則**

人間によるＡＩネットワークシステムの制御可能性を確保すること。

④ **セキュリティ確保の原則**

ＡＩネットワークシステムの頑健性及び信頼性を確保すること。

⑤ **安全保護の原則**

ＡＩネットワークシステムが利用者及び第三者の生命・身体の安全に危害を及ぼさないよう配慮すること。

⑥ **プライバシー保護の原則**

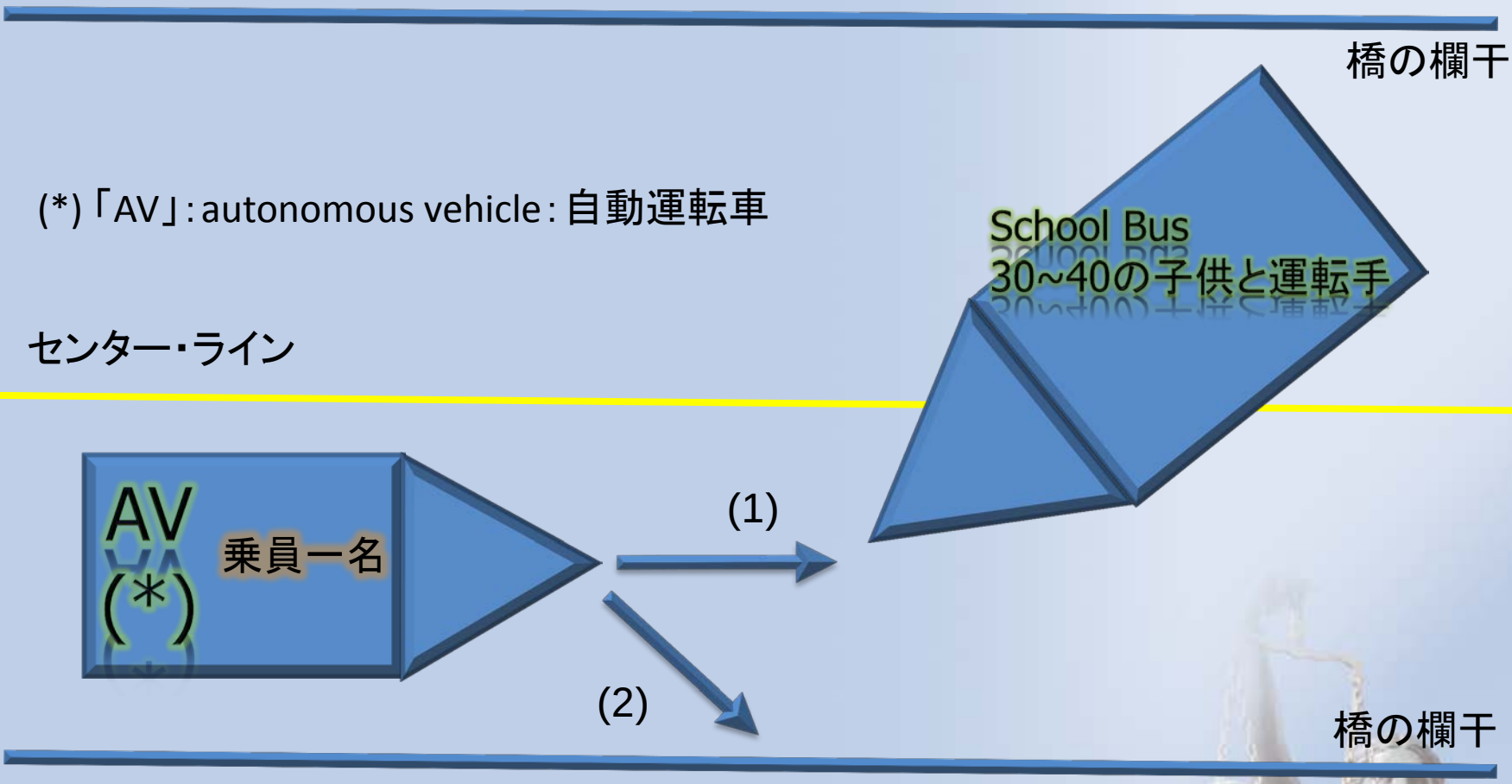
ＡＩネットワークシステムが利用者及び第三者のプライバシーを侵害しないように配慮すること。

⑦ **倫理の原則**

ＡＩネットワークシステムの研究開発において、人間の尊厳と個人の自律を尊重すること。

⑧ **アカウンタビリティの原則**

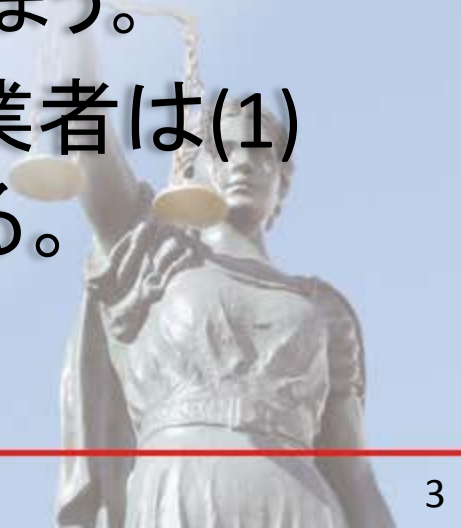
ＡＩネットワークシステムの研究開発者が利用者など関係するステークホルダーに対しアカウンタビリティを果たすこと。



Drawn by Hirano based upon hypos. *in* Clive Thompson, *Relying on Algorithms and Bots Can Be Really, Really Dangerous*, WIRED, Mar. 25, 2013, *available at* https://www.wired.com/2013/03/clive-thompson-2104/ (last visited Oct. 25, 2016) (originally in Gary Marcus, *Moral Machines*, New Yorker Blogs, No. 27, 2012, *available at* http://www.newyorker.com/news/news-desk/moral-machines (last visited Oct. 25, 2016)); Jeffrey K. Gurney, *Crashing into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183, 261 (2015-2016).

「橋問題」

- スクール・バスが突然、自動運転車(AV)の前に飛び出して来て、AVには以下の２つしか選択肢が残されていない。
  - (1) 直進して30～40名の子供とバス運転手を死なせてしまうか; 又は
  - (2) 右折してAVの乗員一名が死んでしまう。
- 功利主義者は(2)を勧めても、製造業者は(1)を選びがちであろう、と云われている。



AI-NW研究開発8原則の当てはめ

- ① 透明性の原則
  - 製造業者は、スクール・バスの乗員達や同じような立場の相手方を犠牲にして迄も、常にAVの乗員を保護するような衝突進路をAVが選択するように、隠れてAIを（設計時に）操作するおそれが指摘されている。（しかも複雑なAIが何故そのような選択をしたのかは明らかにされない！）
- ② 利用者支援の原則 (利用者の選択を尊重)
  - AV乗員の意向を問わずに勝手に選択肢 (1) を（変更不可能なデフォールト：初期値として）製造業者が設定すれば、乗員の選択意向に反するかもしれない。（選択肢(2)を好む乗員も居るかもしれない。）
- ⑦ 倫理の原則(人間の尊厳の尊重)
  - 選択肢(1)を選択することは倫理的に正しいか?
- ⑧ アカウンタビリティーの原則
  - 選択肢(1) は社会に受容され得るか？選択肢(1) は「責任ある行動」と云えるか？

未解決な問題例

- ⑤ 安全保護の原則
  - 「利用者及び第三者の生命・身体の安全に危害を及ぼさないよう配慮する」ように要求されている。(強調付加)
  - しかし選択肢(1)は第三者の生命・身体に危害を及ぼし、他方の選択肢(2)も利用者の生命・身体に危害を及ぼしてしまう。
  - 従って第⑤原則は、今のままでは、「橋問題」への回答になっていない。
- 同様な問題はアイザック・アシモフ著の『われはロボット』でも著わされている.
  - アイザック・アシモフ『堂々めぐり』(1942年).

アシモフのロボット工学3原則

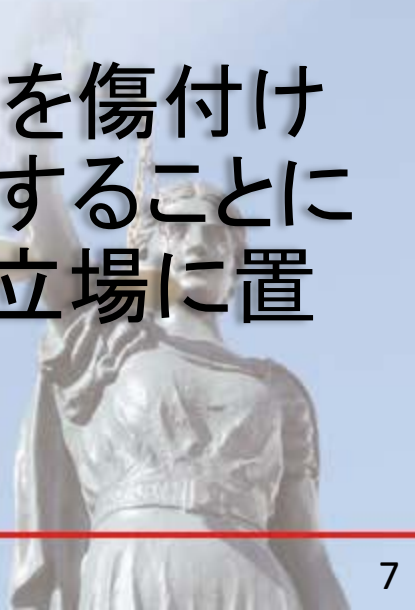
- **The 1st Law: A robot may not injure a human being or, through inaction, allow a human being to come to harm;**
  - **第一原則： ロボットは、ヒトを傷付けてはならず、又は、不作為によってヒトが危害に遭遇する事態を許してはならない。**
- **The 2nd Law: A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law; and**
  - **第二原則：ロボットは、ヒトが与えた如何なる命令にも従わねばならない。但し命令が第一原則に反する場合を除く。**
- **The 3rd Law: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.**
  - **第三原則：ロボットは、自らの存在を守らねばならない。但し、第一原則又は第二原則と抵触しない場合に限る。**

ISAAC ASIMOV, I, ROBOT (1950).

[The Zeroth Law: A robot may not harm humanity, or, by inaction, allow humanity to come to harm.]「第零原則： ロボットは、人類に危害を与えることが許されず、又は、不作為によって人類が危害に遭遇する事態を許してはならない。」 ISAAC ASIMOV, ROBOTS AND EMPIRE (1985).

ロボット工学第一原則を「橋問題」に当てはめると...

- 「橋問題」に遭遇したアシモフのロボットは、選択肢(1)又は(2)の選択を迫られる。
- しかし何れの選択肢を選んでも、多数のヒトを殺すか、又は一人を殺さねばならないように迫られることになる。
- 従って、何れの選択肢を選んでも、ヒトを傷付けてはならないと命じる第一原則に違反することになり、『堂々めぐり』のロボットと同様な立場に置かれてしまう。



規範を起案する難しさ

- 文言の多義性／曖昧性
  - アシモフの第一原則の「ヒトを傷付けてはなら」ないとは、身体・生命への物理的な危害のみならず、感情に対する危害も禁止されると読めるかもしれない。 アシモフ著『うそつき』(1941年)参照。
  - AI-NW研究開発の第⑦「倫理の原則」の「倫理」には、「功利主義」も「義務論」も含まれる。 しかし「功利主義」と「義務論」とはしばしば対立し得る。

- 複数の原則相互が牴触するおそれ
  - 例えばAI-NW研究開発の第⑤「安全保護の原則」と、第⑥「プリイバシー保護の原則」とは、トレードオフな関係になり得る。 ユージーン・ヴォロック「不法行為法対プライバシー」*in*『コロンビア大学ロー・レヴュー』114巻879頁 (2014年)参照。

「制御可能性」と「追跡究明可能性」の重要性

- AIやロボットに特有なリスクは、主に以下であるように思われる:

  (1) 制御不可能性 (un-controllability);

  及び

  (2) 追跡究明不可能性 (un-traceability)。

「制御不可能性」と責任

- 制御不可能性に関して:
  - AIやロボットの民事賠償責任論議に於いては、AIやロボットの制御不可能な特性ゆえに、結果的に生じ得る事故や危険を設計者、プログラマー、及び製造業者でさえも**予見不可能**（**un-foreseeable**）と解釈する指摘が顕著に見受けられる。
  - 予見不可能性は、「過失」や「近因」と呼ばれる因果関係の不存在を意味するので、設計者、プログラマー、及び製造業者に責任を課せなくなり、「**責任空白**」（“**liability gap**”）が生じるという指摘もあり、これがAIやロボットの責任論議に於いて重大な問題点の一つとされている。
  - 制御不可能な特性に由来する予見不可能性の原因は、AIやロボットの「自律的」（**autonomous**）又は「創発的な行動」（**emergence behaviors**）によるとも指摘されている。

「追跡究明不可能性」と責任

- 追跡究明不可能性も製品安全や責任論議に於いて重大な懸念の一つ:
  - アルファ碁が世界チャンピオンとの5連戦に於いて唯一敗れた一戦の際に打った手(誤作動?)が開発者にも理解できないとされている。
  - しかし敗因(何故そのような手を打ったのかという誤作動の理由)が解らなければ、設計者、プログラマー、及び製造業者も改善しようがない。(従って、製品安全を向上させていく為にも、原因を追跡究明できるように仕組んでおくことが重要。)
  - 更に、責任を特定する際にも追跡究明可能性は重要であり、例えばToyotaの2005年型Camry車の「意図せぬ急加速」事件に於いて裁判所は、不具合等の存在・不存在が追跡究明不可能な仕組みであったことを責めている。
    - *See In re Toyota Motor Corp. Unintended Acceleration*, 978 F.Supp.2d 1053 (C.D. Cal. 2013) (*See* the next slide.).

追跡究明可能なソフトの使用が奨励されたPL訴訟例 (1/2)

急加速事故の原因が、運転者の踏み間違えか又は欠陥が原因かを追跡究明できる装置を組み込んでおかなかったToyota社が、裁判所から以下のように云われてしまった:

> Toyota's software does not record software failure. . . . . "To rule that this prevented [the plaintiff] from establishing a prima facie case would be to insulate manufacturers from liability for defective products in any case where the defect causes its own destruction. Such a result would be totally untenable [支持できない]."

Toyota Unintended Acceleration, 978 F.Supp.2d at 1097 (*quoting Firestone Tire & Rubber Co. v. King*, 244 S.E.2d 905 (Ga. App. 1978))(emphasis added). 当事件の概要は、今月中旬発行予定の拙稿「*Toyota Motor Corp. Unintended Acceleration* ～AI・ロボット・自動運転時代の「誤作動法理」適用を示唆する事例～」*in*『国際商事法務』44巻11号__頁 (2016年11月)参照。

〈次スライドに続く.〉

追跡究明可能なソフトの使用が奨励されたPL訴訟例 (2/2)

「*Toyota 意図せぬ急加速事件*」『連邦補足判例第二集』978巻1099頁、1101～02頁 (斜体は原文, 下線付加):

> [A] jury must . . . conclude either that [the driver] mistakenly pressed the accelerator pedal instead of the brake pedal, or that it did not. If the jury finds that she was not mistaken, that necessarily establishes the *existence* of a mechanical malfunction. . . . . A reasonable jury . . . [may] infer . . . existence of [a specific defect that could have opened the Camry's throttle from its idle position]. This is particularly appropriate in light of the fact that the Camry software does nothing to track its own failure. If it did, the lack of any identification of a software failure [by the plaintiff] would support Toyota's position [that Toyota should enjoy summary judgment for itself]; however, absent the ability to trace software failure, the lack of evidence of a specific type of failure is merely inconclusive [決定的ではない].

(参考①)

Article

論　説

「ロボット法」と自動運転の「派生型トロッコ問題」

――主要論点の整理と、AIネットワークシステム「研究開発８原則」

中央大学大学院総合政策研究科委員長

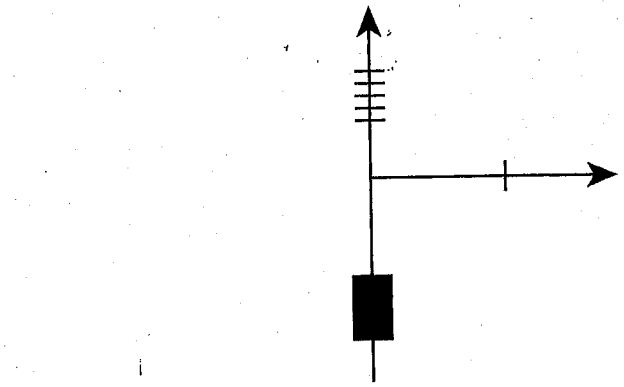
平野　晋 Susumu Hirano

要旨：本誌でその効用を紹介した自動運転が[1]、実用化されつつある。しかしその前には、思考実験であった「[派生型] トロッコ問題」(The Trolley Problem [Variants]) (【図 1】)[2] 等への対応が不可欠と海外では指摘されている。

――自動運転車が前方の５人を轢きそうで、回避の為に右にハンドルを切るとその先の一名を轢いてしまう。――設計者は事故よりも遥か以前に、設計選択を迫られる[3]。この議論の現状を整理・分類し、日本が先にG7に向けて政策提言し賛同を得たAIネットワークシステムの「研究開発８原則」[4] に照らしつつ、解決策を社会全体として検討すべきと指摘する。

目次

【図1】「転轍機に居合わせた人の問題」(The Bystander at the Switch)



出典：J. Thomson・後掲注(2) 1402頁。

はじめに(ロボット法の必要性)

ヒト型ロボットに、右手で左耳に触りなさいと頼んでみたまえ。
奴らは殆どいつも、こんな風に動くんだよ——左耳に触る為に、
頭を突き抜けようとするんだ——[5]。

ロボットは、環境情報を「**認識**」し、ヒトから指令された目的達成の為に自ら最適と「**判断**」した方策に基づいて(自律性)、「**行動**」する(**認識・判断・行動サイクル**：“sense-think-act” cycle)。しかしその「判断」が、ヒトの思いもよらない突飛で危険なものになることが、前掲引用文のように危惧されている。ロボット法の主導者達は、そのような危険な道具が蔓延する前に、ロボットの特性を理解し、生じ得る危険性を予見し、対策を検討することが必要、と主張する。その状況は、嘗て「サイバー法」の必要性が叫ばれていた頃に似ている[6]。

「ロボット・カー」とも呼ばれる自動運転車にも、「認識・判断・行動サイクル」が組み込まれる。不可避的な事故に遭遇することも、稀ではあるが十分に想定される。その際に最適な判断を設計に組み込む為には、倫理的規範を、透明性を維持しながら社会的に議論・検討することが必要である、と海外では主張されている。さらに筆者が座長代理を務めた総務省「AIネットワーク化検討会議」も、倫理の重要性を含む「研究開発8原則」を提言し(【表1】参照)、先のG7サミットでも賛同を得ている[7]。

以下Ⅰでは、倫理問題が思考実験にとどまらない実例として、いわゆる「派生型**トロッコ問題**」を扱ったアメリカ代表判例を紹介する。Ⅱでは、「衝突最適化—crash optimization—」設計における、海外での倫理的課題論議の代表例を紹介しながら[8]、そこに「研究開発8原則」を照らし合わせてみる[9]。

Ⅰ 近未来版「エッカート対ロング・アイランド鉄道事件[10]」

日本では、倫理的課題の優先度は低いという声もいまだに見掛ける[11]。しかし、倫理的課題があながち思考実験にとどまらない事実を理解してもらう為に、「*エッカート事件*」[12]を紹介しよう。ヘンリー・エッカート氏は、目の前の線路に幼子が座って轢かれそうな場面に遭遇する。とっさの判

【表1】「AIネットワークシステム研究開発8原則」

8原則	概要
① 透明性の原則	動作の説明可能性および検証可能性を確保
② 利用者支援の原則	利用者を支援、利用者に選択の機会を適切に提供するよう配慮
③ 制御可能性の原則	制御可能性を確保
④ セキュリティ確保の原則	頑健性および信頼性を確保
⑤ 安全保護の原則	利用者および第三者の生命・身体の安全に危害を及ぼさないよう配慮
⑥ プライバシー保護の原則	利用者・第三者のプライバシーを侵害しないよう配慮
⑦ 倫理の原則	人間の尊厳と個人の自律を尊重
⑧ アカウンタビリティの原則	利用者等へのアカウンタビリティを遂行

断で幼子を救ったけれども、ヘンリー自身は轢かれてその夜に死亡する。遺族が鉄道会社の賠償責任を求める裁判において、ヘンリーの寄与過失（contributory negligence）が問われ、NY州最上級審は注意義務違反がなかったと評価。ヘンリーの英雄的な「とっさの判断」に、過失を認定できなかったのである。本件が「派生型**トロッコ問題**」に似ていると筆者が思う理由は、何れかの命が失われるしかない状況（不可避的事故）における選択・判断の妥当性が問われたからである。

完全自動運転が導入されつつある現代では、ヘンリーの立場に設計者やプログラマーが取って代

1 拙考「製造物責任（設計上の欠陥）における二つの危険効用基準～ロボット・カーと『製品分類全体責任』～」本誌1040号（2014）43頁、44頁。

2 Judith Jarvis Thomson, *The Trolley Problem*, 94 YALE L. J. 1395（1985）（Philippa Footが最初に提起した問題を、「転轍機に居合わせた人の問題」等に改良して分析）. 派生型の方が実験倫理哲学において広く検討されている。Peter Danielson, *Surprising Judgments about Robot Drivers: Experiments on Rising Expectations and Blaming Humans*, 9 NORDIC JOURNAL OF APPLIED ETHICS 73, 75（2015）.

3 Patrick Lin, *The Ethics of Autonomous Cars*, THE ATLANTIC（Oct. 8, 2013）,〈http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/〉（last visited June 6, 2016）; Nick Belay, Note, *Robot Ethics and Self-Driving Cars: How Ethical Determinations in Software Will Require a New Legal Framework*, 40 J. LEGAL PROF. 119, 121 & n. 25（2015）.

4 AIネットワーク化検討会議「報告書2016：AIネットワーク化の影響とリスク―智連社会（WINS）の実現に向けた課題―」（平成28年6月20日）42～47頁〈http://www.soumu.go.jp/main_content/000425289.pdf〉（last visited July 2, 2016）。

5 Curtis E.A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, *in* ROBOT LAW 51, 51（Ryan Calo et al. 2016）. 筆者訳（強調付加）。

6 A. Michael Froomkin, *Introduction*, ROBOT LAW・同上 x, x-xiii等参照。

7 検討会議報告書・前掲注(4)2頁。なお筆者は、ロボットビジネス推進協議会の元保険部会長でもある。

8 字数節約の為に細かな注書を省略するが、四の「衝突最適化」分析や指摘は別段の記述がない限り原則として、以下のすべてまたはいずれかが示しているものを筆者が整理・分類し、または修正・発展させたものである。Patrick Lin, *The Robot Car of Tomorrow May Just Be Programmed to Hit You*, WIRED, May 6, 2014,〈http://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/〉（last visited June 7, 2016）; Noah J. Goodall, *Ethical Decision Making during Automated Vehicle Crashes*, 2424 J. TRANS. RES. BOARD 58（2014）〈http://people.virginia.edu/~njg2q/ethics.pdf〉（last visited June 7, 2016）; Jeffrey K. Gurney, *Crashing into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law*, 79 ALB. L. REV. 183（2015-2016）; Wesley Kumfer & Richard Burgess, *Investigation into the Role of Rational Ethics in Crashes of Automated Vehicles*, 2489 J. TRANS. RES. BOARD 130（2015）; Noah J. Goodall, *Can You Program Ethics into a Self-Driving Car?*, IEEE SPECTRUM, May 31, 2016,〈http://spectrum.ieee.org/transportation/self-driving/can-you-program-ethics-into-a-selfdriving-car〉（last visited June 21, 2016）; Jason Millar, *An Ethical Dilemma: When Robot Cars Must Kill, Who Should Pick the Victim?*, ROBOHUB（June 11, 2014）〈http://robohub.org/an-ethical-dilemma-when-robot-cars-must-kill-who-should-pick-the-victim/〉（last visited July 2, 2016）. 他にも意識調査が複数公表され、新しくはJean-Francois Bonnefon, Azim Shariff, & Iyad Rahwan, *The Social Dilemma of Autonomous Vehicles*, 352 SCIENCE 1573（June 24, 2016）参照。

9 本稿はAIネットワーク化検討会議の意見ではなく、私見である。

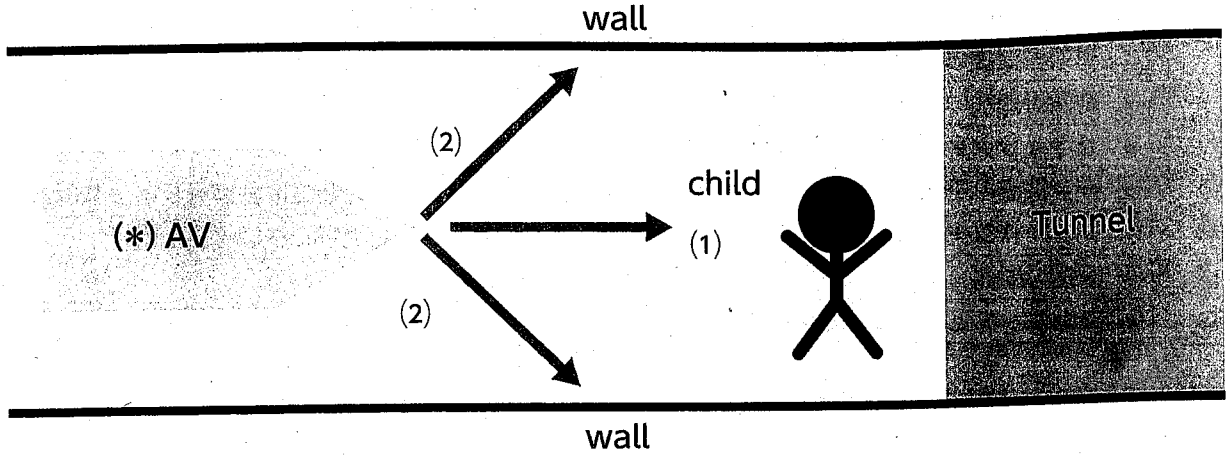
10 Eckert v. Long Island R.R., 43 N.Y. 502（1871）.

11 国土交通省「自動走行ビジネス検討会将来ビジョン検討ワーキンググループ（第2回）議事要旨」（2015年11月10日開催）4頁〈http://www.mlit.go.jp/common/001118839.pdf〉（last visited June 21, 2016）（「倫理問題については、ベテランドライバーでも対応が難しいため、…優先順位は低い…。」）。

12 拙考「補追『アメリカ不法行為法』判例と学説〔3〕」際商35巻12号（2007）1736～38頁（同事件要旨と評価・分析を紹介）。

【図2】「トンネル問題」

(*)「AV」は「autonomous vehicle」(自動運転車)の意。



(1)**子供を轢く**か、または(2)**利他的に壁に激突して自身の死を選ぶか**の、究極の選択を迫られる。
Drawn by S. Hirano based upon a hypo. in Millar, *supra* note 8.

わりつつある。たとえば、【図2】のように自動運転車が、(1)トンネル前方に飛び出した子供を轢いてしまうか、または(2)子供を回避して壁に激突し乗員が死に至るかの選択を、(運転者ではなく)設計者達が迫られつつある。(ヘンリーのように利他的な)選択(2)を設計者達がとれば、彼等が所属・関係する製造業者等が、後日、乗員の遺族から製造物責任や過失責任を問われ、(1)を選択しても子供の遺族から同じく責任を問われ得る[13]。その際、製造業者等が、「とっさの判断」ゆえに責任なしと評価されることはない。責任を問われているのは、十分に検討する時間が与えられていたプログラミング段階の「設計選択」(design choice)だからである。すなわち(ロボットに法人格を付与等しない限り[14])争点は、ロボット・カーの事故時の判断ではなく、そこに至った「設計上の欠陥」の有無になる。

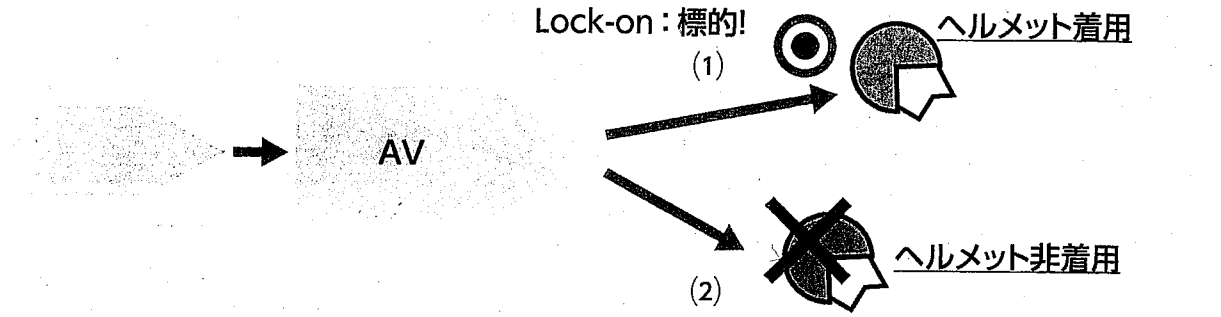
このように「*近未来版エッカート事件*」とでも表し得る衝突を最適化する前に検討すべきと海外で指摘されている倫理的諸課題を、以下で整理・分類し、「研究開発8原則」に照らし合わせてみたい。

Ⅱ いくつかの諸問題(例示)と「研究開発8原則」の当てはめ

1. 不可避的事故の瞬間には、自動運転ではなくヒトに判断を委ねるべきという案の抱える問題点

そもそも機械に判断を委ねること自体が反倫理的である、という主張は多く見受けられる[15]。これに対しては、そもそもヒトの能力が劣るから自動運転を導入する趣旨に反する、という批判も考え得る(後掲9も参照)。社会的議論抜きで自動運転に委ねない選択を決すれば、より良い認識・判断・行動能力を有する[かもしれない]自動運転に選択を委ねたい利用者から自律的選択の機会(**②利用者支援の原則**)を奪うかもしれず、安全性を軽視し(**⑤安全保護の原則**)、果たして本当に倫理的か否か(**⑦倫理の原則**)について疑義が残り、かつ利用者への説明が欠ければ説明責任(**⑧アカウンタビリティの原則**)も果たさないことになるのではないか。

【図 3】「バイク問題」



不可避的事故に至る状況に遭遇した「AV」は、**対人損害を極小化する目的からは、(1)ヘルメット着用者を「標的」にして衝突することが選択される**。が、しかしこの選択は ... 。

Drawn by S. Hirano based upon a hypo. in Goodall, *Program Ethics, supra* note 8; Lin, *Programmed to Hit You, supra* note 8.

2. 安全性の高い相手が「標的」にされる問題

「**バイク問題**」（【図 3】）において残された進路は、(1)ヘルメット着用バイクへの衝突か、または(2)非着用バイクへの衝突である。社会全体の損失額最小限化こそが望ましいという功利／結果主義（utilitarianism／consequentialism）的目的を実現するならば、(1)ヘルメット着用者が「標的」（target）とされ得る[16]。

製造物責任法的に分析すれば、レベル 4 時代[17]には製造業者等の責任比率が高まり[18]、損害賠償責任リスクへの「露出」（exposure）を最小限化しようとする製造業者等も安全性の高い相手を標的とする設計を選択しがちといわれている。運行供用者が責任を負う場合でも、そのリスクを引き受ける損害保険会社としても賠償額の最小限化を願うから[19]、損保業界としては安全性の高い相手を標的とすることに利益があり、その影響力が懸念されよう。

以上の設計選択は、ヘルメット着用を怠る望ましくない行為者を優遇し、逆に望ましい者を懲らしめるので、不公正であると指摘されている。（説明責任を果たせるか否かの⑧**アカウンタビリティの原則**）さらにはヘルメット非着用の方が標的にされないから望ましいという誤ったメッセージが広まるおそれ（⑤**安全保護の原則**）も指摘されてい

13 子供の命を救える代替設計案（RAD：reasonable alternative design）不採用の理不尽さが設計上の欠陥基準である旨は、拙考「製造物責任法リステイトメント起草者との対話―日本の裁判例にみられる代替設計『RAD（ラッド）』の欠陥基準―」本誌 1014 号（2013）40 頁、42 頁、45～49 頁参照。拙考「走行情報のプライバシーと製造物責任と運転者の裁量」『知財研フォーラム』103 巻（2015）26 頁、27～28 頁も参照（Eugene Volokh, *Tort Law vs. Privacy*, 114 Colum. L. Rev. 879（2014）を紹介しながら、速度制限違反自動感知報告装置や飲酒運転自動感知報告装置の採用を怠った製造業者等が設計上の欠陥を問われると指摘）。

14 その蓋然性は（少なくとも当分の間は）低いであろう。

15 Clive Thompson, *Relying on Algorithms and Bots Can Be Really, Really Dangerous*, Wired, Mar. 25, 2013 〈http://www.wired.com/2013/03/clive-thompson-2104/〉（last visited June 7, 2016）; Danielson・前掲注(2) 82 頁; Millar・前掲注(8). この抵抗感は特にロボット兵器の論議で活発である。Human Rights Watch, Losing Humanity: The Case Against Killer Robots（2012）38～39 頁〈https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf〉（last visited July 2, 2016）.

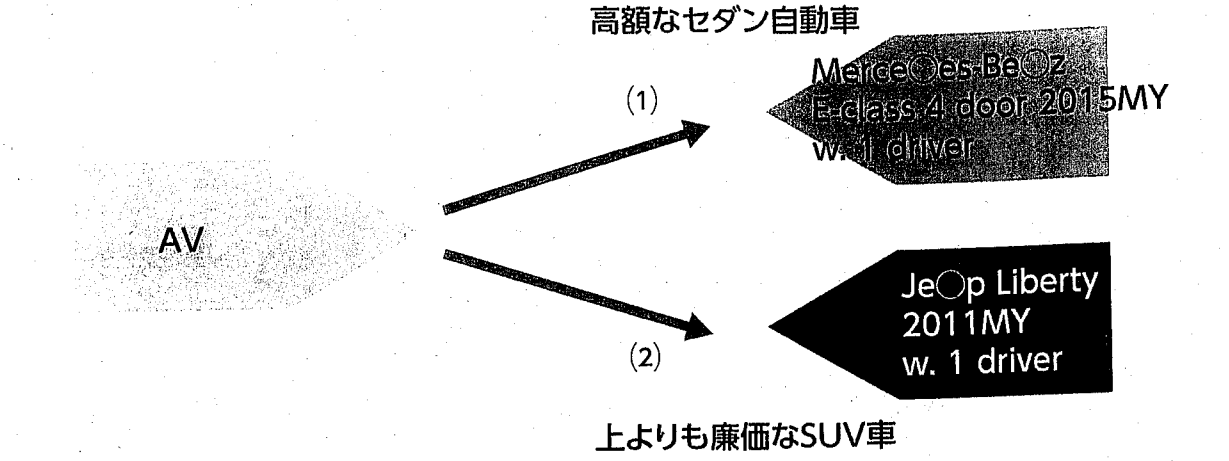
16 もっともGurney・前掲注(8) 213 頁参照（規則功利主義者は異なり得る）。

17 現在の倫理規範論議の主な例示は、完全自動運転車（レベル 4）と、ヒトが運転する自動車等が混在する不可避的事故の衝突最適化である。Kumfer & Burgess・前掲注(8) 135 頁も参照。

18 拙考「ロボット・カー」・前掲注(1) 44 頁参照。

19 Belay・前掲注(3) 127～28 頁参照。

【図4】「自動車問題」



Drawn by S. Hirano based upon a hypo. in Gurney, *supra* note 8, at 198-202.

る。（さらに差別的取扱いの問題については後掲 5）

3. 富裕者優遇の問題

「**自動車問題**」（【図4】）は、乗員の人身損害が生じないかまたは軽微ゆえに、対物損害が焦点になる。(1)は高額な対向車で、(2)は低額である。製造業者等は、財物損害額の低い(2)を標的とする設計を選択する。運行供用者が賠償責任を負う場合でも、損保会社は（も）(2)への衝突を望むであろう。

このハイポ（hypo：仮想事例）は、衝突対象から回避されがちな富裕者が優遇されるという問題を示している。（説明責任が果たせるか否かの**⑧アカウンタビリティの原則**）（さらに差別的取扱いの問題については後掲 5）

4. 所有者優遇の問題

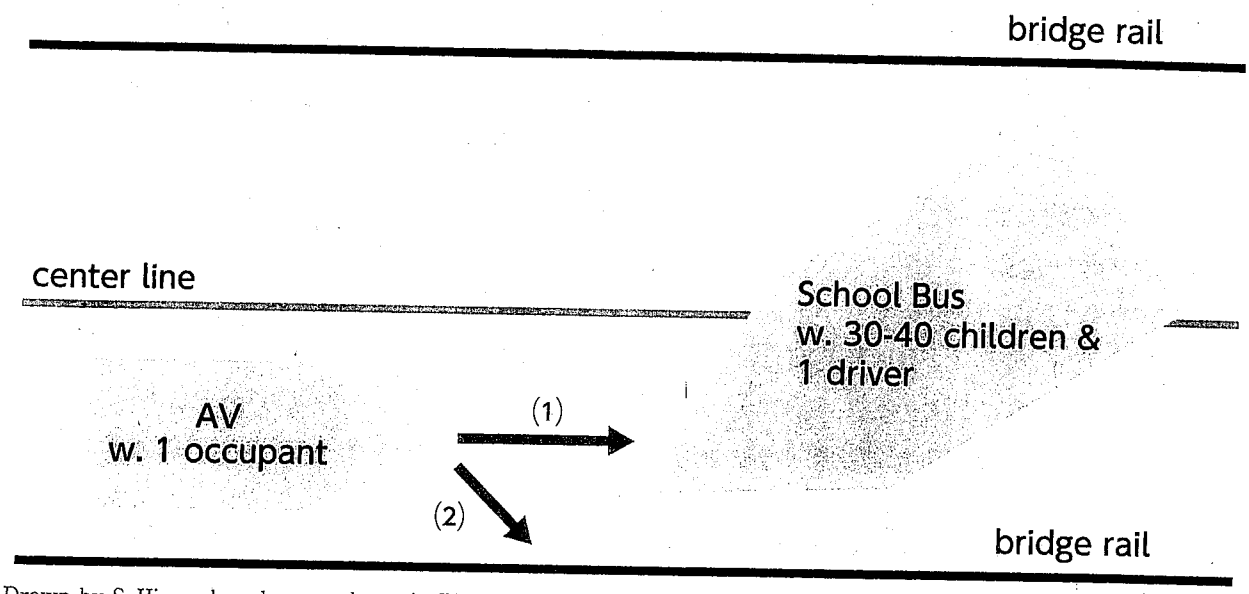
「**橋問題**」（【図5】）では、子供が（多数）乗車したスクール・バスが対向車線から突っ込んで来て、(1)バスと衝突して子供達が死に至るか、または(2)利他的に橋から落ちる選択肢しか残されていない。功利／結果主義からは、(2)の利他的な自己犠牲が求められる。しかし製造業者に設計選択を委ねてしまうと、自社製品の乗員を犠牲にして赤の他人を利するようなクルマは売れないから[20](1)が選択される問題が指摘されている。（**⑤安全保護の原則**）もっとも、利他性を強いる選択(2)が正しいか否かは、前掲「**トンネル問題**」（【図2】）同様に**⑦倫理**および**⑧アカウンタビリティの原則**上の難しい問題である。

製造業者等のみに設計選択を任せて、特に人工知能のような複雑な仕組みを用いた場合、ある進路をとって衝突に至った「判断」の理由がヒトには理解し難い為に、隠れて自社製品の乗員を優遇する設計選択が組み込まれても、その「マニピュレイション」（manipulation—操作—）が不明なままの状態が長く存続する危険性も指摘されている[21]。これは、**①透明性の原則**に反し、乗員に自律的選択の機会を与えないので**②利用者支援の原則**に反し、**⑦倫理の原則**に反し、かつ説明責任を欠くので**⑧アカウンタビリティの原則**にも反するであろう。

5. 差別的取扱いと、命を秤にかける問題

そもそも差別的取扱いは倫理的に避けるべきと

【図5】「橋問題」



Drawn by S. Hirano based upon a hypo. in Thompson, *supra* note 15 (originally in G. Marcus, *Moral Machines*, New Yorker Blogs, Nov. 27, 2012 〈http://www.newyorker.com/news/news-desk/moral-machines〉 (last visited June 10, 2016)).

いう指摘も見られる。(⑦**倫理**および⑧**アカウンタビリティの原則**) たとえば前掲「**自動車問題**」(【図4】) の条件を変えて、人身損害が生じ得ると考えてみよう。(1)は高額なセダン自動車で衝突耐性が高く、他方(2)もSUV車ゆえに通常のセダン自動車よりも衝突耐性は高い。甲乙付け難いので、顔認識装置や、クルマ同士が通信し合って相手方情報を理解し合う「V2V通信」(vehicle-to-vehicle com.) 等を通じて、先方乗員の属性等を見極めて致死率の把握を試みるかもしれない。統計上、女性の方が男性よりも28%、老人の運転者の方が若者よりも3倍、酩酊者はシラフの2倍、さらには乗員が一人の場合の方が複数乗員の場合よりも14%、死亡率が高い[22]。しかし、これら属性等次第で標的を決めることは、ヒトの命を皆平等に扱うべきという倫理からは、差別的取扱いになり説明に窮するのではないか。

さらに「**自動車問題**」の登場人物を修正して、たとえば衝突相手の選択肢が(1)子供か、または(2)老人かに置き換えてみた場合、果たして余命寿命の長短による評価が許されるであろうか。これは年齢による差別に当たるけれども、そのように高度な倫理的判断も、「命の価値評価」(value-of-life estimates) や臓器移植を待つ者の優先権決定 (identification of organ transplant recipients) 等の医学の世界において学ぶべき知見が存在するかもしれない[23]。

6. 相手方の過誤(fault)を考慮しない問題

前掲「**トンネル問題**」(【図2】) において、(1)飛び出してきた子供を救う為に、(2)自動運転車が

20 同仮説に関連する実証実験の結果はBonnefon et al.・前掲注(8) 1575頁参照。
21 特にGoodall, *Decision Making*・前掲注(8) 63頁参照。
22 同上62頁。
23 同上63頁; Goodall, *Program Ethics*・前掲注(8); Millar・前掲注(8)も参照。

利他的に乗員を自己犠牲させた場合、そもそも事故原因である子供こそが非難されるべきで、落ち度がないにも拘わらず自動運転車側の乗員が損失を被るのは不公正とも考えられる。すなわち、非難可能性や帰責性等を判断の要素に入れない問題も、「**トンネル問題**」は示唆している。(説明責任を果たしたか否かの**⑧アカウンタビリティの原則**)

7. 明確化しにくい倫理規範の限界と対応策

功利／結果主義だけを行動指針とすると、多くの者にとって納得のゆかない設計選択となることは、「衝突最適化」を検討する際の障害の一つである。(**⑤安全保護**、**⑦倫理**、および納得のゆく説明が難しい**⑧アカウンタビリティの原則**) これは、Judith Thomsonが「**転轍機に居合わせた人の問題**」(前掲【図1】)の比較対象としたハイポ、「**太った男の問題**」(The Fat Man)[24]を思い起こせば自明である。すなわちトロッコ線路上の陸橋に居合わせた太った男を突き落とせば暴走トロッコを止めて5人を救える場合でも、太った男を殺すことは許されない。社会全体にとっての命の損失を最小限化する功利／結果主義的な倫理規範だけでは、問題を解決し得ないのである。

そこで、エマニュエル・カントに代表される「義務論」(deontology)がしばしば対比される。人を道具としてのみ利用することは許されないという規範である。しかしこれも抽象的過ぎて、当てはめが難しい。その問題点は、いわゆるアシモフのロボット工学3原則が機能しない、とアシモフ自身が短編『堂々めぐり』(Runaround)や『うそつき』(Liar!)にて示唆している[25]。

しかし、現時点においても抽象的で表明することが難しい倫理規範を、複雑過ぎてヒトには理解できない人工知能の問題——ロボット法が特に関心を寄せる問題——を治癒した上で教え込む提案を、Goodallが表明しているので興味深い[26]。(**①透明性**、**②利用者支援**、**③制御可能性**、**⑤安全保護**、**⑦倫理**、および**⑧アカウンタビリティの諸原則**)

8. 意図的にセンサーを機能させない(目をつぶる)問題

差別的取扱いを避ける方策として、自動運転車のセンサーを機能させず、衝突対象候補者の属性情報を敢えて収集しない提案も考えられる。しかしこの情報は事故解析や今後の安全性向上等(**⑤安全保護の原則**)の為に有用な情報であるから、それを収集しない選択は、現実的ではないという指摘もある。

9. ランダムに運命を決める問題

衝突する相手がその都度ランダムに決せられる設計を採用して、差別的取扱いの批判を回避する案も示されている。いわば意図的に運に任せる設計であるが、利用者に適切な選択の機会を付与しないので、**②利用者支援の原則**上問題かもしれない。さらに、そもそも90%超の事故原因たるヒューマン・エラーを解消すべく優れた自動運転を導入しようとしているところ、その能力を意図的に使用しない設計選択は自動運転導入の趣旨に反するとの批判がある。加えて、事故時よりも遥か以前のプログラミング段階で悲惨な結果を回避し得たのに(**⑤安全保護の原則**)、その選択の機会を意図的に捨て去ることは社会的に許されないという批判もある[27]。加えて、倫理規範の問題が難しいから判断に至る思考さえも捨て去ることは、倫理的に許されないとの批判もある。(**⑦倫理**および**⑧アカウンタビリティの原則**)

10. 所有者に運命を決めさせる問題

購入者に販売店で選択させれば、自律性を重んじるので**②利用者支援の原則**には適うが、しかしたとえば社会規範に反する選択肢を所有者が選ぶおそれがあるという批判がある。(個人の自律尊重

の限界としての**⑦倫理の原則**）

もっとも筆者には、たとえば社会規範に反する問題は選択肢を所有者に付与しないように設計段階で組み込んでしまえば、そのような批判を回避できると思われる。しかしそれでも、いまだ次のような問題が残ると筆者には思われる。たとえば「**自動車問題**」（［図4］）で、損害賠償額が高額な選択肢――衝突(1)――を、購入者の信念や単なる気まぐれ等から選んでしまうと、損害保険料が高額になり得る。そこで保険料を安く抑えたい所有者にも、富裕者への衝突回避を自発的に選ぶような経済的インセンティヴが働いてしまう。（前掲3）

さらに、これは論者達があまり論じていないが、V2V通信が導入された場合に自動運転車同士が「取引」する事態においては、所有者達の異なる（勝手な）意向を受けて機械同士がどのような取引を行うのかが――結果的にはヒトの命のやりとりなので――、大きな関心事にならざるを得ない。（**①透明性、③制御可能性、⑦倫理**、および**⑧アカウンタビリティの諸原則**）すなわち、本稿の取り上げる諸問題（の少なくとも幾つかについて）は、各所有者の（身勝手な）選択に委ねることが許されず（**②利用者支援の原則**および**⑦倫理の原則**［の限界］）、社会規範の確立こそが求められるのではなかろうか。

おわりに

以上により、衝突最適化に関し海外で指摘されている倫理諸課題においては、説明責任を果たしているか否かの**⑧アカウンタビリティの原則**や**⑤安全保護の原則**が多く問題となることが判明したのではあるまいか。加えて、ある判断に至った理由が不明なままAI等を用いたり、開かれた議論抜きで恣意的に設計が選択されれば、**①透明性、②利用者支援**（利用者自律性）、**③制御可能性**、および**⑦倫理の諸原則**も問題になることが明らかになったであろう。やはり学際的知見を結集しつつ、透明性のある社会的合意形成の努力は必要であろう。

【本稿は2016年7月7日に脱稿した。】

24 Thomson・前掲注(2)1409頁、1415頁。同様に、5人の臓器移植を待つ患者を救う為に一人の健康な人を犠牲にすることも許されない（移植問題）。同上1396頁、1401頁、1408頁。「切り札としての権利は効用に勝る」ゆえに、一人の命の権利を侵害してまで5人を救うことは許されず、転轍機の場合は右折先の一人がその場を立ち去っても5人が助かることに変わりはないから、**太った男・移植問題**と異なる。Christopher Hitchcock, *The Metaphysical Bases of Liability: Commentary on Michael Moor's Causation and Responsibility*, 42 RUTGERS L. J. 377, 401-02 (2011) も参照（同旨）。

25 抽象的概念や文言の解釈が難しい点に関し、ロボット工学のいわゆる第三原則「［ヒトに危害を加えずヒトの命令に服従すること］に反するおそれのない限り、自己を守らねばならない」の中の「反するおそれのない限り」の理解が難しく、「危害」の理解も難しいことが、両作品に表れていよう。瀬名秀明「『ロボット学』の新たな世紀へ」*in*アイザック・アシモフ（小尾芙佐訳）『われはロボット〔決定版〕』414頁（早川書房、2004）および両作品参照。

26 Goodall, *Decision Making*・前掲注(8)63～64頁。

27 Adam Kolber, *Will There Be a Neurolaw Revolution?*, 89 IND. L. J. 807, 844 (2014) も参照（"we can no longer hide behind ambiguous facts" と指摘）。

☆アメリカ・ビジネス判例の読み方☆ ⑳

In re Toyota Motor Corp. Unintended Acceleration

～AI・ロボット・自動運転時代の「誤作動法理」適用を示唆する事例～

平野 晋*

はじめに

いわゆる急加速問題の誤作動法理に関する事例を紹介する。

【事件名】 *In re Toyota Motor Corp. Unintended Acceleration*, 978 F.Supp.2d 1053 (C.D.Cal. 2013).

【裁判所】 連邦地方裁判所カリフォルニア中央区担当

【判決日・決定日】 2013年10月7日

【事件の概要】 本件は，Toyota社（π）製2005年型Camry車の「意図せぬ急発進：sudden, unintended acceleration (SUA)」が主張された連邦広域継続訴訟—— multi-district litigation (MDL) ——に属する事件で，故人（本件事故が死亡原因ではない）St. John夫人の遺言執行者（π）を通じて提起された訴えである。

事故時，St. John夫人が運転するCamry車が停止表示に従って完全に停車した後，夫人がブレーキから足を離して右折しようと思ったところ暴走して制御不能に陥り，体育館入口の円柱に突っ込んだ。夫人はブレーキを掛けて停止しようと努力したけれども，車は勝手に加速していった，と夫人は証言。目撃証人は，Camryが走行した路上にタイヤ痕があることを目撃している。

πと⊿は多くの専門家証人の証拠を提出し，双方がその不採用を申し立て，かつ⊿はサマリー・ジャッジメント（SJ）も併せて申し立てた。（本稿は主に後者を紹介。）

【主な争点】 欠陥が原因で急加速したか否か等の記録が残らないソフトウエアを搭載した車が急加速した事故に於いて，踏み間違えの可能性が残存し——すなわち誤作動の発生自体さえ十分に立証されておらず——，かつ欠陥の特定及びそれが急加速の原因である旨の［直接的な］証拠が欠けていても，状況証拠のみで，⊿勝訴のSJを免れて事件を陪審員に委ねられるか。

【判決・決定等】 委ねられる。⊿勝訴のSJ申立を却下する。

【主な関連法規・裁判例・学説等】

- *Rose v. Figgie Int'l*, 495 S.E.2d 77 (Ga. App. 1997).
- RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 3 & cmt. *b* [hereinafter referred to as『製造物責任リステイトメント』].
- *Jenkins v. General Motors Corp.*, 524 S.E.2d 324 (Ga. App. 1999).
- *Miller v. Ford Motor Co.*, 653 S.E.2d 82 (Ga. App. 2007).
- *Stanley v. Toyota Motor Sales, U.S.A., Inc.*, 2008 WL 4664229 (M.D. Ga. Oct. 20, 2008).

【理由】 原告は欠陥の性格を正確に特定する必要はなく，「装置が意図されたようには機能しなかった…do not operate as intended…ことを示せば」足りる。先ず製造上の欠陥は，欠陥を証明する為の証拠が破壊され，又は原告側の落ち度なしに利用不可能な場合には，状況証拠によって証明することが許される。本件に於いても，⊿のソフトウエアが不具合を記録しない仕組みであった。記録が無いから原告が一応の証明責任を果たしていないと裁判所が評価［し

*ひらの すすむ，中央大学教授・大学院総合政策研究科委員長，博士（総合政策・中央大学），米国弁護士（ニューヨーク州）

て請求を棄却］すれば，「欠陥自体がその存在の証拠を破壊する場合…where the defect causes its own destruction…」には常に欠陥製品の責任からメーカーを免れさせてしまう。探知…trace…不可能な事象までも探知せよ等と，裁判所は原告に要求しない。状況証拠による製造上の欠陥証明を許容してきた判例の論理は，設計上の欠陥にも容易かつ論理的に適用できる。『製造物責任リステイトメント』§3のコメントb.も，原告による設計上の欠陥特定を不要としている。

誤作動の原因が［欠陥以外にも］複数考えられる場合に，欠陥こそが誤作動の原因であるという推認を許す為には，他原因を原告が排除しなければならないと判断した裁判例「*Jenkins*」，「*Miller*」，及び「*Stanley*」を⊿は挙げている。

しかし，上記３つの事件では誤作動の「原因…cause…」が争点であったけれども，本件では誤作動の「存在…existence…」自体が争点なので異なる。すなわち本件では，ひとたび陪審員がアクセルを間違って踏んではいないと認定すれば，一つだけ存在する［欠陥以外の］他原因が排除されるから，自動的に機械的誤作動が「存在」したことに成り，他原因排除の欠如という欠点が治癒されてしまう。従って，設計上の欠陥主張をトライアルに進める前には，踏み間違えの可能性をπが排除せねばならないという⊿の主張を当裁判所は認めない[1]。

πは特定のソフトウエア設計上の欠陥又は製造上の欠陥を絞り込めなかったし，運転者が踏み込んだアクセル・ペダルからのアナログ情報なしに，欠陥ゆえにスロットルがアイドリング位置から勝手にもっと広く開いた位置に成った旨の物的又はその他の［直接的］証拠を提示できなかった。⊿のフェール・セーフ機構が機能しなかった旨も［直接には］証明できなかった。そして大量の証拠が双方から提出されたにも拘わらず，踏み間違えたという推認しか認めるべきではないと⊿は主張している。しかし当裁判所はそのように結論できない。St. John 夫人の証言と他の証拠から，ブレーキを踏んだにも拘わらず加速し続けたと，理に適った陪審員は推認することが出来る。

設計上の欠陥ではなく踏み間違えが原因である可能性をπは否定できなくとも良く，状況証拠によって設計上の欠陥を証明できる[2]。設計上の欠陥に基づく請求を支持する大量の証拠をπは提出している。尤も特定の欠陥ゆえにスロットルがアイドリング位置から勝手に開いた旨の許容可能な証拠提出をπは怠ってはいるけれども，その存在を推認することが理に適った陪審員に許されるに足る証拠をπは示している。［本件のように］⊿のソフトウエアが自らの不具合を探知する為の工夫を何ら行っていない事実に照らせば，上のような推認を許容することが特に適していると云える。

設計上の欠陥に於ける危険効用基準に関係する代替設計案についてもπは，少なくとも２つの設計案を示している。

⊿は，たとえπの主張が正しかったと仮定しても，フェール・セーフ設計が事故を防止すると主張している。しかしこの主張は，(1)フェール・セーフ設計そのものが誤作動しないこと，及び(2)これが作動する為の前提条件が全て事故前に満たされたこと，という仮定に基づいている。しかしフェール・セーフが作動しなかったか又は不完全にしか作動しなかったことを［間接的に］示す証拠を，πは少なくとも２つ示している。

おわりに（ビジネス上の留意点）

現在の AI 開発論議に於いても，問題原因のトレース可能性や透明性が開発者に強く求められている。

〔注〕

1　なお SJ 審査の段階では，踏み間違えていないという St. John 夫人の証言が正しかったという前提で審査される，と法廷意見は付言している。978 F.Supp.2d at 1099 n.77.

2　この点については『製造物責任リステイトメント』§3の例示５が説得的である，と法廷意見は付言している。*Id.* at 1101 n.78.