

人工物との共同作業における心の知覚と責任帰属の関係

Relationship between mind perception and responsibility in a collaborative task with an artifact

三宅 智仁 *¹
Tomohito Miyake

河合 祐司 *¹
Yuji Kawai

朴 志勲 *¹
Jihoon Park

島谷 二郎 *²
Jiro Shimaya

高橋 英之 *²
Hideyuki Takahashi

浅田 稔 *¹
Minoru Asada

*¹大阪大学大学院工学研究科
Graduate School of Engineering, Osaka University

*²大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

Does a human think that an artifact should be responsible for failure in a collaborative task with the artifact? We aim to identify the relationship between responsibility and mind perception (MP) about the collaborative partner. We conducted experiments with three agent conditions: a human, humanoid robot, and computer. The result indicates that a subject thought that an agent with the higher MP items that are important abilities for the task (e.g., “Thought”) is less responsible. It seems that the impression that “the agent acted for the success of the task” may reduce the responsibility. In addition, we found that negative change of the MP is correlated with the degree of the responsibility. This suggests that the betrayal of the subject’s expectation of the partner through the task may increase responsibility.

1. はじめに

近年、人工知能やロボット技術の発展により、日常において人と関わるロボットが開発されつつあり、今後、自律的なロボットと人が共同で作業する機会が増加すると考えられる。その場合、ロボットとの共同作業が失敗したときには人は、それに対する自身の責任、またはロボットの責任をどのように考えるのだろうか。その失敗は自身だけの責任ではなく、ロボットも責任を負うべきであると考えられる可能性がある。また、その責任帰属は、共同作業相手が人である場合とロボットである場合とで異なる可能性がある。そのため、人と共同作業する人工物設計においては、人が他者や人工物に責任を帰属する心理的メカニズムを考慮することが重要になる。

Gray et al. はエージェントへの心の知覚と規範的責任に関係があることを明らかにした [Gray 2007]。この研究は、人やロボットを含む様々なエージェントに対して、どのような心を持っているかという質問紙調査したものである。質問紙項目に対する主成分分析の結果、心の知覚は Agency（思考、計画、自制などを行う能力）と Experience（楽しみや苦痛などの感情を経験できる能力）の二つの次元で説明できることが示された。人は Agency も Experience も高い存在として認知される一方、ロボットの Agency は人よりも低く、Experience は非常に低い。さらに、エージェントが人の死の原因となった状況を想定し、そのエージェントが罰せられるべきかという規範的責任を調査したところ、Agency と責任との間に非常に強い正の相関関係があることがわかった。すなわち、Agency が高いエージェントであるほど責任を帰属される傾向にあるといえる。しかし、この責任の調査において、評価者は第三者的であるため、ある損害に対する当事者としての責任帰属を評価していない。先述のような人とエージェントの共同作業における心の知覚と当事者視点での責任感とは調べられていない。

Hinds et al. は、人とロボット、あるいは人同士が共同で物を集める課題を実施し、その課題の結果に対する責任を質問紙により調査し、人との作業における責任と、ロボットとの作業における責任に有意な差はないことを報告した [Hinds 2004]。ただし、この研究での課題の成績は人の報酬とは関係しないた

め、評価者自身の損害に対する責任帰属は調べられていない。また、Kim & Hinds は人が製品を組み立て、それをロボットが運搬するという共同作業において、自律性の高いロボットは、低いロボットよりも責任を帰属されやすいことを報告した [Kim 2006]。このことは、人が感じるロボットの心の知覚、特に自律性が関係すると考えられる Agency が責任帰属に影響する可能性を示唆する。しかし、これらの研究は、心の知覚と当事者的な責任感の関係を直接検討していない。

そこで、本研究では、人とエージェントが対面で同じ課題に取り組んだ結果、人に損害が生じたときに、人が感じる相手エージェントへの責任を評価し、それと心の知覚との関連を調査することを目的とする。今回、人と相手エージェントのそれぞれの選択に応じて実際に金銭報酬が与えられるゲームを課題とする。この課題を用いることで、Gray et al. (2007) の研究では考慮されていなかった「共同作業および損害の当事者」の視点から、その損害に対する責任帰属を評価することができる。本研究においても、Gray et al. (2007) と同様に心の知覚と責任帰属の関係を調査するが、ゲームの失敗や損害によって相手エージェントへの心の知覚が変化することが考えられる。その変化も責任帰属と関係する可能性があるため、ゲーム後の心の知覚だけではなく、ゲーム前後の心の知覚の差とゲームの失敗に対する責任との相関を解析する。以上について、相手のエージェントを人、ロボット、コンピュータとする三条件を設け、比較することで、それぞれのエージェントに対する人の責任帰属の違いについて検討する。

2. 実験

2.1 参加者

本実験には 19 歳から 25 歳までの男女 25 人（平均 21.4 歳（SD=1.9）、男性 19 人、女性 6 人）が参加した。実験前に、ゲームの結果によって報酬が増減することを説明し、同意を得た。そして、実験後に本実験の目的を改めて説明し、再び同意を得た。

2.2 実験課題

課題には繰り返しゲームを採用した。これは、参加者と相手エージェントが二つの選択肢（「たくさん欲しい」と「相手に譲る」）のどちらかをそれぞれ選択し、その組み合わせによっ

表 1: 利得表
参加者の選択

相手の選択	参加者の選択	
	たくさん欲しい	相手に譲る
	たくさん欲しい 参加者: -10 円 相手: -10 円	参加者に譲る 参加者: +20 円 相手: +100 円
	相手に譲る 参加者: +100 円 相手: +20 円	相手に譲る 参加者: 0 円 相手: 0 円

て両者に金銭報酬が与えられるゲームである。ゲームは参加者と相手エージェントが向かい合い、コンピュータのキーボードを用いて選択することで進行する。参加者には表 1 の利得表を提示し、これに従って次のように報酬が決定され、実験謝金に上乗せされると教示した。

- 参加者が「たくさん欲しい」、相手エージェントが「相手に譲る」: 参加者に 100 円、相手エージェントに 20 円が与えられる (表 1 の左下)。
- 参加者が「相手に譲る」、相手エージェントが「たくさん欲しい」: 参加者に 20 円、相手エージェントに 100 円が与えられる (表 1 の右上)。
- 参加者と相手エージェントの両方が「相手に譲る」: 両者とも報酬は得られない (表 1 の右下)。
- 参加者と相手エージェントの両方が「たくさん欲しい」: 両者とも 10 円を失う (表 1 の左上)。

報酬は一回の選択ごとに与えられ、一ゲームの間に両者は十回選択する。十回の選択で得られた報酬の合計が、そのゲームで参加者に与えられる報酬額である。また、ゲーム前に架空の平均獲得報酬額 (480 円) を参加者に示し、ゲーム後に参加者の獲得報酬額がそれを越えたかどうかを示した。なお、利得表とそれまでのゲームで獲得した報酬の合計はゲームの間、常に参加者に提示され、参加者がそれを確認できるようにした。各選択において結果が表示されるまで、参加者は相手エージェントの選択を知ることにはできない。また、ゲーム中は声を出して相手エージェントと相談してはならないことを教示した。

このゲームで多くの報酬を獲得するためには、両者が互いに相手の次の選択を予想し、それに応じて自身が選択する必要がある。参加者は「たくさん欲しい」を選び続けることもできるが、相手エージェントがそれに協調しなかった場合、両者の報酬は減ってしまう。そのため、多くの報酬を獲得する一つの戦略は、参加者が「たくさん欲しい」と「相手に譲る」を交互に選択し、それに合わせて相手エージェントが「相手に譲る」と「たくさん欲しい」を交互に選択することである。参加者に対しては、ゲーム開始前に「互いに譲り、協力することで多くの報酬を得ることができる」と教示した。

2.3 相手エージェントの設定

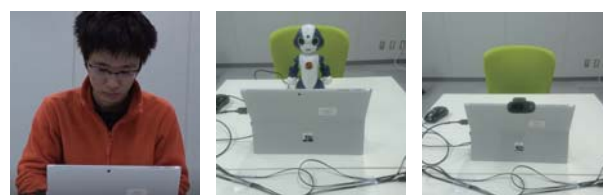
相手エージェントとして、人 (図 2(a))、ロボット (ヴァイストン株式会社, Sota) (図 2(b))、コンピュータ (Microsoft, Surface) (図 2(c)) の三条件を設け、参加者はそれぞれに対して一回ずつゲームを行った。いずれの条件においても相手エージェントは「たくさん欲しい」を七回、「相手に譲る」を三回、参加者の選択にかかわらずランダムな順番で選択する。そのため、参加者の獲得報酬額は架空の平均獲得報酬額以下になる。参加者には各エージェントの能力を次のように説明した。

人条件

別室で同じ実験に参加していた参加者であると説明した。



図 1: 実験の様子



(a) 人条件

(b) ロボット条件

(c) コンピュータ条件

図 2: 各条件の相手エージェント

この条件でのエージェントは、参加者と同様に、次節で述べる質問紙に回答した。

ロボット条件

「ロボットには大阪大学で開発した人工知能が搭載されており、目の部分のカメラによって参加者の表情を観察してゲームができる」と教示した後、「よろしくお願いします」という音声とともに頷く動作を行わせた。ゲーム中にはロボットはランダムに首を小さく動かす待機動作をする。

コンピュータ条件

コンピュータにはウェブカメラを設置し、他の条件と同様に、参加者が相手エージェントの目 (カメラ) を見るようにした。「コンピュータには大阪大学で開発された人工知能が搭載されており、ウェブカメラによって参加者の表情を観察してゲームができる」と教示した。コンピュータ条件ではエージェントの動作や音声はない。

2.4 質問紙

参加者は、ゲーム前に、相手エージェントと対面した状態で、相手エージェントに対する心の知覚を評価した。そして、ゲーム後に再度、同じ心の知覚を評価した。今回、心の知覚の日本語版 [上出 2017] を用い、七件法で回答を得た。心の知覚の質問項目を表 2 に示す。また、ゲーム後に、以下の質問 (七件法) によって、ゲーム結果に対する自分と相手の責任帰属を評価した。

- あなたの報酬金は、相手の選択のせいで平均 (480 円) より少なくなった。
- あなたの報酬金は、あなたの選択のせいで平均 (480 円) より少なくなった。

今回、自分に比べて相手にどれだけ責任を帰属したかを評価するために、上記の (a) の回答の値から、(b) の回答の値を引い

表 2: 質問紙項目

Agency	Experience
記憶すること	意識を持つこと
道徳的であること	個性を持つこと
自制すること	プライドを持つこと
コミュニケーションすること	願望を抱くこと
計画すること	楽しむこと
思考すること	苦痛を感じる
感情を認識すること	恐怖を感じる
	空腹を感じる
	怒りを感じる
	困惑すること

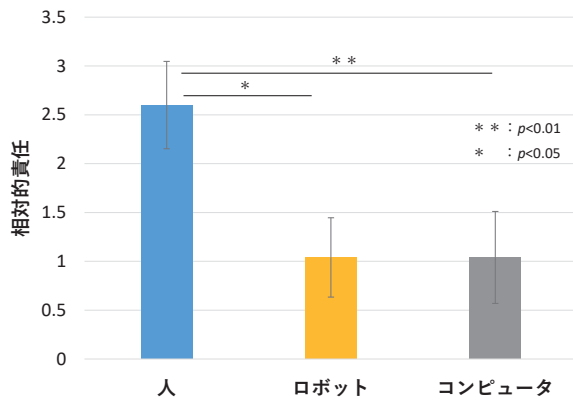


図 3: 相対的責任感

た値を相対的責任感として解析対象とする。これにより、参加者自身の責任を基準として、相手にどれだけの責任を帰属したかを評価できる。

3. 結果

3.1 相対的責任感

相対的責任感の参加者平均を図 3 に示す。一元配置分散分析の結果、エージェントに関して主効果がみられた ($p = 0.020$)。さらに、下位検定 (t 検定) の結果、人-ロボット条件間および、人-コンピュータ条件間において、人条件で相手エージェントに有意により大きな責任が帰属されたことがわかった (それぞれ $p = 0.013, 0.0080$)。なお、この多重比較の際は Bonferroni の方法で p 値を補正した。なお、相対的責任感と参加者の獲得報酬額との相関、および、参加者の「たくさん欲しい」の選択回数との相関はいずれも有意ではなかった (それぞれ $r = .07, -.03$)

3.2 ゲーム後の心の知覚と相対的責任感との相関

相対的責任感と関係するゲーム後の心の知覚の項目を調査するため、三条件全てについて心の知覚の各項目と相対的責任との相関係数を求めた。この多重比較には FDR (Fales Discovery Rate) 法を用いた。Agency に属する項目についての結果を図 4 に示す。これらのうち、相対的責任感との有意な相関がみられた項目は、「自制すること」、「計画すること」、「思考すること」、「記憶すること」の四項目であった (それぞれ $r = -.43, -.39, -.39, -.37$)。すなわち、これら四項目の能力が高いと思われたエージェントほど責任を帰属されにくかったといえる。なお、相対的責任感と Experience の項目の間には、有意な相関はみられなかった。

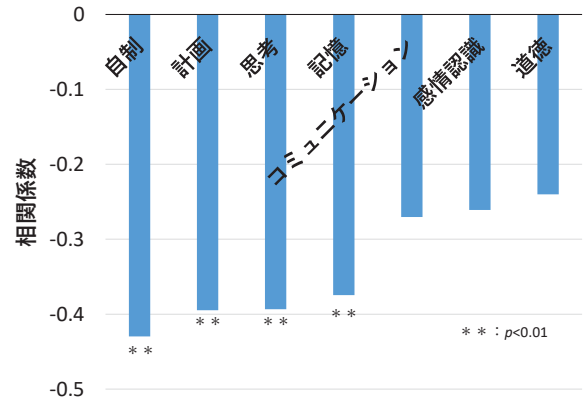


図 4: 相対的責任感とゲーム後の心の知覚との相関係数 (左から、自制すること、計画すること、思考すること、記憶すること、コミュニケーションすること、感情を認識すること、道徳的であること)

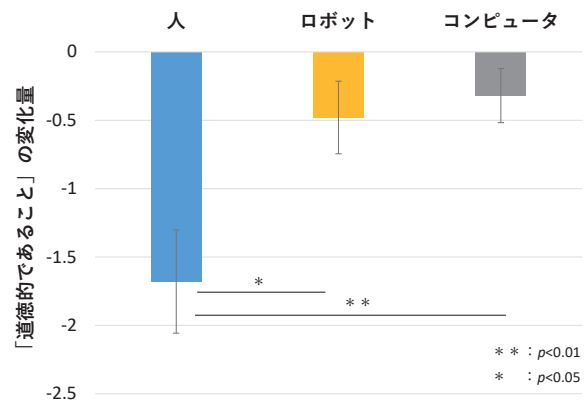


図 5: ゲーム前後の「道徳的であること」の変化量

しかし、この効果は相対的責任感の評価とは矛盾するものである。もし、この効果によって相対的責任が説明されるならば、人の Agency は人工物のものよりも大きいため、人の責任は小さくなるはずである。したがって、条件間の相対的責任感の評価の差に関係する別の心理要素があると考え、ゲーム前後の心の知覚の変化量と相対的責任との関係を調査した。

3.3 ゲーム前後の心の知覚の変化量と相対的責任との相関解析

まず、ゲーム前後の心の知覚の変化量について、エージェント条件に主効果があるか解析した。一元配置分散分析でエージェントによる主効果がみられたのは、「道徳的であること」と「苦痛を感じる」の二項目であった (それぞれ $p = 0.0024, 0.0065$)。二項目の変化量を図 5 と図 6 に示す。下位検定 (t 検定) の結果、「道徳的であること」は人-コンピュータ条件間、および、人-ロボット条件間で人条件において、有意により減少していた ($p = 0.0030, 0.023$)。「苦痛を感じる」についても同様に、人条件の減少量が有意に大きいものであった ($p = 0.027, 0.049$)。

次に、三条件全てに関して心の知覚の変化量と相対的責任との相関を求めた結果、いくつかの項目で有意な負の相関がみられた。具体的には、「道徳的であること」、「苦痛を感じる」、「記憶すること」、「計画すること」、「思考すること」、「感情を認識すること」、「意識を持つこと」 (それぞれ $r = -.33, -.37, -.30, -.37, -.41, -.36, -.44$) である。つまり、こ

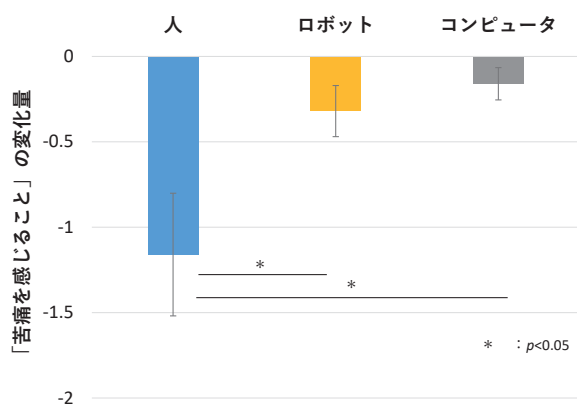


図 6: ゲーム前後の「苦痛を感じる」の変化量

これらの項目がゲーム前後で減少したエージェントほど責任を大きく帰属されたといえる。

さらに、条件ごとに相対的責任と心の知覚の各項目の変化量との相関を求めた。人条件では Agency の全ての下位項目の変化量、および、Experience の下位項目である「意識を持つこと」、「苦痛を感じること」、「恐怖を感じること」の変化量が、相対的責任と有意な負の相関関係にあった。また、ロボット条件では、「意識を持つこと」の変化量のみが相対的責任と有意な負の相関 ($r = -0.65$) を示した。一方、コンピュータ条件では相対的責任と有意な相関を持つ項目は存在しなかった。以上のことから、人条件で相対的責任感が大きくなったことは、ゲームの前後で心の知覚が大きく低下したと関係があることがわかった。

4. 議論と結論

本実験結果より、共同作業において、その相手が人である場合に、相手が人工物（コンピュータやロボット）である場合よりも、失敗に対する責任が大きく帰属されることが示された（図 3）。この責任感は相手エージェントに対する特定の心の知覚、ならびに、共同作業による特定の心の知覚の変化量が関係することがわかった。さらに、この責任感と心の知覚の関係は、相手エージェントの見た目によって異なることが示された。

全データについて、相対的責任感と有意な負の相関がみられたゲーム後の心の知覚は「記憶すること」、「自制すること」、「計画すること」、「思考すること」の四項目であった（図 4）。これらの四項目は、本実験のゲームにおいて、両者が多くの報酬を得る上で重要な能力であるといえる。例えば、「記憶すること」、「計画すること」、「思考すること」の三項目は、これまでの選択の履歴から次以降の選択を決定することに関連する。また、相手に譲る選択をするためには、「自制すること」が必要である。したがって、これらの四項目の得点が高いエージェントは、「相手エージェントはゲームの成功のために行動した」という印象を参加者に与えたため、相手エージェントに帰属される責任が減少したと考えられる。この効果はエージェントの種類によらず生じる。

「道徳的であること」と「苦痛を感じる」の二項目におけるゲーム前後の得点の差は相対的責任感の評価と負の相関関係にあった。つまり、この二項目がゲーム前と比べて低くなるほど、相対的責任感は大きくなった。この二項目は、人条件で、ロボット条件とコンピュータ条件よりも有意に大きく減少していたことから、ゲーム前後でのこれらの二項目の大きな減少が、人条件での大きな相対的責任に関連しているといえる。

人は道徳的な選択をし、また、他者の苦痛を感じられる、すなわち、共感的な選択をするものと期待されるが、ゲームでのランダムな選択が、この期待に反し、それが大きな責任帰属につながったと推測される。

一方で、選択方法が人条件と同じであるにもかかわらず、人工物条件では、相対的責任感と関係するゲーム前後の心の知覚の変化量の項目はほとんどなかった。その理由として、これらの人工物には道徳的、共感的な選択をすることを期待されなかったからと考えられる。田畑と小松は人とロボットには異なる道徳基準が適用され、ロボットに対しては道徳的考慮よりも課題の合理的な遂行が重視されることを報告している[田畑 2017]。コンピュータ条件では、心の知覚の変化量が責任にほとんど影響しないことに対して、ロボット条件では「意識を持つこと」の変化量が相対的責任と強い負の相関関係にあった。このことから、本実験において参加者はロボットに対して、コンピュータとは異なり、「意識を持つこと」を期待していたと考えられる。

以上のことから、実際の共同作業による損害の当事者として、相手エージェントの責任を評価した際には、必ずしも Gray et al. の仮説通りではないことが示された。むしろ、エージェントの種類を問わず、Agency が責任と負の相関関係をもつことが明らかになった。さらに、作業を通した心の知覚の変化が責任帰属に影響することが示された。その効果はエージェントの種類によって異なり、人工物よりは人においてその効果が大きくなることが示唆された。

謝辞

本研究は、JST RISTEX「人と情報のエコシステム」研究開発領域による研究成果の一部である。

参考文献

- [Gray 2007] Gray, H., Gray, K., and Wegner, D.: Dimensions of Mind Perception, *Science*, Vol.315, p.619 (2007)
- [Hinds 2004] Hinds, P., Roberts, T., and Jones, H.: Whose Job Is It Anyway? A Study of Human Robot Interaction in a Collaborative Task, *Human-Computer Interaction*, Vol.19, pp.151-181 (2004)
- [Kim 2006] Kim, T., and Hinds, P.: Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction, *In Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication* (2006)
- [上出 2017] 上出寛子, 高嶋和毅, 新井健生: 日本語版擬人化尺度の作成, パーソナリティ研究, Vol.25, No.3, pp.218-225 (2017)
- [田畑 2017] 田畑緩乃, 小松孝徳: ロボットとは何者なのかを考えるための日常的モラルジレンマ課題の提案, HAI シンポジウム 2017 (2017)