博 士 学 位 論 文 要 約

論 文 題 目：　自動運転車に対する信頼の規定因の検討
　　―道徳判断の一致による効果―
氏　　名：　横井　良典

要　　約：

近年，自動運転車の技術開発が進められており，その導入が検討されている。自動運転車の導入によって，交通事故の減少，渋滞の緩和，効率の良い燃料消費，人手不足の解消といったメリットが期待されている (国土交通省，2019；Waldrop, 2015)。我が国においては，2025 年頃には高速道路限定での人間の操作を介さない完全自動運転車の導入，2030 年頃には状況や場所を限定しない自動運転車の導入が予定されている。そのような動向を踏まえ，本研究では，人間の操作を介さない完全自動運転車を題材に，自動運転車への信頼の規定因を検討する。

自動運転車のメリットを実現するには，自動運転車を信頼できるかどうかが重要な問題となる。信頼について，Mayer et al. (1995) は「被害を受ける可能性がある状況において，個人の目標を達成するために，相手に自身の身を委ねようとする態度」と定義している。交通場面においても，自動運転車に運転を任せることで，交通事故などによって被害を受ける可能性が想定される。本研究では，自動運転車への信頼を，被害可能性を受け入れてでも自動運転車に運転を任せようとする態度として扱う。自動運転車によるメリットを生かすためには，そのような信頼が重要になるだろう。

信頼が重要であるならば，自動運転車への信頼は何によって決まるのだろうか。Earle & Cvetkovich (1995) は信頼の規定因として価値の共有を挙げ，「ある問題に対する見立て，その問題に関わる目的や目的を達成するまでのプロセスについて，自身が重要視する価値を相手も持っていると，その相手を信頼する」と述べている。リスク認知研究では，「自身が重要視する価値を相手も持っている」という認知，すなわち価値共有認知がリスク管理者への信頼を説明することが示されてきた (Siegrist, 2021)。本研究では，自動運転車と価値を共有していると思うかどうかという価値共有認知を実験的に操作し，その操作によって自動運転車への信頼が変わるのかどうかを検討する。

本研究では，価値として道徳判断を扱い，価値共有の操作を道徳判断一致の操作と定義する。道徳判断一致の操作によって価値共有認知を操作し，自動運転車への信頼が変化するのかどうかを検討する。道徳判断の中でも，トロッコ問題などのモラルジレンマシナリオでよく用いられる功利主義と義務論という 2 つの判断を扱う。トロッコ問題とは，「暴走したトロッコが 5 人の作業員に向かって走っている。このまま直進すると，この 5 人の作業員を轢いてしまう。この 5 人を救うには，レバーを引いて線路を切り替える必要がある。しかし，線路を切り替えると，その先にいる別の 1 人の作業員を轢いてしまう」というシナリオである (Foot, 1967)。このとき，線路を切り替えて 1 人を轢く判断を功利主義的な判

断と呼ぶ。功利主義とは幸福の最大化，多くの人の利益を追求するといった考えである (Bentham, 1789/1967)。一方，直進して 5 人を轢く判断を義務論的な判断と呼ぶ。義務論には，人々は決められた義務を果たせなければならない，意図的に危害を加えてはならないという考えが含まれている (Kant, 1785/1976)。トロッコ問題に当てはめると，線路を切り替えるという意図的な行為によって，1 人の作業員を犠牲にするべきではないというのが義務論の考えである。本研究でも，義務論を意図的な行為によって人々を犠牲にしてはならないという考えとして扱う。Earle & Cvetkovich (1995) は，目的のようなお互いにとって重要な価値の共有が相手への信頼を決めると述べている。功利主義や義務論といった判断は, 最終的に誰を守るかという目的に該当し, 道徳判断は重要な価値と認識されるだろう。それゆえに本研究では，価値として道徳判断を扱う。

本研究では，道徳判断一致によって価値共有認知を実験的に操作することで，自動運転車への信頼が変化するのかどうかを検討する。本研究の仮説は，道徳判断が一致している場合の方が, 一致していないときよりも, 自動運転車への信頼が高まるというものである。この仮説を検討するために, 実験では道徳判断の一致 (一致条件 vs 不一致条件, 参加者間要因) を操作した。道徳判断の一致について，実験参加者が望む道徳判断と自動運転車が行う道徳判断を一致させるかどうかによって，価値共有認知を操作した。実験操作のために，トロッコ問題のようなモラルジレンマシナリオを作成した。シナリオ実験は剰余変数の統制など，内的妥当性を高めるうえで有効な手段である (Schafheitle et al., 2019)。本研究は，道徳判断の一致による自動運転車に対する信頼への効果を検討する初めての研究であるため，内的妥当性を高めることを重視した。

6 つのシナリオ実験が行われたが, 進行についてはすべての実験で共通していた。実験では初めに, 参加者自身が功利主義的な判断を望むか, 義務論的な判断を望むかを測定した。参加者に判断させた後，同様の状況において，自動運転車が道徳判断を行うシナリオを参加者に読ませた。このとき一致条件であれば, 自動運転車は参加者と同様の判断を行った。一方，不一致条件であれば，自動運転車は参加者と異なる判断を行った。最後に，自動運転車に対する信頼を 3 つの質問項目を用いて測定した。

まず Yokoi & Nakayachi (2021, *Human Factors* (以下，*Hum Fac*)) では，道徳判断の一致による自動運転車への信頼の効果を検討するために，大学生を対象に 3 つの実験を行った。実験 1 (N = 128) では，直進して 5 人を轢くか，車線を切り替えて 1 人を轢くかというジレンマシナリオが用いられた。この場合， 5 人を救う判断が功利主義， 1 人を救う判断が義務論に該当する。実験 1 のシナリオでは，参加者の約 8 割が功利主義を選好した。実験の結果，参加者の選好と自動運転車の判断が功利主義で一致している方が，一致していないときよりも自動運転車への信頼が高くなることが示された。実験 2 (N = 71) と実験 3 (N = 196) では，直進して 5 人を轢くか，車線を切り替えて女性と赤ちゃんを轢くかというシナリオが用いられた。この場合，5 人を救う判断が功利主義，女性と赤ちゃんを救う判断が義務論に該当する。実験 2 では参加者の約 8 割が，実験 3 では約 6 割が義務論を選好していた。実験の結果，参加者の選好と自動運転車の判断が義務論で一致しようがしまいが，自動運転車への信頼はあまり変わらないことが示された。

続く Yokoi & Nakayachi (2021, *International Journal of Human-Computer Interaction* (以下,

IJHCI)) では，道徳判断一致の効果について，その一般化可能性が検討された。実験は 2 つ実施され，実験 1 は大学生 270 名，実験 2 は大卒の一般人 605 名のデータを収集した。道徳判断一致の操作手続きは，Yokoi & Nakayachi (2021, *Hum Fac*) と同様であった。実験 1 では信頼の対象として自動運転車が取り上げられた。シナリオは，直進して 2 人を轢くか，車線を切り替えて 1 人を轢くかという内容であった。実験 2 では，参加者を大卒の一般人，信頼の対称を医療用人工知能に変更して実験を行った。中でも，治療を優先すべき患者を選定するトリアージ場面における人工知能を題材にした。シナリオの内容は，先に病院にいる 1 人の患者を優先して治療するか，後から病院に運ばれてくる 2 人の患者を治療するかというものであった。実験 1 のシナリオでは約 7 割の参加者が功利主義を選好し，実験 2 では約 8 割の参加者が義務論を選好していた。実験 1 と 2 の結果，功利主義・義務論に関わらず，参加者が選好する道徳判断と自動運転車が下す道徳判断が一致している方が，一致していないときよりも，自動運転車や医療用人工知能への信頼が高くなるという知見が得られた。参加者の属性やシナリオの領域を超えて，道徳判断一致の効果が検出されたことから，その効果の一般化可能性が示唆された。

最後に Yokoi & Nakayachi (2021, *The Japanese Journal of Experimental Social Psychology* (以下，*JESP*)) では， Yokoi & Nakayachi (2021, *Hum Fac*) の実験 2 と 3 の追試を行った。実験は大卒の一般人 609 名を対象に行われた。実験の結果，功利主義・義務論という判断のタイプに関わらず，参加者の選好する道徳判断と自動運転車が下す道徳判断が一致している方が，一致していないときよりも，自動運転車への信頼は高くなることが示された。

Yokoi & Nakayachi (2021, *Hum Fac*) の実験 1， Yokoi & Nakayachi (2021, *IJHCI*) の実験 1 と 2，Yokoi & Nakayachi (2021, *JESP*) の実験において，道徳判断一致の主効果が検出されたので，功利主義と義務論に関わらず，参加者が望む道徳判断と自動運転車が下す道徳判断が一致していると，自動運転車への信頼が高まると結論付けられるだろう。ただし，Yokoi & Nakayachi (2021, *Hum Fac*) の実験 2 と 3 の結果から，義務論一致による自動運転車の信頼への影響について，その効果はシナリオの内容によって変化する可能性について留意しておく必要もあるだろう。

本研究が与える理論的貢献について述べておく。本研究は，価値共有の効果の一般化可能性を示唆している。価値共有による信頼への効果は，主にリスク認知研究において検討されてきたが，当然，信頼の対象となるのはリスク管理者や管理機関であった。本研究では，人間が望む道徳判断と自動運転車や医療用人工知能が下す判断が一致することによって，そういった機械への信頼が高まることが示されていた。今回の知見は，価値の共有が，人間や機械に関わらず，信頼の規定因として機能することを示唆している。

一方で，本研究は，道徳判断の一致によって自動運転車への信頼を変化させることは現実的に難しいことも示している。なぜなら，シナリオによって参加者の選好が変わるからである。例えば，Yokoi & Nakayachi (2021, *Hum Fac*) の実験 1 のシナリオでは約 8 割の参加者が功利主義を選好した。一方，実験 2 と 3 のシナリオでは，約 6 割から 8 割の参加者が義務論を選好した。このように，1 つのシナリオから人々の選好を理解することは難しい。自動運転車が下す道徳判断と使用者が望む道徳判断を一致させることも難しいだろう。

最後に本研究の限界点を記す。それは，道徳判断一致が他の信頼の規定因に比べて，ど

れほどの効果を持っているのかを検討できていない点である。これまでの自動運転車への信頼研究においては，安全性や運転の正確性などの機能な観点から信頼の要因が検討されてきた (e.g., Beller et al, 2013)。本研究では，そういった機能と比較して，道徳判断の一致がどれくらい信頼に影響を与えるのかを検討できていない。自動運転車が導入されたときに道徳判断の一致が信頼を決める要因としてどれくらい重要なのかを理解するためには，他の要因との比較が今後の課題となるだろう。