

目標と手段の共有が人工知能への信頼に及ぼす影響

米杉拓真^a 佐野実沙^a 小西桃香^a 一山了^a

要約

本論文の目的は、目標共有と手段共有の違い、及び両方共有が情報提供者に対する信頼に影響するかどうかを検討するものである。堀江・犬塚・井川（2009）によると組織目標のコミットメントと自由度の付与が内発的モチベーションに影響を与えるとされている。そのため目標共有と手段共有を両方行うことで、自由度が付与されず、内発的モチベーションが低下し信頼が低下すると考えた。情報提供者と価値分配条件の違いを考慮して273名の参加者を対象に2要因計画で調査を行った。その結果、情報提供者の違いに関わらず、目標と手段の両方を共有する場合が最も信頼度が高くなり、仮説は全て棄却され意図せぬ結果となった。情報提供者が人の場合における差は有意に見られたことから人には両方を共有する方が良いと解釈できる。この点は更なる検証が求められるが、本研究の結果は企業組織においてAIと人間との協働を考えるうえでも一定の示唆があるといえる。

JEL 分類番号： D91, O33, J24

キーワード：人工知能, 信頼, 価値共有, 内発的動機付け

^a 米杉拓真 同志社大学商学部 tagulabo@gmail.com

^a 佐野実沙 同志社大学商学部

^a 小西桃香 同志社大学商学部

^a 一山了 同志社大学商学部

1. イントロダクション

現在の AI 技術は様々な分野において進歩しており、未来社会を研究するうえでは AI という要素を考慮しなければならない。特に今後、AI が社会で普及するためには、人々が技術を受け入れることが求められ、それは人々との信頼と強く紐付いているといわれる Lee & See (2004)。したがって、企業組織において AI の普及が進んでいる状況では、人々が信頼できる AI の運用方法を確立することが喫緊の課題である。本稿では、経営の目標とそれを達成するための手段又はこれら両方を AI と共有することが信頼に影響を与えるのかを把握する。これにより、企業組織で働く人々が AI 技術を受容する基盤の形成に資することができる。企業組織で AI が普及している現状から人々が信頼して働けるような AI の運用方法を確立する必要があり Lee and See (2004)、AI と経営の目標を共有するか目標を達成するための手段を共有するかで信頼が変わるのかを検証した。

信頼の定義は、多くの先行研究では能力と動機づけ（誠実さ）が要因となって信頼が決まるとされているが、これら以外にも、中谷内 and Cvetkovich (2008) は、能力と誠実さで信頼を決定するモデルを伝統的信頼モデルとする一方、人は相手の主要価値が自分のものと類似していると認知するとき相手を信頼するモデルも存在するとし、主要価値とは提示されたリスク問題の見立て方や、そこで何を重視し、どのような結果を選好するかであるとしている。これは人に対する信頼の前提だが、Nass and Moon (2000) では、人はロボットや機械に対しても人に対する反応と似た反応をすることを明らかにしており、AI との価値共有を検証している研究として Verbene et al. (2012) は、自動運転自動車の運転スタイルを被験者が望むものと共有する実験を行い、価値共有が信頼を高めると述べている。したがって、AI に対しての信頼も人と同様に能力、誠実さ、価値共有認知が決定要因となると定義する。

横井・中谷内 (2018) は、治療方針（手段）の共有が AI への信頼を高めるのかサーベイ実験を用いて検証しており、実際に信頼の高まりに影響を与えていた。Yokoi and Nakayachi (2019) は、AI と投資戦略（目標）を共有することの信頼への影響について実験しているが、影響は確認されなかった。

そこで Yokoi and Nakayachi (2019) では、目標と手段の共有の違いによって信頼に与える影響が変わってくるのではないかということが示唆されていることと横井・中谷内 (2018) では、AI 条件に比べ、医師条件の方が信頼の効果が大きく出ていることから、前回の研究では目標又は手段を AI 又は人と共有するときの信頼を検証した。その結果、人とは目標共有を、AI とは手段共有を行うほうが信頼されるという事がわかった。そして、本研究では、目標と手段の両方を共有すれば信頼がより促進されるかもしれないと考えたが、堀江・犬塚・井川 (2009) によると組織目標のコミットメントと自由度の付与が内発的

モチベーションに影響を与えるとされているため、目標共有と手段共有を両方行うことで、自由度が付与されず、内発的モチベーションが低下すると考えた。また、DR. Okello, L. Gilson (2015), 堀江・犬塚・井川 (2007) によると信頼関係が動機づけに影響を与えることが示されていることから、両方共有することで自由度が下がり、内発的モチベーションが下がれば信頼の促進が低下するといえる。以上の記述から、以下の3つの仮説を設定し場面想定法を用いたサーベイ実験で検証する。

仮説1：人との両方共有の方がAIとの両方共有より信頼が高い。

仮説2：人と目標を共有する方が両方を共有するより信頼が高い。

仮説3：AIと手段を共有する方が両方を共有するより信頼が高い。

2. 実験

本実験では、目標共有と手段共有の違いが情報提供者に対する信頼に影響するかどうかを検討する。実験実施期間は、2020年9月3日から9月23日の間であり、最終的なサンプル数は、273となった。被験者の平均年齢は、25.0歳(SD = 11.0歳)、男性148名、女性125名であった。

本研究では、独立変数を価値共有条件(目標共有条件、手段共有条件、両方共有条件の3水準、参加者間要因)と情報提供者(AI条件と人条件の2水準、参加者間要因)とする2要因計画で実験を実施した。従属変数は、情報提供者への信頼であった。

実験は、web調査サイト Google フォームを用いて行われた。本研究では、全6パターンのシナリオを用意している(付録A参照)。シナリオは前半と後半に分かれていた。シナリオの前半を読んだ後、被験者は、自分が望ましいと考える経営の在り方として、「利益を多く見込むが、比較的タイトな予算計画(業務改善策または両方)」と「利益見込みはわずかだが、比較的余裕のある予算計画(業務改善策または両方)」のどちらか一方を選択するよう求められた。ここで、企業活動は予算に沿って実行されることから、経営における目標としての位置づけにあると考えられる一方、手段は目標を達成するために行われるものであり、例えば業務改善策が挙げられることから本研究のシナリオに採用している。経営方針の選択後、目標共有条件では、シナリオに登場するAI(あるいはアドバイザー)が被験者の望ましいと考える予算計画を、手段共有条件では被験者の望ましいと考える業務改善策を、両方共有条件ではその両方を提案するというシナリオの後半が提示された。シナリオをすべて読み終えた後、従属変数を測定するための質問に移った。測定は、シナリオ読後の質問では情報提供者への信頼、能力認知、動機づけ認知、価値共有認知が尋ねられた。

これらの質問は、「1: 全くそう思わない」から「7: 非常にそう思う」までの7件法リッカード・スケールで回答するものだった。具体的な質問項目は表1に示す。質問項目の文言は横井・中谷内（2018）、Yokoi and Nakayachi（2019）を参考にした。

表1 実験で用いた質問項目と各項目の信頼係数		表2 各条件における記述統計量					
信頼	$\alpha=.82$	サンプル数	信頼	能力認知	動機付け認知	価値共有認知	
・このAIシステム(アドバイザー)を信頼できる		AI・目標共有	47	4.33(1.29)	4.94(1.35)	4.89(1.51)	4.58(1.39)
・このAIシステム(アドバイザー)は頼りになる		人・目標共有	42	4.11(1.28)	4.07(1.24)	4.32(1.31)	4.56(1.35)
・このAIシステム(アドバイザー)に判断を委ねることができる		AI・手段共有	46	4.49(1.38)	4.96(1.32)	4.59(1.38)	4.68(1.35)
価値共有認知	$\alpha=.77$	人・手段共有	45	4.36(1.20)	4.33(1.38)	4.63(1.29)	4.67(1.42)
・このAIシステム(アドバイザー)と私は経営について同じ意見である		AI・両方共有	50	4.58(1.49)	4.89(1.52)	4.72(1.67)	4.66(1.50)
・このAIシステム(アドバイザー)と私は経営について最も重要な考えを共有している		人・両方共有	43	4.60(1.36)	4.54(1.47)	4.71(1.48)	4.98(1.47)
・このAIシステム(アドバイザー)と私は経営についての考えは、私の好み合っている							
能力認知	$\alpha=.83$						
・このAIシステム(アドバイザー)は高い専門性を持っている							
・このAIシステム(アドバイザー)は豊富な知識を持っている							
・このAIシステム(アドバイザー)は有能である							
動機付け認知	$\alpha=.65$						
・このAIシステム(アドバイザー)は業務に一生懸命である							
・このAIシステム(アドバイザー)は誠実さを持っている							
・このAIシステム(アドバイザー)は公正に判断する							

※括弧内は標準偏差

3. 結果

各条件における記述統計量を図2に示す。図2から、傾向として、情報提供者の違いにかかわらず、目標と手段の両方を共有する両方共有を行うほうが信頼されることがわかる。

ここで、両方共有を行う場合の情報提供者の違いによる差の検定を行ったが、有意差は見られなかった($U=1046(n_1=50, n_2=43)p=0.825$)。また、情報提供者が人の場合における目標共有と両方共有の違いによる差を検定したところ、こちらは5%有意がみられた($U=676(n_1=42, n_2=44)p<.05$)。さらに、情報提供者がAIの場合における手段共有と両方共有の違いによる差の検定を行ったが、有意差は見られなかった($U=676(n_1=46, n_2=50)p=0.432$)。このことから仮説1、仮説2、仮説3すべての仮説において棄却され、意図せざる結果となった。

4. 考察

本研究において、両方を共有する場合が最も信頼の値が高くなった理由としては、両方を共有した場合において、自由度の低下はなかったと考えられる。つまり、予算計画とそれに対する業務改善案を作成するというシナリオでは自由度の低下はなかった可能性がある。しかし、情報提供者が人の場合における目標共有と両方共有の違いによる差は有意に見られたことから、人には両方を共有するほうが良いと解釈できる。この結果は、企業組織においてAIと人間との協働を考えるうえでも一定の示唆があるといえる。信頼度の高い組織で働く従業員は、生産性が高く、業績向上につながるということがいわれているZak(2017)。このことから、企業組織内で人とAIとで適当な情報を共有することにより、人から人へ、あるいは人からAIへの信頼を促進し、企業パフォーマンスの向上にポジティブな影響を与える可能性がある。この点については、さらなる検証が求められる。しかし、本研究において、目標共有または手段共有における情報提供者の違いによる差は、先行研

究である磯川・米杉・石田・佐野・小西・一山（2020）とは異なり、情報提供者が人の場合においても手段共有のほうが高い数値となったため、情報提供者が人の場合において、目標の共有を行うよりも、両方の共有を行うほうが信頼を促進させるとはならない可能性がある。また、異なった結果になった要因として考えられるものは2点ある。1点目は、平均年齢の上昇によるものである。補助分析として、従属変数を信頼度、独立変数を年齢、性別、AIへの関心度、企業経営への関心度として重回帰分析を行った結果、年齢に10%有意がみられた。そのため、平均年齢が高いシナリオの信頼度が高くなり、先行研究と結果が異なってしまった可能性がある。また有意差は出ないもののAIの関心度も信頼に影響する傾向が見られる。2点目は、共有状況の可視化による影響である。E. Mayr et al. (2019)では、情報の可視化により、その情報の利用に何らかのリスクが伴う場合には信頼に関連すると述べられており、今回のシナリオでは、予算計画、業務改善案のどちらか、あるいはその両方の共有内容の画像を添付し、可視化した。その結果、目標共有よりも手段共有を行った場合のほうが信頼度が高くなっていたため、手段共有のほうが利用にリスクが伴うと考えられた可能性がある。

5. 本研究の限界と今後の展望

第1の限界は、実験研究を採用したことに起因する外的妥当性の問題である。外的妥当性とは「実験を通して得られた情報や知見が、実験室という特殊な状況の範囲を越えて一般的な状況に当てはまる程度」のことをいう佐藤（2015）。本研究は、現実の企業組織を念頭において調査を行っているが、今回の結果がそのまま現実の環境を捉えているかどうかについては、一定の留保が必要である。第2の限界は、学生を主な被験者に用いていることである。この点は、企業で働く社会人を対象にサーベイ実験を行うことで、本研究の結果の説得力も高まると考えられる。これは、今後の課題としたい。今後の展望としては、年齢及びAIへの関心度が信頼に影響する傾向がみられるため、検討していきたい。

付録A 本研究で用いられたシナリオ

シナリオ前半(目標共有条件)

あなたは企業Aに所属しています。企業Aの置かれた業界では、近年、競争が激しさを増しています。あなたが所属する部署では、次年度の年次予算計画を立てるように指示されています。

シナリオ後半(AI条件)

あなたの所属する企業Aでは、AIシステムが導入され、ビッグデータ(数値やテキスト、音声情報などの多様な膨大なデータ)をチェック・分析し、経営の意思決定に活かしています。

AIシステムは過去のデータをもとに、より最適な予算計画(業務改善案または両方)を提案しようとしています。あなたには、あなたの考える予算計画(業務改善案または両方)とAIシステムの提案する予算計画(業務改善案または両方)を考慮し、最終的な予算計画(業務改善案または両方)を作成することが求められています。

シナリオ前半(手段共有条件)

あなたは企業Aに所属しています。企業Aの置かれた業界では、近年、競争が激しさを増しています。あなたが所属する部署では、次年度の年次予算計画を達成するための業務改善案を立てるように指示されています。

AIシステムは分析の結果、「利益を多く見込むが、比較的タイトな予算計画(業務改善案または両方)」(利益見込みはわずかながら、比較的余裕のある予算計画(業務改善案または両方))を提案しました。このAIシステムは正確な予測に努めますが、ビジネス環境は不確実であり、常にリスクを伴います。

シナリオ後半(A条件)

あなたの所属する企業Aでは、アドバイザーの助言を参考にして、経営の意思決定に活かしています。アドバイザーは過去の業績をもとに、より最適な予算計画(業務改善案または両方)を提案しようとしています。あなたには、あなたの考える予算計画(業務改善案または両方)とアドバイザーの提案する予算計画(業務改善案または両方)を考慮し、最終的な予算計画(業務改善案または両方)を作成することが求められています。

シナリオ前半(両方共有条件)

あなたは企業Aに所属しています。企業Aの置かれた業界では、近年、競争が激しさを増しています。あなたが所属する部署では、次年度の年次予算計画とそれを達成するための業務改善案を立てるように指示されています。

アドバイザーは分析の結果、「利益を多く見込むが、比較的タイトな予算計画(業務改善案または両方)」(利益見込みはわずかながら、比較的余裕のある予算計画(業務改善案または両方))を提案しました。このアドバイザーは正確な予測に努めますが、ビジネス環境は不確実であり、常にリスクを伴います。

引用文献

- Dickson, R. O. Okello, and L. Gilson, 2015. Exploring the influence of trust relationships on motivation in the health sector: a systematic review. *Human Resources for Health*
- E. Mayr, N. Hynek, S. Salisu, and F. Windhager, 2019. Trust in Information Visualization. *Euro Vis Workshop on Trustworthy Visualization*
- 堀江 常稔, 犬塚 篤, 井川 康夫, 2007. 研究開発組織における知識提供と内発的モチベーション. *Japanese Journal of Administrative Science* 20(1), 1-12
- 堀江 常稔, 犬塚 篤, 井川 康夫, 2009. 内発的モチベーションの知識創造行動への効果. *技術と経済* 504, 66-71.
- 磯川 雄大, 米杉 拓真, 石田 朋己, 佐野 実沙, 小西 桃香, 一山 了, 2020. 共有された目標と手段の違いが人工知能への信頼に及ぼす影響. *Annual Conference of the Japanese Society for Artificial Intelligence*
- Lee, J. D., and See, K. A., 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors* 46, 56-80
- Nakayachi, K., and G. Cvetkovich, 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* 23(3), 259-268
- Nass, C., and Moon, Y., 2000. Machines and mindlessness: Social responses to computers. *The Journal of Social issues* 56, 81-103
- 佐藤郁哉, 2015. 社会調査の考え方[下], 東京大学出版会, 東京
- Verberne, F. M., Ham, J., and Midden, C. J., 2012. Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors* 54, 799-810
- 横井良典, 中谷内一也, 2018. 治療方針の共有が人工知能の信頼に及ぼす影響. *社会心理学研究* 34(1), 16-25
- Yokoi, R., and Nakayachi, K, 2019. The effect of shared investing strategy on trust in artificial intelligence. *実験社会心理学研究* 59(1), 46-50
- Zak, P. J., 2017. The neuroscience of trust. *Harvard Business Review* 95(1), 84-90