

ロボットによる課題進行は創造的認知を促進するか： 評価不安・信頼・ロボットへの印象に着目して

Do the robot experimenter facilitation human creative cognition? Focusing on fear of negative evaluation, trust, and impressions of the robot

安陪 梨沙[†], 服部 雅史[‡], 林 勇吾[‡]

Lisa Abe, Masasi Hattori

[†]立命館大学大学院人間科学研究科, [‡]立命館大学総合心理学部

[†]Graduate School of Human Science, Ritsumeikan University

[‡]College of Comprehensive Psychology, Ritsumeikan University

lt0839ie@ed.ritsumei.ac.jp

概要

本研究では、創造性課題において、ロボットまたは人による実験の進行が、作品の独創性に対してどのように影響するのか比較検討した。実験中に作品が評価されることについて考えたか(状態評価不安)を調整変数とし、独創性に与える影響を検討した結果、ロボット群では状態評価不安が高いほど、独創性が低くなることが明らかになった。作品評価に対する疑問を発言しづらい状況が独創性の発揮に対して不利に働いたことが考えられる。

キーワード：創造性, ロボット, 評価不安

1. 目的

近年、文章制作における文例の提示やイラスト作成における構図の提案など、創造的活動において、AIが活用され始めている。また、音声認識や自然言語処理技術の発展により、ロボットやチャットボットのような、自動的な対話による応対システムの社会普及が進んでいる。特に、ロボットは表情や会話のトーン、身振り手振りを用いることで親しみやすさを演出することが可能であることから、今後、仕事仲間や友人のように振る舞い、人間の活動を補助していくことができるようになると考えられる。そのため、本研究では人の創造的な認知活動、つまり、アイデア生成や作品制作を支援するロボットの開発を目指している。

実験者の働きかけによって創造性が向上する例として、地球外の惑星に住む生き物を作成させる際に質・量を促進させるような付加的教示を行うことで独創性(各部位に新奇な表現が用いられているか)と作品の量が向上したという研究がある(吉田・服部, 2002)。また、この効果はロボットにおいても部分的に確認されており、ロボットが同様の教示を行うことで、人間の実

験者が何も教示しない時よりも作品の類似度(地球上に存在する生物とどの程度似ているか)の低下がみられている(林・下條, 2021)。このことから、人、あるいはロボットによって課題中に創造性を促進するような教示を与えることで作品の質が向上する可能性がある。しかし、林・下條(2021)では人とロボットのどのような要因が創造性を向上させているのか、具体的には検討されていない。そのため、本研究では、付加的教示がない場合に、ロボット実験者の存在が参加者の創造性にどのように影響するのか検討する。

また、人が創造性を発揮する際、他者から受ける影響の中でも、創造性を低下させると考えられるものに評価不安(他者からの否定的評価に対する不安; 笹川ら, 2004)がある。Diehl & Stroebe(1987)では評価不安を誘発するとアイデアの算出量が減少したことから、評価不安はアイデアの生成量、質を抑え、無難なアイデアを生み出すことにつながると考えられる。しかし、評価不安は他者から受ける評価に対する恐れを主としている。このことから、実験の進行において人が存在しない、ロボットが実験進行を行う状況では評価不安が軽減される可能性がある。また、独創性に対する評価不安の影響しやすさには、参加者の持つ、評価不安への感受性が関連すると考えられる。そこで、本研究ではこれを「特性評価不安」として、社会的不安測定尺度(笹川ら, 2004)を用いて測定し、実際に実験中に感じる評価不安(状態評価不安)と区別することとした。特性評価不安が高い参加者ほど、人間群において実験中不安が誘発され、作品の独創性が低くなることが考えられる。

以上のことから、本研究では創造的な課題を行う実験環境において、ロボットのどのような要因が創造性に影響を与えるのか、ロボットをどのように用いることで人の創造性を支援できるか検討する。そのため、参

加者をロボットによる実験進行群(ロボット群)と人間による実験進行群(統制群)に分け、創造性課題における作品の独創性に差が生じるか、また評価不安の高低によって人とロボットの課題進行により作品の成績に差が生じるかを検討する。

2. 方法

2.1 参加者

立命館大学生 53 名(男性 15 名, 女性 38 名, M: 19.3, SD: 0.93) が作品の作者として実験に参加した。また, CrowdWorks で募集された 20 名(男性 16 名, 女性 4 名, M: 39.6, SD: 7.01) が作品の評価者として実験に参加した。

2.2 装置

実験室の机に実験実施用タブレット (Surface Go 3) を設置し, キーボード (Surface Go Type Cover) と Surface ペンを接続した。実験の進行状態を把握するため天井に設置されたモニターカメラを用いた。カメラの操作は別室に設置されたコンピュータによって行った。実験用ロボットは Palmi (DMM. Make ROBOTS; 図 1) を使用した。Palmi の操作には Palmi What' s up? というアプリケーションを使用し, 実験者が使用するノートパソコン(Lenovo, ThinkPad E450) で作動させた。実験の課題を行う際には JavaScript により作成された描画用システムを使用した。操作パネルには線の色(黒・白)や, 太さと透明度の変更, 描画を一段階戻す, 全て消去する機能があった。

2.3 尺度

特性評価不安の尺度として 5 段階のリッカート尺度である短縮版社会不安測定尺度(笹川ら, 2004)を使用した。実験中に感じる評価不安(「状態評価不安」)は, 自分の作品が評価されることを想起する程度と関係すると考え, 「地球外生命体の作成中に自分の作品がどのように評価されるかについて考えましたか(1: 全く考えていない~5: 何度も考えた)」という質問項目を用意した。信頼のしやすさの測定に, 5 段階のリッカート尺度である一般的信頼尺度(Yamagishi & Yamagishi, 1994)を使用した。実験中の心身の状態の測定に, 6 尺度に対して 1~100 点で点数付けを行う日本語版 NASA-TLX (芳賀・水上, 1996)を使用した。ロボットへの印象を評定するため, ロボット群のみ Godspeed Questionnaire (Bartneck et al., 2009) の下位尺度である知的印象 (Perceived

図 1 Palmi (DMM. Make ROBOTS)



Intelligence)項目を SD 法で実施した。この項目は事前に行った予備調査で独創性と顕著に相関がみられたため使用した。

2.4 手続き

作者は各実験者要因(ロボット群・人間群)に無作為に割り当てられた。

作者が自発的に入室するよう, 実験室の扉に「時間前でも部屋に入っていて問題ありません。部屋に入ったら扉を閉めてください。(ロボット群のみ)中にいるロボットの指示に従ってください」と書かれた張り紙を掲示した。作者が中に入ると, 実験者またはロボットが, タブレットが設置された机の前に座るよう促した。タブレットに実験の概要, および個人情報の取り扱い, 実験の辞退が自由であることについて記述した同意書を表示した。記入後, 実験者またはロボットの進行により描画ツールに慣れるための練習試行を 5 分間行った。タブレットの画面に「描画ツールを使う練習をしてもらいます。自分の名前や○△□などを描いてみてください」と書かれた画像を提示した。続いて, 本番試行として, 出来るだけ多くの「地球外の惑星に住む新しい生き物」を生成する課題を 20 分間行った。その際, 5 分ごとに経過時間が提示された。終了後, Google フォームに Godspeed Questionnaire (Bartneck et al., 2009; ロボット群のみ), 短縮版社会不安測定尺度(笹川ら, 2004), 実験中の評価不安, 日本語版 NASA-TLX (芳賀・水上, 1996), 一般的信頼性尺度 (Yamagishi & Yamagishi, 1994) を翻訳したものが表示され, 回答した。

ロボットの発話タイミングの調整は実験者が別室から実験室の様子をモニターして行った。人間群では, 発話内容とタイミングがロボットと同様であるよう, 台本を作成し, 時間の管理と課題の説明を行った。

2.5 作品の選定と評価

評価対象となる作品数が 398 個となり, 評価者の負担が大きくなることが予想されたため, 実験終了から

10日後、作者に自身の作品のうち最も独創的だと思うものを1つ選ぶよう依頼した。

依頼に対して返信があった作品について、評価者に評価を依頼した。評価項目は「独創的：独創的であるか（1：全く陳腐である～7：非常に独創的である）」「精巧さ：丁寧に描いているか（1：非常に雑に描かれている～非常に丁寧に描かれている）」とした。

3. 結果

実験者がロボットを操作したと察知した作者、作品選定の依頼に対して返信がなかった作者を除いた36名（男性6名、女性30名）を分析対象とした。また、評価者については、平均回答時間が1作品につき10秒以下であった者、分散が大きかった($SD > 5000$)者、作品の提示から回答までの分散が大きかった($SD > 1000$)者、合わせて4名を除外し、16名の評価を分析対象とした。分析にあたっては、評価者ごとに得点を標準化した。

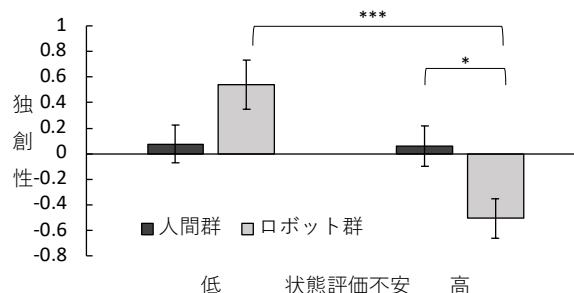
各評価項目での実験者要因間の差を t 検定によって分析した。その結果、参加者の一般的信頼のみ有意となった ($t = 2.48, df = 33.83, p = .02, d = 0.83$)。

独創性 ($t = 0.71, df = 28.10, p = 0.49, d = 0.24$)、精巧さ ($t = 0.72, df = 34.00, p = 0.48, d = 0.24$)、作品数 ($t = 0.18, df = 21.44, p = 0.86, d = 0.061$)、状態評価不安 ($t = 1.26, df = 28.63, p = 0.22, d = 0.42$)、NASA-TLX ($t = 0.64, df = 32.91, p = 0.53, d = 0.21$)、特性評価不安 ($t = 0.88, df = 34.00, p = 0.39, d = 0.29$) は有意でなかった。

実験者要因の傾向について検討するため、尺度間の相関を求めた (表1,2)。主な結果として、ロボット群のみ、独創性と状態評価不安の間に正の相関がみられた ($r = .70, p < .01$)。また、ロボット群でのみ、状態評価不安と特性評価不安の間に正の相関がみられた ($r = .56, p < .05$)。

続いて、状態評価不安の高低を中央値で分類した。実験者要因（人間群・ロボット群）と状態評価不安（評定値低：3以下・評定値高：4以上）が独創性得点に与える影響を参加者間分散分析により検討した (図2)。その結果、実験者要因間の主効果は有意でなく $F(1, 32) = 0.09, p = .76, \eta^2 = .003$ 、状態評価不安の主効果 $F(1, 32) = 10.73, p = .002, \eta^2 = .25$ 、と交互作用 $F(1, 32) = 10.19, p = .003, \eta^2 = .24$ 、が有意であった。交互作用について単純主効果検定を行

図2 事前課題ごとの平均創造性評定値



注) エラーバーは標準誤差を示す。*: $p < .05$, ***: $p < .001$

った結果、状態評価不安が高い時、実験者要因間に差が認められた $F(1, 32) = 6.58, p = .02, \eta^2 = .17$ 。また、ロボット群において、状態評価不安の高低間に差が認められた $F(1, 32) = 4.51, p < .001, \eta^2 = .38$ 。

4. 考察

本研究では、ロボット、人間という実験者の違いが創造性課題で作成された作品の独創性に与える影響を検討した。その結果、独創性得点には実験者要因間で差がみられなかったにもかかわらず、ロボット群のみ状態評価不安と独創性得点の間に正の相関が見られた。また、独創性得点に対する実験者要因と状態評価不安の高低の影響を分析したところ、状態評価不安の程度が高いほど、ロボット群において独創性得点が低くなることがわかった。このことから、実験中に、評価を受けることを考える参加者にとって、実験者がロボットであることは不利に働く可能性がある。

ロボット群で状態評価不安の高い参加者の独創性が低下した要因として、参加者の持つ、ロボットとのインタラクションへの抵抗が考えられる。実験中、人間群では作業内容の確認や質問紙の内容等に関する質問が複数あったにも関わらず、ロボット群ではそのような質問がほとんどみられなかった。状態評価不安の高い参加者にとっては、作品評価や実験内容について疑問を抱いても口に出しづらい状況が独創性の発揮に対して不利にはたらいた可能性がある。さらに、ロボット群でのみ状態評価不安と特性評価不安の間に正の相関がみられたことから、ロボット群において特性評価不安が高い参加者、つまり元々評価不安を感じやすい参加者は、上記の状況によって状態評価不安が誘発されやすくなっていた可能性がある。

また、実験後に実施した一般的信頼尺度の得点について、実験者要因間で差が見られていた。一般的信頼は相手についての情報がない場合の、相手の信頼性に対する推定値である（小杉, 1998）。群を分けた時点で一般的信頼に差があった場合、人間群では初対面である実験者をより信頼し、状態評価不安の高低によらず、創造性課題に取り組むことができた可能性がある。一方、ロボット群では、初対面であるロボットを信頼しにくく、状態評価不安の強さによって独創性に差が出たのかもしれない。また、一般的信頼尺度への回答を、実験手続きの後半に実施したことから、一般的信頼尺度得点の差には実験手続きが影響したことも考えられる。特に、ロボット群では、インタラクションに気を使うこと、会話が円滑に進まないことなどから一般的信頼が損なわれた可能性も考えられる。

以上から、ロボットが創造性課題を進行する際に特有の傾向として、状態評価不安が強く影響することがわかった。

文献

[1] Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81. <http://dx.doi.org/10.1007/s12369-008-0001-3>

[2] Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3), 497–509.

<https://doi.org/10.1037/0022-3514.53.3.497>

[3] 芳賀 繁・水上 直樹 (1996). 日本語版 NASA-TLX によるメンタルワークロード測定：各種室内実験課題の困難度に対するワークロード得点の感度 *人間工学*, 32(2), 71–79. <https://doi.org/10.5100/jje.32.71>

[4] 林 勇吾・下条 志敏 (2021). ロボットを用いた創造的認知の支援に向けたファシリテーション方法に関する実験的検討 *知能と情報*, 33(4), 768–776. https://doi.org/10.3156/jsoft.33.4_768

[5] 小杉 素子・山岸 俊男 (1998). 一般的信頼と信頼性判断 *心理学研究*, 69(5), 349–357. <https://doi.org/10.4992/jjpsy.69.349>

[6] 笹川 智子・金井 嘉宏・村中 泰子・鈴木 伸一・嶋田 洋徳・坂野 雄二 (2004). 他者からの否定的評価に対する社会的不安測定尺度(FNE)短縮版作成の試み：項目反応理論による検討 *行動療法研究*, 30(2), 87–98. https://doi.org/10.24468/jjbt.30.2_87

[7] Yamagishi, T., & Yamagishi, M. (1994). Trust and commitment in the United States and Japan. *Motivation and Emotion*, 18(2), 129–166. <https://doi.org/10.1007/BF02249397>

[8] 吉田 靖・服部 雅史(2002). 創造的問題解決におけるメタ認知的処理の影響 *認知科学*, 9(1), 89–102. <https://doi.org/10.11225/jcss.9.89>

表 1 人間群における測定項目間の相関

	独創性	精巧さ	特性評価不安	状態評価不安	一般的信頼	NASA-TLX	作品数
独創性							
精巧さ	0.44						
特性評価不安	-0.12	-0.04					
状態評価不安	-0.02	↙					
一般的信頼	-0.04	↘					

表 2 ロボット群における測定項目間の相関

	独創性	精巧さ	特性評価不安	状態評価不安	一般的信頼	NASA-TLX	作品数	Perceived Intelligence
独創性								
精巧さ	0.55 *							
特性評価不安	0.51 *	0.38						
状態評価不安	0.70 **	0.32	0.56 *					
一般的信頼	-0.26	-0.29	-0.22	-0.24				
NASA-TLX	0.14	0.30	0.23	0.28	-0.18			
作品数	0.33	0.33	-0.05	0.23	-0.01	-0.28		
Perceived Intelligence	0.37	0.14	0.00	0.24	0.14	0.14	0.00	

*: $p < .05$ **: $p < .01$