

擬人化傾向及び情報科学知識が対話エージェントへの信頼形成に与える影響と対話方略の検討

The Influence of Anthropomorphization Tendencies and Information Science Knowledge on Trust Formation Toward Dialogue Agents and Examination of Dialogue Strategies

中間 音羽*
Towa Nakama

宮本 友樹
Tomoki Miyamoto

内海 彰
Akira Utsumi

電気通信大学

The University of Electro-Communications

Abstract: 本研究では、対話エージェントの家庭導入初期における信頼形成メカニズムを解明するため、ユーザーの情報科学知識量および擬人化傾向が信頼感に与える影響を多角的に検討した。クラウドソーシングを通じて募集した184名を対象に、コミュニケーションロボット Sota を用いた4つの対話戦略（社会的、知的、身体的、道徳的）を提示する被験者間実験を実施した。分析の結果、情報科学知識量と擬人化傾向の間には有意な負の相関 ($B = -0.07, p = .031$) が認められ、技術的リテラシーが高いほど対象を分析的に捉え、擬人化が抑制されることが示された。信頼形成プロセスにおいては、ユーザーの擬人化傾向と対話戦略の間に有意な交互作用 ($p < .001$) が確認された。具体的には、擬人化傾向が低い層では道徳的・知的戦略が有効であるのに対し、高い層では社会的・身体的戦略がより高い信頼を獲得するという逆転現象が明らかになった。一方で、誠実さやデータの透明性を強調する道徳的戦略は、ユーザーの特性に関わらず安定した信頼を得る普遍的な有効性を示した。また、科学的好奇心の分析では、日常的な行動頻度が信頼を促進する一方で、能動的な選択が信頼を抑制するという、興味に対する受容性と審美眼の二面性が示唆された。本研究の成果は、ユーザーの属性に応じたパーソナライズされた対話デザインの重要性を提示しており、AIの円滑な社会実装に向けた実践的な指針を与えるものである。

1 はじめに

近年のAI技術の進展に伴い、対話エージェントやコミュニケーションロボットが家庭環境へ導入される場面が増加している。これらと人間が健全に共生するためには、技術的な正確性のみならず、心理的な信頼関係の構築が不可欠である。特に、ロボットが初めて家庭に導入される初期段階において、ユーザーがどのように信頼を寄せるかは、その後の継続的な利用を左右する。

先行研究では、ロボットに対する信頼形成にはユーザーの個人特性、特に情報科学に関する知識量が重要であることが示されている。専門知識を持つユーザーは性能や透明性といった認知ベースの信頼を重視する一方、非専門的なユーザーは親しみやすさや人間らしい振る舞いといった感情ベースの信頼を形成しやすい傾向があ

る。しかし、知識量の相違がどのような人間らしさへの期待に変容し、いかなる対話方略が信頼形成に有効であるかというメカニズムは未だ十分に解明されていない。

本研究は、個人の情報科学知識量と擬人化傾向が信頼形成にもたらす影響を明らかにすることを目的とする。そのうえで、先行研究のレビューに基づいて設計した4つの対話方略の効果を検討する。なお、本研究は電気通信大学人を対象とする研究に関する倫理委員会の承認（管理番号：[H23071]）を得て実施した。

2 関連研究

2.1 人間とAIにおける信頼形成

人間とロボットやAIとのインタラクションにおいて、信頼は受容性を決定付ける中核的な概念である。Mayerら[1]は、人間同士の信頼関係を構成する要素と

*連絡先：電気通信大学 I 類 経営・社会情報学プログラム
〒182-8585 東京都調布市調布ヶ丘 1-5-1
E-mail: n2210461@gl.cc.uec.ac.jp

して、有能性 (Competence)、誠実性 (Integrity)、善意 (Benevolence) の三次元モデルを提案した。このモデルは、対話エージェントに対する信頼形成プロセスの基礎としても広く応用されている。さらに、Hengstlerら [2] は、ユーザの特性によって信頼の拠り所が異なることを指摘した。専門的な知識を持つユーザは、対象の技術的性能や論理的透明性に基づく認知ベースの信頼 (Cognitive-based trust) を重視する。これに対し、知識が乏しい一般的なユーザは、人間らしい振る舞いや外見の親しみやすさに起因する感情ベースの信頼 (Affective-based trust) を形成しやすい傾向がある。本研究の信頼形成プロセスに対する構想は、この知識量による信頼の判断材料の相違を前提として構築されている。

2.2 ユーザの個人特性：知識量と科学的好奇心

ユーザの独立変数としての知識量やリテラシーの影響についても、多くの知見が蓄積されている。Shahら [5] が提案した TOSLS (Test of Scientific Literacy Skills) は、科学的情報を批判的に評価する能力を測定するものであり、リテラシーの高さが情報の受容性に直結することを示している。さらに、Kahanら [6] は「科学的好奇心 (SCS)」の概念を提唱し、科学的な情報に対して自発的に触れようとする熱量が高い個人は、たとえ自身の信念に反するデータであっても客観的に受け入れる可能性が高いことを示した。また、Dennett [7] の意図的スタンスと設計的スタンスの理論は、対象を「心を持つもの」として捉えるか「設計された仕組み」として捉えるかの認知的境界を示唆しており、これがユーザの知識量に起因するという本研究の視点に関連している。

2.3 社会的・知的戦略の設計根拠：擬人化の理論的基盤

人間が非人間に対して人間的な特性を帰属させるプロセスは、単一の現象ではなく多層的な構造を持つ。本研究では、擬人化を「心の知覚 (Mind Perception)」と「心理的動機 (SEEK モデル)」の観点から定義する。

2.3.1 心の知覚の二次元モデル

Grayら [4] は、人間が他者に「心」を見出す際の次元として、主体性 (Agency) と体験性 (Experience) の2軸を提唱した。

- 主体性 (Agency)：計画を立てる、論理的に考える、自己を制御するといった「思考・実行」の能力に関連する。本研究の対話方略の一つである知的戦略は、エージェントをこの主体性を持つ存在として捉える人間に対して有効であると考えられる。
- 体験性 (Experience)：痛みや喜びを感じる、空腹を覚える、意識を持つといった「感情・感覚」の能力に関連する。本研究における社会的戦略の有効性は、エージェントにこの体験性を見出すことに相当する。

2.3.2 擬人化を誘発する心理的動機：SEEK モデル

Epleyら [3] によれば、擬人化は以下の3つの主要な動機によって引き起こされる。

- 社会性動機 (Sociality Motivation) 他者とのつながりを求める欲求。
- 効果動機 (Effectance Motivation) 自身の環境を理解・予測・制御したいという欲求。
- 誘発的擬人化 (Elicited Agent Knowledge) 自身の知識構造を対象に適用すること。

以上の理論を踏まえ、本研究ではユーザの信頼形成を促すためのアプローチとして擬人化の2つの側面に対応した対話方略を定義し、実験に採用する。

- 社会的戦略：エージェントを感情や共感能力を持つ「体験的な存在 (Experience)」としてユーザに認識させ、社会性動機に働きかける。
- 知的戦略：エージェントを高度な推論や計画能力を持つ「主体的な存在 (Agency)」として認識させ、効果動機に働きかける。

2.4 身体的戦略の理論的根拠：CASA パラダイムと身体性

ロボットの振る舞いが人間に与える心理的影響について、Nassら [8] はCASA (Computers are Social Actors) パラダイムを提唱した。これは、人間が相手を機械だと認識していても、無意識に社会的・擬人化的な反応を示してしまうという強力な知見であり、本研究における対話方略設計の理論的基盤となっている。この社会的反応をより強固にする要因として、身体性の重要性が指摘されている。Kieslerら [9] による身体化認知 (Embodied Cognition) の観点からは、視線や身振りといった非言語情報の提示が、エージェントの实在感と信頼性

を高めることが示されている。また、Schömbbsら [10]の研究では、画面内のエージェントと比較して、物理的な動きを伴うロボットの存在が信頼構築に直接的に寄与することが実証されている。さらに、Schreibelmayerら [11])は、音声の人間らしさ (Human-likeness) が特定の状況下における受容性に影響を与えることを報告している。これは、単なる外見的な擬人化だけでなく、多動的な動作と音声制御の動的な組み合わせが、信頼形成において相乗的な意義を持つことを示唆している。

以上のレビューに基づき、本研究では、エージェントの物理的実在感と動的な表出能力を最大限に活用する身体的戦略を定義、採用し、信頼形成における重要な対話方略の一つとして有効であるか実証を試みる。

2.5 道徳的戦略の設計根拠と理論背景

人間が対象を道徳的な文脈で捉える際、Grayら [12]が提唱した道徳的類型化 (Moral Typecasting) の理論が重要な役割を果たす。この理論によれば、人間は他者を道徳的行為者 (Moral Agent: 行動の責任を負う者) または道徳的享受者 (Moral Patient: 影響を受ける者) のいずれかとして認識する傾向がある。

Waytzら [13]の先行研究では、対話エージェントが高度な知性や予測可能性を示す場合、ユーザはエージェントを単なる道具ではなく、ある種の道徳的行為者として扱うことが示唆されている。エージェント自らがシステムの限界を認め、データの取り扱いや公平性に関する宣言を行うことで、ユーザに対して「このエージェントは自身の行動に責任を持つ誠実な存在である」という知覚を誘発する。

以上の理論背景をもとに本研究の最後の対話方略として「道徳的戦略」を定義、採用する。

情報技術への理解が乏しい低知識層においては、複雑なアルゴリズムに基づく能力を正しく評価することは困難である。これに対し、誠実性 (約束を守る、一貫した主義を持つ) の知覚は、専門的な知識を介さずとも、エージェントの言動の一貫性や倫理性から直感的に判断可能である。したがって、道徳的戦略は、高リテラシー層には論理的な安心感を、低リテラシー層には直感的な信頼感を与えるという、重層的な効果を持つと考えられる。

2.6 関連研究を踏まえて

本研究では、これら4つの方略を独立変数として実装し、ユーザの知識量との交互作用を明らかにすることで、ユーザの特性に応じた最適な信頼形成プロセスの解明を試みる。次章では、これらの方略を具現化し



図 1: 実験で使用した Sota の外観・対話例

た対話システムの実装および実験デザインについて詳述する。

3 実験方法

3.1 ユーザ特性 (独立変数・調整変数)

第一に、情報科学に関する客観的な知識量を測定するための指標として TOSLS[5] を用いる。第二に、人間以外の事象に人間的な性質を見出す傾向 (擬人化傾向) を測定するため、IDAQ (日本語版) を用いる。第三に、科学技術に対する日常的な関わりを評価するため、科学的行動尺度 (SCS) を用いる。SCS は、科学的な情報に触れる行動頻度と、科学的な分野を自ら選ぶ能動的選択の2つの次元から構成される。

3.2 信頼感 (従属変数)

エージェントに対する信頼の評価には、全体的な信頼感を測る TIAS および、機能的・対話的側面から多次元的に信頼を測る TPS-HRI を用いた。

3.3 実験システムと対話方略の設計

実験にはコミュニケーションロボット Sota を用いた。ロボットの制御には gRPC を用いた Python スクリプトを使用した。本研究では、CASA 理論及び信頼形成と擬人化に関する先行研究レビューに基づいて、以下の4つの対話方略を設計指針に基づいて実装した。

- 社会的戦略: 共感や親和性に基づく発話
発話例: 「朝だけでも少し気持ちが楽になるように、僕がお手伝いしましょうか」
- 知的戦略: 論理的かつ効率性を重視する発話
発話例: 「情報量を整理するだけで、行動開始までの所要時間は10~15%短縮されます」

- 身体的戦略：非言語動作、実在感の提示
発話例：「支度の流れを教えてくださいと、僕も動きを合わせやすくなります。」
- 道徳的戦略：誠実さや透明性を強調する発話
発話例：「予定データは外部に共有せず、処理はすべて本体内で完結します。不利益な操作は含みません。」

3.4 実験デザインと参加者

対話方略を操作変数とした4条件の被験者間計画を採用した。クラウドソーシングを通じて募集した184名（男性107名、女性75名、未回答2名、 $M_{age} = 42.4$ 、 $SD_{age} = 9.48$ ）を対象とした。

3.5 実験手順

実験参加者はGoogleフォームにて情報科学知識および擬人化傾向を問う質問に回答後、図1に示すようなロボットを家庭に導入する場面を想定した共通動画（約60秒）と割り当てられた発話戦略において特定の機能を提案する場面の動画（約90秒）を視聴した。最後に、提示されたロボットに対する信頼感や、提案への納得感に関する事後アンケートに回答し、一連の手続きを完了した。

4 結果・考察

4.1 操作チェック

各実験条件が参加者に設計意図通り知覚されているかを確認するため、操作チェック項目について条件間での1要因分散分析を行った。その結果、すべての質問項目において対話方略の提示は適切に行われており（ $p = .006$ ）、参加者は各戦略の違いを正しく認識していたことが確認された。

4.2 共分散分析

対話エージェントの対話方略およびユーザの個人特性が信頼感に与える影響を検討するため、対話方略を独立変数、TIASを従属変数、個人特性（TOSLS, IDAQ, SCS）およびその交互作用を共変数とした共分散分析を行った。各個人特性の影響を統計的に統制した上で、対話方略が信頼感に与える主効果を検討した。分析の結果、対話方略の主効果は有意ではなかった（ $F(3, 166) = 1.652, p = .179, \eta_p^2 = .029$ ）。このことから、個人特性の影響を統計的に統制した条件下において、対話

方略のみでは信頼感に有意な影響を及ぼさないことが示唆された。

4.3 重回帰分析

対話エージェントへの信頼感に影響を及ぼす要因を包括的に検討するため、対話方略、個人特性（TOSLS, IDAQ, SCS）、および交互作用（TOSLS, IDAQ, 対話方略）を投入した重回帰分析を行った。分析の結果、決定係数は $R^2 = .176$ （調整済み）となり、モデル全体として有意であった（ $F(7, 175) = 5.35, p < .001$ ）。

特筆すべき点は、SCSを構成する2つの次元、すなわち行動頻度と能動的選択が、信頼感に対して正反対の影響を与えていたことである。科学的行動の頻度は信頼感に対して有意な正の影響を示し（ $\beta = .239, t = 2.995, p = .003$ ）、一方で能動的選択は有意な負の影響を示した（ $\beta = -.243, t = -2.996, p = .003$ ）。情報科学知識量は信頼感を有意に負に予測した（ $\beta = -.201, t = -2.431, p = .016$ ）。

対話方略と擬人化傾向においては、両者の交互作用項において有意な負の影響が認められた（ $\beta = -.235, t = -3.380, p < .001$ ）。この交互作用は、採用された対話方略が信頼感に与える効果が、ユーザの擬人化傾向の強弱によって異なることを示唆している。

有意な交互作用が認められた対話方略×IDAQについて、単純傾斜分析を行った結果を図2に示す。本実験では、擬人化傾向の強弱によって、信頼を獲得するのに有効な対話方略が完全に逆転するという結果が得られた。

次に、情報科学知識と信頼感の関連について検証した結果を図3に示す。IDAQが低い群（ $-1SD$ ）においては、TOSLSスコアが高いほど信頼感が低下する有意な負の傾斜が認められた。これに対し、IDAQが高い群（ $+1SD$ ）では、TOSLSスコアが高いほど信頼感が高まる正の傾斜が示された。

4.4 考察

科学的行動尺度において、行動頻度が信頼感に正の影響を与える一方、能動的選択が負の影響を与えるという対照的な結果が示された。これは、日常的に科学に接している層は対話エージェントに対しても受容的であるが、自律的に情報を取捨選択する層は、エージェントの提示する情報の妥当性を厳格に評価するため、「専門知識が批判的検討を促す」という先行研究の知見と整合する。

単純傾斜分析により、擬人化傾向×対話方略、情報科学知識量×擬人化傾向が明らかとなった。擬人化傾向が低いユーザにおいては、倫理的な発話を行う道徳

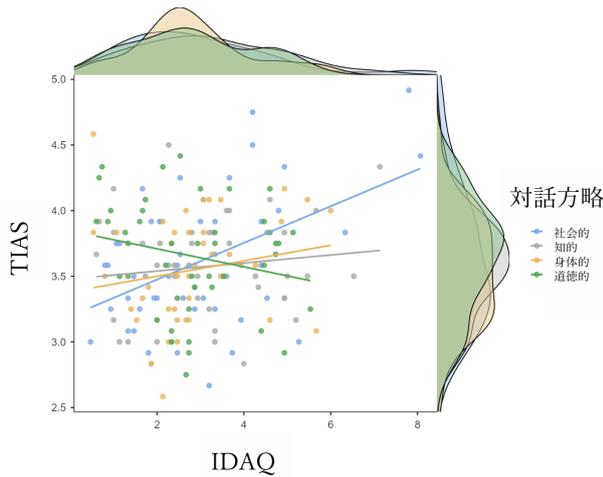


図 2: 対話戦略と擬人化傾向の交互作用

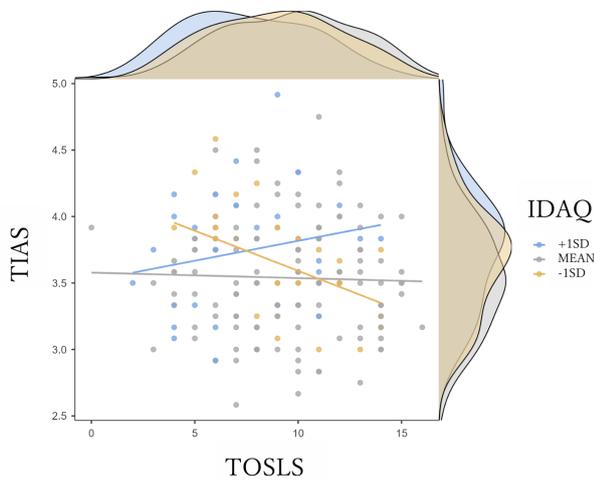


図 3: 情報科学知識量と擬人化傾向の交互作用

的戦略よりも、共感を示す社会的戦略をとる場合に信頼感が低下する傾向がみられた。これは、対話エージェントを道具として捉えるユーザにとって、過度な感情表現はかえって不自然さや不信感を招く不気味の谷に近い現象が生じている可能性を示唆している。

また、擬人化傾向が低いユーザは知識が高いほど AI を批判的に評価する一方で、擬人化向が高いユーザは知識が高いほど AI の提示する論理を好意的に受容し、信頼を寄せるという対照的な調整効果を示唆している。

5 おわりに

本研究では対話ロボットにおいて、対話戦略とユーザの特性が信頼感に及ぼす影響を明確化することを目的とした。その結果、個人の知識量や擬人化傾向によって信頼を得やすい対話方略が異なるという知見が得ら

れた。

謝辞

本研究は、JSPS 科研費 (JP23K16923) の支援を一部受けました。記して感謝いたします。

参考文献

- [1] Mayer, Roger C., Davis, James H., Schoorman, F. David, 「An Inte-grative Model of Organizational Trust」, *Academy of Management Review*, Vol.20, No.3, pp.709-734, 1995.
- [2] M. Hengstler, E. Enkel, and S. Duerr, "The role of trust in the adoption of intelligent assistants and smart personal devices," J. Bus. Res., vol. 69, no. 8, pp. 2848-2856, Aug. 2016.
- [3] N. Epley, A. Waytz, and J. T. Cacioppo, "On seeing human: A three-factor theory of anthropomorphism," Psychol. Rev., vol. 114, no. 4, pp. 864-886, Nov. 2007
- [4] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," Science, vol. 315, no. 5812, p. 619, Feb. 2007
- [5] C. Gormally, P. Brickman, and M. Lutz, "Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments," CBE—Life Sci. Educ., vol. 11, no. 4, pp. 364-377, Dec. 2012
- [6] D. M. Kahan, A. Landrum, K. Carpenter, L. Helft, and K. H. Jamieson, "Science Curiosity and Political Information Processing," Polit. Psychol., vol. 38, no. S1, pp. 179-199, Feb. 2017
- [7] D. C. Dennett, The Intentional Stance. Cambridge, MA, USA: MIT Press, 1987.
- [8] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94), Boston, MA, USA, Apr. 1994, pp. 72-78
- [9] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic interactions with a robot and robot-like agent," Soc. Cogn.,

vol. 26, no. 2, pp. 169–181, Apr. 2008, doi: 10.1521/soco.2008.26.2.169

- [10] S. Schömbbs, J. Klein, and E. Roesler, "Feeling with a robot—the role of anthropomorphism by design and the tendency to anthropomorphize in human-robot interaction," *Front. Robot. AI*, vol. 10, Art. no. 1149601, Jun. 2023, doi: 10.3389/frobt.2023.1149601
- [11] S. Schreibelmayr and M. Mara, "Robot Voices in Daily Life: Vocal Human-Likeness and Application Context as Determinants of User Acceptance," *Front. Psychol.*, vol. 13, Art. no. 787499, May 2022, doi: 10.3389/fpsyg.2022.787499
- [12] K. Gray and D. M. Wegner, "Moral typecasting: Divergent perceptions of moral agents and moral patients," *J. Pers. Soc. Psychol.*, vol. 96, no. 3, pp. 505–520, Mar. 2009, doi: 10.1037/a0013748
- [13] A. Waytz, J. Cacioppo, and N. Epley, "Who sees human? The stability and importance of individual differences in anthropomorphism," *Perspect. Psychol. Sci.*, vol. 5, no. 3, pp. 219–232, May 2010, doi: 10.1177/1745691610369336.