

信頼ゲームを用いた秘密の共有による人工知能への信頼に与える影響

Impact on confidence in artificial intelligence by sharing secrets using trust games

永田大貴^{*1}
Daiki NAGATA

篠田夏映^{*1}
Kae SHINODA

沼田知里^{*1}
Chisato NUMADA

磯川雄大^{*1}
Katsuhiro ISOKAWA

濱中杏香^{*1}
Kyouka HAMANAKA

^{*1}同志社大学商学部商学科
Faculty of Commerce, Doshisha University

In this study, we examined impact on trust for artificial intelligence system in situation of secret sharing. Study on human relationship building have found that secret sharing affect the construction of trust relationship. Also, studies on human attitude for robots and computers found that human show reactions similar to human to robots and computers. Therefore, we focused on whether affect trust in artificial intelligence due to secret sharing. We had survey experiment incorporating trust game and the result showed secret sharing affects trust in artificial intelligence.

1. 問題意識と仮説

近年のロボットや人工知能(AI)の普及に伴い、人間は機械的な人工物の自動化された判断に依拠することが増えた。人間が何らかの”モノ”を利用する際には不確実性が伴うため、一定程度の信頼が必要になる。信頼の定義はさまざまであるが、本研究では、”不確実性があるのに何かに委任しようとする心理的な状態”[中谷内 08]と定義する。今後、人間が AI を利用していくために、人間の AI に対する信頼を高めることが重要になると考えられる。

多くの先行研究によって、能力と誠実さ及び価値の共有が人間同士の信頼を決める主要因だとする見解が支持されている。ここでの価値共有とは、”自身にとって重要な目的や、その目的を果たすまでの過程、あるいはその目的の背景にある問題の見方を共有している”[Earle & Cvetkovich 95]状況である。価値共有のなかには、秘密を共有することも含まれると言ってよいだろう。秘密を共有することで、ウチとソトとの間に境界が設定され、信頼が高まるという[小此木 80]。すなわち、本来他者に漏らす必要のない秘密を相手に漏洩する行為は、相手を信頼しているサインであり、秘密を打ち明けられた側は自分が信頼されていることを知り、相手に対してより大きな信頼を抱くのである[正村 95]。

さらに、AI 搭載ロボットや機械を対象にした研究も行われており、人間と同様に能力と誠実さが AI への信頼に影響することが示唆されている。しかしながら、価値共有に焦点を当てた研究はほとんど存在していない。無論、秘密の共有に関しても同様である。一方で、人間はロボットやコンピュータといったものに対して、対人間と同じ反応を示すことが多くの研究からも明らかとなっている[Nass & Moon 00]。そうであるならば、相手が AI で

あったとしても、秘密の共有が人間の AI に対する信頼に影響するのではないかと考えられる。

以上の議論を踏まえると、秘密の共有が人間の AI に対する信頼を高めるとの仮説が導き出せる。

また、[横井・中谷内 18]は本稿と同様の問題意識で研究を行っている。彼らは、人工知能に対する信頼度を測る際、場面想定法を用いているものの、これは、医療現場を想定したものであり、限定的である。筆者らはゲーム理論を用いて、より普遍的な人工知能に対する信頼の研究を実施した。ゲーム理論の有用性について、[田口 15]では、「ゲーム理論は、複数プレイヤーによる相互依存的な意思決定問題の帰結を均衡ないし解明する分析ツールであり、経済学や政治学においても、制度選択や制度比較の分析をするのに用いられている。というのも、ゲーム理論は、ある社会のしくみを抽象化し、かつ多くの選択肢の中の 1 つとして捉えることで、制度比較を容易になしうという特徴を有しているからである(ゲーム理論の制度比較可能性)」と述べられている。

これらの点を踏まえ、本研究では秘密の共有が人間の AI に対する信頼を高めるとの仮説を検証するために、信頼ゲームを用いた調査を行った。

2. サーベイ実験

2.1 実験計画

(1)目的

本実験では、人と人工知能の関係において、秘密の共有が人々の人工知能に対する信頼を高めるかどうかを検討する。

(3)回答者

サーベイ実験は、私立大学生 230 名を対象に行われた。回答者の平均年齢は 20.6 歳(SD=1.26 歳)、男性 115 人、女性 110 人、その他 5 人だった。

(4)実験デザイン

実験の目的に対応させ、独立変数を秘密の共有の有無(共有条件と非共有条件の 2 水準、参加者間要因)とし、従属変数を

連絡先: 永田大貴,同志社大学商学部商学科, 京都府京都市上京区今出川通烏丸東入, 075-251-3200, tagulabo@gmail.com

信頼ゲームにおける配分者から受益者への配分額とするサーベイ実験を実施した。

(5)手続き

サーベイ実験は、オンライン調査サイトである google form を用いて行われた。質問項目は 2 つあり、1 問目に信頼度の測定のための信頼ゲーム、2 問目に年齢・性別のステータスを聞いた。

(6)信頼度の測定:信頼ゲーム

1 問目には、信頼ゲームを設定し信頼度を測定した。なお、質問文の文言は[Berg et al. 95]を参考に作成した。配分額は、0~1000 円まで、100 円間隔で選択できるように設定した。

(7)条件(condition)の設定

サーベイ実験の条件として 2 条件(秘密を共有する条件と共有しない条件)を設定した。両者の違いは、シナリオ中盤部分の、画像に表示した 6 枚のトランプから 1 枚を選択させる質問と「あなたが選んだトランプは人工知能とあなただけの秘密です。」という一文の有無だけである。

2.2 実験結果

各条件の信頼度（信頼ゲームにおける配分者から受益者への配分額）について分析を行う。ただし、Figure1.そして Figure2.から、各条件ともに配分額が正規していないことがわかる。そのため、分析は正規分布を前提としないマン=ホイットニーの U 検定を行った。¹

平均配分額の結果は、秘密を共有する条件では 632.03 円、秘密を共有しない条件では 528.43 円であった。

(Figure.3)統計的にも両者の有意さが見られた。(U=5469.5, Z=2.177, P=0.029, 0.01<P<0.05)

つまり、秘密の共有をすることによって人と人工知能の間で信頼度が高まることが示唆された。

3. 考察

本研究での秘密の内容は文章や文言によるものではなく、トランプという数字や図画といった抽象的で個人の解釈の差が生まれにくいものを採用した。これは、[小此木 80]の、“秘密は、秘密にされる内容の如何によってではなく、むしろ、何ごとかを「秘密にしよう」とし、その「秘密を保とう」とする主体の「意思」の働きを、その本質としている”という言葉に基づいたものである。このことから、本研究結果はある程度の一般性を持つことが保証されると考える。しかしながら、被験者は大学生に限定されていたので、実験結果の外的妥当性に関しては慎重に勘案していく必要がある。

また、本研究において、実験結果から秘密を共有することが人間と同様に人工知能に対しても信頼度を高めることが示唆された。人間同士の信頼関係構築に関する先行研究にもあったように、人工知能に対しても秘密を共有することで親密さが生まれたためであると考えられる。これから来る人工知能時代には、iPhone の Siri 機能やネット上で採用されている様々なレコメンド機能のような身近なものはもちろん、医療や投資といった分野でのめざましい進化に対してもその信頼性を高めて利用できるようになるならば、その貢献性を高められると考える。

本研究での限界としてあげられる点は 2 つある。1 つはデータの信頼性である。本研究のサーベイ実験において組

み込んだ信頼ゲームは、本来実験室で参加者にゲームの結果に応じ謝礼を支払って行うものである。しかし本研究ではあくまで信頼ゲームを組み込んだサーベイ実験を実施したに過ぎないため、実験室実験と比較してその信頼性は劣ると考えられる。もう 1 つは実験パターンの偏りである。本研究のサーベイ実験は、共有する秘密は配分者側から与えるパターンのみを設定して実施したが、これが受益者側から与えられるパターンも無論考えられるため、その検証も今後必要になるだろう。

参考文献

[Berg et al 95] Berg, Joyce, dickhaut, John, McCave, Kevin (1985) Trust, reciprocity, and social history. *Games and economic behavior* 10 (1), pp.122-142

[Earle & Cvetkovich 95] Earle, T. C. & Cvetkovich, G. T. (1995). *Social trust: Toward a cosmopolitan society*. Greenwood Publishing Group.

[Nass & Moon 00] Nass, C. & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *The Journal of Social Issues*, pp.56,81-103.

[小此木 80] 小此木啓吾(1980)『笑い 人みしり 秘密-心理現象の精神分析』創元社.

[田口 15] 田口聡志(2015)『実験制度会計論: 未来の会計をデザインする』中央経済社.

[中谷内 08] 中谷内一也(2008)「リスク管理機関への信頼:SVSモデルと伝統的信頼モデルの統合」社会心理学研究,第 23 巻,第 3 号,pp.259-268.

[正村 95] 正村俊之(1995)『秘密と恥』勁草書房.

[横井・中谷内 18] 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,第 34 号,第 1 号,pp.16-25.

図表

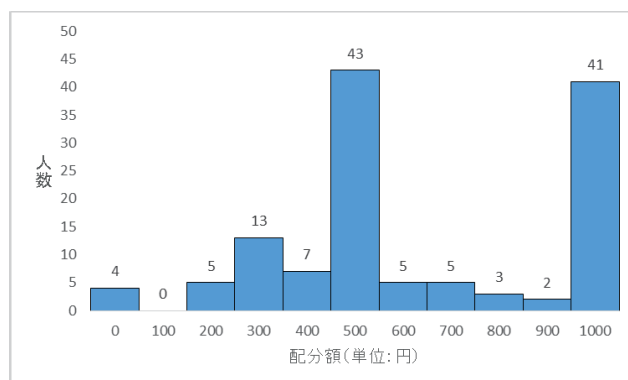


Figure1.配分額のヒストグラム（秘密の共有をする条件）

¹分析に際し、<http://www.gen-info.osaka-u.ac.jp/MEPHAS/wilc1.html> (最終閲覧日:2019年2月7日)を使用した。

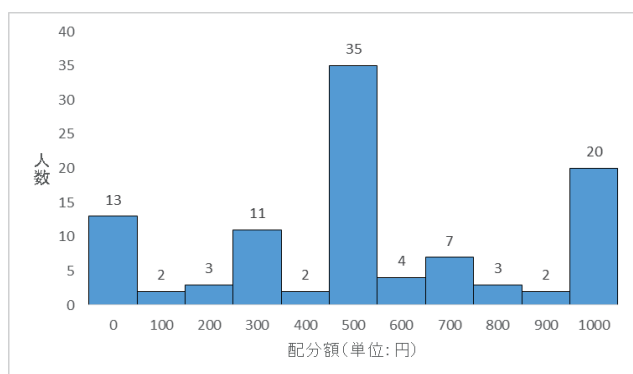


Figure2.配分額のヒストグラム（秘密の共有をしない条件）

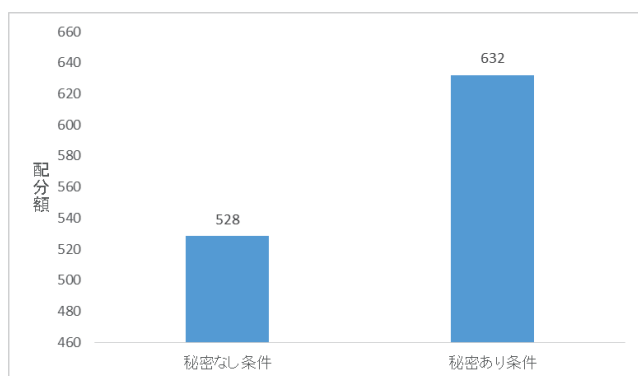


Figure3.各条件における配分額の平均値

本研究で用いられたシナリオ

シナリオ前半(全ての条件で共通)

<ルール>

これから2人1組でゲームを行います。2人にはそれぞれ役割があり、1人は送り手、もう1人は受け手です。送り手には現金1000円が与えられます。送り手は1000円の中の好きな額を受け手に配分することが出来ます。配分されたお金は3倍に増やされてから受け手に渡されます。その後、受け手は3倍にされた金額のうち好きな額を送り手に返すことが出来ます。

<役割>

あなたの役割は「送り手」です。「受け手」となるのは人工知能です。この人工知能は過去に行われたゲームの結果に基づいた合理的な判断を下します。

シナリオ後半(秘密の共有を行う条件)

<条件>

次に表示される6枚のトランプから1枚を選択してください。選んだトランプは人工知能とあなただけの秘密です。

<配分額>

あなたは1000円のうちいくらを相手に渡しますか。

シナリオ後半(秘密の共有を行わない条件)

<配分額>

あなたは1000円のうちいくらを相手に渡しますか。