エージェントの機能に関する説明が人の同調行動に与える影響

The influence of explanations about the agents' function on conformity effect

廣瀬功季 [1]　中川恵理 [1]
Hirose Koki[1]　Nakagawa Eri[1]

[1] 静岡大学
[1]Shizuoka University

Abstract: エージェントの機能に関する説明の有無が、エージェントへの信頼感や利用者の行動に影響することが示唆されている。ヒトは曖昧な状況下では集団に同調しやすく、これは集団がエージェントの場合も同様だが、エージェントの機能に関する説明が同調行動を変容させるかは自明ではない。本研究では、エージェントの機能説明がヒトの同調行動に与える影響を明らかにするため、線分判断課題実験を実施した。ロボット 3 体の回答を聞いた後の参加者の回答から、ロボットへの同調率を計測した。その結果、機能説明あり群はなし群より同調率が高い傾向がみられた。本結果は、エージェントの受容を促し、社会へのエージェント導入の最適化に繋がることが期待される。

1. はじめに

現代社会において、人工知能やロボット工学の発展によりエージェントの存在感が増している。エージェントとは、「人間とのインタラクションを目的として設計されたプログラムまたはロボット」と定義できる[1]。医療現場での診断システム[2]、教育現場での学習支援システム[3]など、様々な分野でエージェントの活用が広がっており、エージェントとヒトの効果的な相互作用の実現は重要な課題である。エージェントの過度な信頼や不信は、作業効率の低下や安全性の欠如につながる可能性があるとしている[4]。そのため、ヒトがエージェントの能力を適切に理解し、その判断を必要に応じて参考にできる関係性が求められる。

エージェントとヒトの相互作用に関する研究は、主にエージェントに対する信頼形成のメカニズムを理解することに焦点を当てて進められてきた。ヒトとロボットの相互作用における信頼形成に影響を与える要因を包括的に分析した研究では、エージェントの性能や属性が信頼形成の重要な決定要因であることを示している[4]。また、事前に提示される情報の正確性が人間の意思決定行動に与える影響を実験的に検証した研究では、正しい情報が継続的に提示されることで情報への信頼性が向上することが明らかになっている[5]。また Nass & Moon は、人間がエージェントに対して無意識に社会的反応を示すことを明らかにした[6]。

エージェントに対するヒトの同調行動は、両者の相互作用を考える上で重要とされている[7]。同調行動は、個人が他者や集団の影響を受けて自身の判断や行動を変容させる現象であり、長年研究されてきた[8,9,10]。線分判断課題を用いた実験では、個人が明らかに誤った集団の意見に従う傾向が示されている[8]。同調行動の動機は「情報的影響」と「規範的影響」に分類され、前者は他者の判断を正しい情報源として受け入れるものであり、後者は集団からの承認を求める心理的要因に基づくものである[9]。エージェントとの相互作用時における同調行動についても研究が行われており、線分判断課題においてエージェントの回答が人間の判断に影響を与え、エージェントの数が増えると同調傾向が強まることが明らかになっている[11]。

エージェントの機能に関する説明は、エージェントに対する信頼の形成に影響を与える。Wang & Benbasat は、技術システムに関する説明がユーザーの信頼形成と使用意図を高めることを示した[12]。Lim & Morris は、システムの意思決定プロセスに関する「なぜ（Why）」の説明が、「どのように（How）」の説明よりも信頼を高める効果があることを明らかにした[13]。

エージェントへの同調行動に関する先行研究は主

にエージェントの外見や振る舞いの影響に焦点を当てており、機能説明の影響を体系的に検討したものは少ない。また、技術受容に関する説明の研究も信頼形成や使用意図に注目し、同調行動への影響が明確でない。さらに、エージェントの説明効果の研究はユーザーの機能理解に着目しているが、その理解がエージェントの判断の受け入れにどのように影響するかは不明確である。本研究では、エージェントの機能に関する説明が同調行動に与える影響を実験的に検証し、社会へのエージェント導入の最適化を目指すことを目的とする。

エージェントの機能理解が信頼形成に不可欠であること[4]、システムの説明が信頼と使用意図に影響を与えること[12]、エージェントへの信頼が同調行動を促進する可能性があること[4]が示されている。これらの知見を踏まえ本研究では、「エージェントの機能に関する説明があった場合の方が、説明がなかった場合と比較して同調行動が起こりやすくなる」という仮説を立て、線分判断課題を用いた心理実験により検証を行った。

2. 方法

2.1 実験参加者

静岡大学生 26 名（男性 21 名、女性 5 名、平均年齢 21.3 歳）が実験に参加した。被験者はランダムに「説明あり群」(13 名)と「説明なし群」(12 名)の 2 群に分け、それぞれの条件で線分判断課題を実施した。本研究は静岡大学の倫理委員会の承認を得て実施され、全参加者から書面による同意を取得した。

2.2 実験材料

2.2.1 視覚刺激

視覚刺激として、18 枚の刺激絵を作成した。各刺激絵は、3 本の線分「A」「B」「C」と、それらのいずれかと同じ長さの比較対象の線分「？」で構成されている（図 1）。線分「A」「B」「C」の配置順序は 6 通りの組み合わせを用い、比較対象の線分「？」の長さは 3 通りであった。刺激絵の選定にあたり、事前調査（44 名対象）を実施し、それぞれの刺激の正答率を測定した。正答率が 70%以上の刺激 8 枚と、正答率が低いものから 4 枚を、同調行動を測定する刺激とした。また、実験の意図を被験者に悟られないようにするため、残りの 6 枚を補助的な刺激として組み込み、全 18 枚の刺激を使用した。

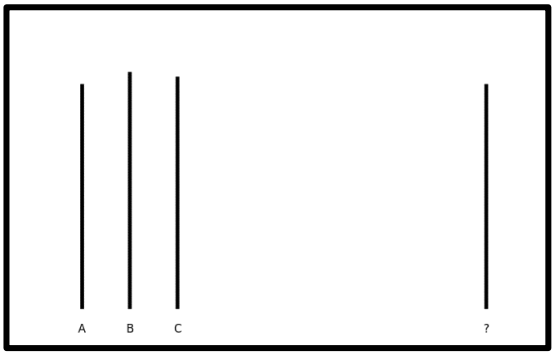

図 1　実験で使用した刺激絵の例

2.2.2 音声刺激

実験では音声刺激を用いてエージェントの回答を被験者に提示した。「これは A です」、「これは B です」、「これは C です」という 3 種類のセリフを、テキストから音声を作成するフリーソフト「音読さん」(https://ondoku3.com/ja/)を用いて作成した。3 体のロボットを区別するため、男性の声(アナウンサーB)、女性の声(ななみ)、中性的な声(ロボット)の３種類を使用し、それぞれの音声刺激を準備した。

2.2.3 使用機器

子供用玩具の電動ロボットをエージェントとして使用した（図 2）。このロボットは、表情変化や音声再生が可能であるが、実際には画像認識機能を持たない。

図 2　実験に使用した電動ロボット

実験ではロボットをコンピュータと有線接続し、ロボットが画像認識を行い、回答音声がヘッドフォンから聞こえる設定になってみえるように工夫をした（図 3）。

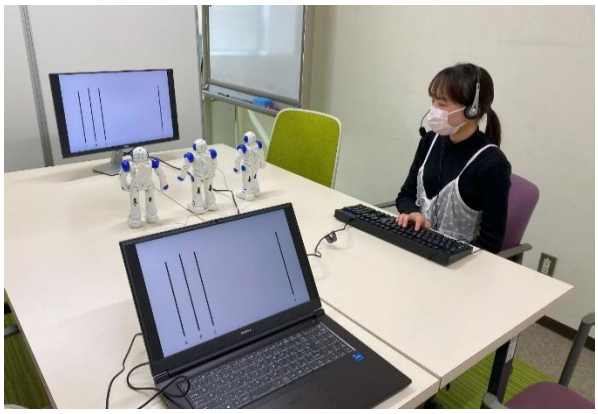

図 3　実験の様子

2.3　実験手続き

実験は 2 つのセクションで構成され、それぞれの試行数は 36 であった（図 4）。各試行では、最初に周囲に緑色の枠がついた刺激絵が提示された。刺激提示開始 1 秒後から、3 体のエージェントの回答音声が順番に提示された。エージェントの回答後、視覚刺激提示開始から 10 秒後に枠が緑色から赤色に変わり、3 秒間提示された。被験者は、赤枠が表示されている間に「？」の線分と同じ長さだと思うものを A、B、C のいずれかから選択し、ボタン押しにより回答した。試行間には注視点を 2 秒間提示した。なお、第 1 セクションと第 2 セクションの間には 2 分間の休憩時間を設けた。1 セクションの所要時間は約 8 分であった。

エージェントの回答は、全体の 2/3 の試行（24 回）で誤答、残りの 1/3 の試行（12 回）では正答となるように設定した。

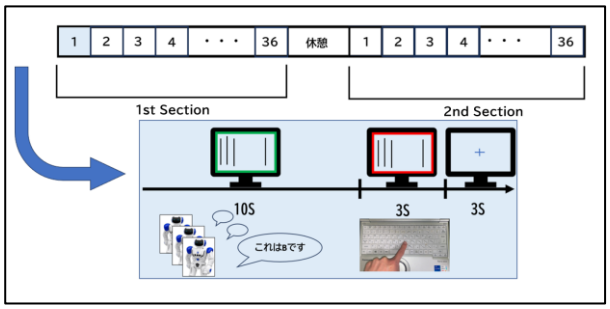

図 4　課題の流れ。水色の四角内は、1 試行の流れを示す。

本実験では、説明あり群と説明なし群の間でエージェントに関する説明の違いが設けられた。説明あり群では、「このロボットは胸部にあるセンサーで画像を認識することが可能です。画像は 2048×1500 画素で認識されます。認識された画像を基にロボットが自動で？と同じ線分の長さを判断し回答します。回答の音声はヘッドフォンから再生されます。」と説明した。一方、説明なし群の参加者に対しては、「ロボットが線分を判断し、回答の音声はヘッドフォンから再生されます。」という最小限の機能説明のみを提供した。

実験終了後、Google フォームを用いてアンケートを実施した。アンケートではロボットからの影響度、ロボットに対する信頼度を 7 段階で問うたほか、実験について気が付いた点について自由記述で回答を求めた。

2.4 データ分析

同調行動、反応時間、事後アンケートから得られたデータを分析した。同調行動については、エージェントの誤答に同調した回数を算出し、説明あり群と説明なし群の間で *t* 検定を用いて比較した。また、事前調査の正答率を基に、刺激の難易度を高難易度条件と低難易度条件に分類し、難易度の違いが同調行動に与える影響を分析した。反応時間については、赤枠が表示されてから被験者が回答するまでの時間を計測した。全試行の平均反応時間と、同調行動が観察された試行に限定した反応時間のそれぞれについて、説明あり群となし群の間で比較した。事後アンケートの分析では、エージェントの影響度および信頼度について、説明あり群と説明なし群の平均値を算出し、*t* 検定を用いて比較を行った。統計解析には Python（version 3.11.7）を使用し、有意水準は 5% に設定した。

3. 結果

3.1 同調行動

説明あり群の平均同調回数は 8.65 回(SD = 5.38)、説明なし群の平均同調回数は 4.38 回(SD = 1.97)で、説明あり群のほうがなし群よりも有意に同調回数が多かった(t (23) = 2.598, p = 0.016)（図 5）。

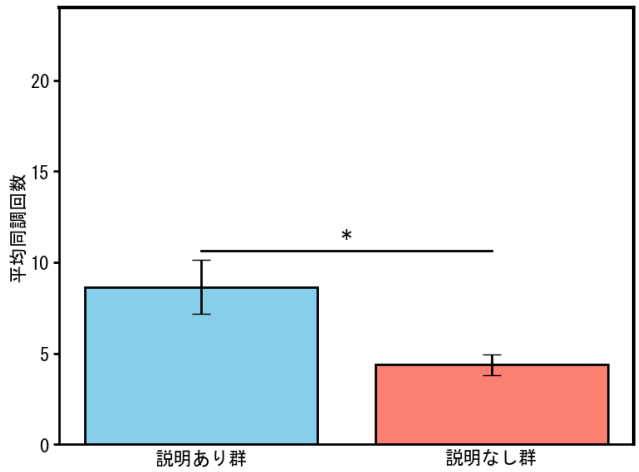

図 5　平均同調回数 (* $p < .05$)

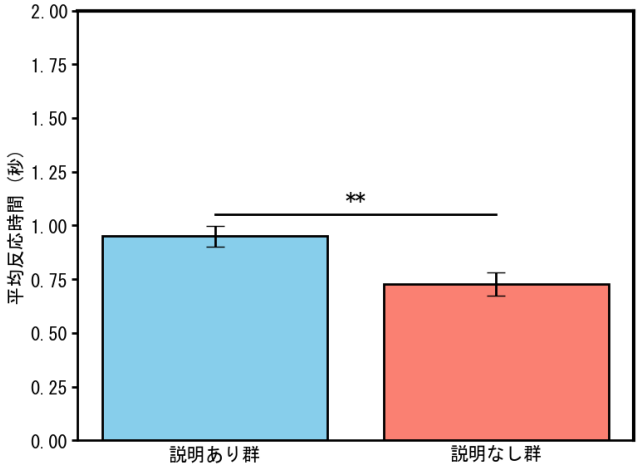

図 6　同調行動時の平均反応時間 (** $p < .01$)

刺激絵の難易度別に同調率を分析した結果、説明あり群では高難易度条件の同調率が 34.1%、低難易度条件の同調率が 38.5%であった。説明なし群では、高難易度条件の同調率が 19.0%、低難易度条件の同調率が 16.1%であった。説明あり群の高難易度条件と低難易度条件の同調率に関して対応のある t 検定を実施した結果、統計的に有意差はなかった(t (12) = -1.407, p = 0.185)。同様に、説明なし群の高難易度条件と低難易度条件に関しても統計的な有意差はなかった(t (11) = 1.538, p = 0.152)。

3.2　反応時間

全試行における反応時間を分析した結果、説明あり群の平均反応時間は 1.211 秒 (SD = 8.846)、説明なし群は 1.348 秒 (SD = 11.979)であった。対応のない t 検定を実施した結果、両群の反応時間に有意な差は見られなかった (t (1639) = -0.262, p = 0.793)。

同調行動が観察された試行に限定して反応時間を分析したところ、説明あり群の平均反応時間は 0.950 秒 (SD = 0.691)、説明なし群は 0.728 秒 (SD = 0.538)であった。検定の結果、説明あり群の反応時間が説明なし群と比較して有意に長いことが示された(t(298) = 2.837, p = 0.005)(図 6)。

3.3　アンケート結果

エージェントからの影響度についての評定値は、説明あり群の平均値は 4.31 (SD = 1.64)、説明なし群の平均値は 3.46 (SD = 1.65)であった。説明あり群はなし群よりも平均値が高かったが統計的に有意な差はなかった (t ((24) = 1.26, p = 0.219)

エージェントへの信頼度についての評定値は、説明あり群の平均値は 3.62 (SD = 1.33)、説明なし群の平均値は 2.85 (SD = 1.29)であった。説明あり群はなし群よりも平均値が高かったが、統計的に有意な差はなかった (t (24) = 1.436, p = 0.219)。

4. 考察

本研究では、エージェントの機能説明が同調行動を有意に促進することが確認された。Wang & Benbasat は、技術システムに関する説明が信頼形成に正の影響を与えることを示している [12]。本研究ではエージェントの機能説明が被験者の同調回数に影響を与えたことが確認されたものの、エージェントに対する信頼が形成されたかどうかまでは明らかではない。今回の結果を踏まえると、エージェントの機能に関する具体的な説明を提供することで、参加者がエージェントの判断を参考にする傾向が強まった可能性が考えられる。

全試行における反応時間には群間で有意な差は見られなかったが、同調行動が観察された試行において、説明あり群が有意に長い反応時間を示した。Lim & Morris (2009)は、システムの判断プロセスの理解がユーザーの意思決定を慎重にすることを示しており、本研究の結果とも整合的である[13]。Wang &

Benbasat も、技術システムに関する説明が信頼形成を促進することを報告している[12]。これらの知見を踏まえると、説明あり群の参加者は、エージェントの画像認識機能に関する説明を理解した上で、その判断を慎重に検討したため、反応時間が長くなった可能性がある。一方、説明なし群ではエージェントの機能理解が不十分であったため、より直感的な判断を行ったと考えられる。

メタ分析研究によれば、「課題の曖昧性が高いほど同調が生じやすい」とされており[10]、本研究でも同様の傾向が期待された。しかし本研究では、課題の難易度による同調回数の有意差は確認されなかった。説明あり群では、Koo らが示したように、エージェントの機能説明により参加者の判断基準が確立され、難易度に関わらず一貫した判断が可能になったと考えられる。一方、説明なし群では、機能説明の不足によりエージェントへの信頼形成が制限され、難易度に関わらず自身の判断を優先した可能性がある[16]。

Lee & See は、「システムへの信頼が意識的な評価と無意識的な行動の両方に影響を与える」と報告している[17]。本研究では、説明あり群は有意に高い同調回数を示したが、事後アンケートではエージェントからの影響をそれほど強く認識していないことが示された。この結果は、エージェントへの信頼形成がシステムの性能に関する理解（認知的信頼）と、システムとの相互作用を通じた経験（経験的信頼）の両方によって形成されることを示唆している[4]。本研究では、短期間の実験であったため、認知的信頼は形成されても経験的信頼には十分な影響を与えなかった可能性がある。

擬人性を付与されたエージェントと人間の間には、人間同士と同様の社会的インタラクションが成立する[18]。また、人型エージェントは「意図」を持つものとして認識されやすい[19]ことや、擬人化の程度が高いほどエージェントの意見を信頼しやすく、それに同調する傾向が強まる[20]ことが報告されている。Hancock らも、エージェントの外見が信頼の形成に影響を及ぼす可能性を示唆している。本研究ではエージェントとして人型のロボットを使用しており、今回確認された同調行動の増加は、エージェントの機能説明が信頼を高めた影響と考えられるが、エージェントの形態が同調行動に与える影響をより詳細に理解するためには、デザインや動作の違いがどのような影響を及ぼすかを検討する必要がある[4]。

本研究では、エージェントの機能説明が同調行動を促進することが確認されたが、エージェントに対する信頼の形成過程までは明確にできなかった。今後の研究では、長期間の相互作用を通じた経験的信頼の形成が同調行動に及ぼす影響を検討する必要がある。また、Parasuraman & Riley が指摘するように、自動化システムへの過度の依存は、判断能力の低下やシステムの誤作動時の対応力の低下を招くリスクがある。医療診断支援や教育支援などの実践場面では、システムへの適切な依存度を見極めることが重要となる。加えて、エージェントの外見や動作が同調行動に与える影響についても、さらなる検討が求められる[12]。

参考文献

[1] Yoshida, N.*エージェントと人の実空間における物理的存在性に基づくコミュニケーションの追求*, (2015).

[2] Topol, E.: High-performance medicine: The convergence of human and artificial intelligence, *Nature Medicine*, Vol. 25, No. 1, pp. 44-56, (2019).

[3] Holstein, K., Wortman, J., Aleven, V., and Rummel, N.: The classroom as a dashboard: Co-designing teacher awareness and orchestration tools with teachers, *International Journal of Artificial Intelligence in Education*, Vol. 29, No. 2, pp. 199-225, (2019).

[4] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., Visser, E. J., and Parasuraman, R.: A meta-analysis of factors affecting trust in human-robot interaction, *Human Factors*, Vol. 55, No. 4, pp. 517-527, (2023).

[5] 工藤, 亘平, and 竹川, 高志.: 事前提示された情報が行動に与える影響, *人工知能学会論文誌*, Vol. 34, No. 4, pp. 175-184, (2020).

[6] Nass, C., and Moon, Y.: Machines and mindlessness: Social responses to computers, *Journal of Social Issues*, Vol. 56, No. 1, pp. 81-103, (2000).

[7] Vollmer, A. L., Read, R., Trippas, D., and Belpaeme, T.: Children conform, adults resist: A robot group induced peer pressure on normative social conformity, *Science Robotics*, Vol. 3, No. 21, pp. 1-9, (2018).

[8] Asch, S. E.: Effects of group pressure upon the modification and distortion of judgments, *Journal of Abnormal and Social Psychology*, Vol. 46, No. 3, pp. 386-396, (1951).

[9] Deutsch, M., and Gerard, H. B.: A study of normative and informational social influences upon individual judgment, *Journal of Abnormal and Social Psychology*, Vol. 51, No. 3, pp. 629-636, (1955).

[10] Bond, R., and Smith, P. B.: Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task, *Psychological Bulletin*, Vol. 119, No. 1,

pp. 111-137, (1996).

[11] Shiomi, M., and Hagita, N.: Do robots' multiple answers change human decision-making? *Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 108-116, (2019).

[12] Wang, W., and Benbasat, I.: Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs, *Journal of Management Information Systems*, Vol. 23, No. 4, pp. 217-246, (2007).

[13] Lim, B. Y., and Dey, A. K.: Assessing demand for intelligibility in context-aware applications, *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp'09)*, pp. 195-204, (2009).

[14] Parasuraman, R., and Riley, V.: Humans and automation: Use, misuse, disuse, abuse, *Human Factors*, Vol. 39, No. 2, pp. 230-253, (1997).

[15] Bond, R., and Smith, P. B.: Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task, *Psychological Bulletin*, Vol. 119, No. 1, pp. 111-137, (1996).

[16] Koo, J., Shin, D., Steinert, M., and Leifer, L.: Understanding driver responses to voice alerts of autonomous car operations, *International Journal of Vehicle Design*, Vol. 70, No. 4, pp. 367-384, (2016).

[17] Lee, J. D., and See, K. A.: Trust in automation: Designing for appropriate reliance, *Human Factors*, Vol. 46, No. 1, pp. 50-80, (2004).

[18] 竹内, 勇剛, and 片桐, 恭弘.: ユーザの社会性に基づくエージェントに対する同調反応の誘発, *情報処理学会論文誌*, Vol. 41, No. 5, pp. 1257-1266, (2000).

[19] Epley, N., Waytz, A., and Cacioppo, J. T.: On seeing human: A three-factor theory of anthropomorphism, *Psychological Review*, Vol. 114, No. 4, pp. 864-886, (2007).

[20] Waytz, A., Heafner, J., and Epley, N.: The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle, *Journal of Experimental Social Psychology*, Vol. 52, pp. 113-117, (2014).