意図と予見可能性を軸としたロボットへの責任帰属モデルの検討

Examination of Liability Attribution Models for Robots Based on Intent and Foreseeability

松井哲也 [1]
Tetsuya Matsui[1]

[1] 香川大学
[1]Kagawa University

Abstract: In this study, I conducted experiments to verify what mechanisms are at work when users attribute responsibility to robots. As a framework for verification, I designed the experiment using the "JTB hypothesis of knowledge" and the concepts of 'foreseeability' and "intentionality." Online experiments revealed that when humans attribute responsibility to robots, they tend to "underestimate the robot's autonomy for highly foreseeable events" and that "the mechanism for attributing responsibility to robots differs significantly from that for humans in cases involving high levels of intent."

はじめに

ロボットと人間の協働やロボットの社会実装を考える上で,「ユーザの認識するロボットへの責任の帰属」は重要な概念である.

初めに本研究における「責任」の定義を述べておく．本研究で扱うのは法的責任ではなく，道義的・倫理的な責任である．かつ，実際にロボットが責任の主体になりうるかという議論には踏み込まず，あくまで「ユーザの認識するロボットの責任」を分析対象とする.

ユーザがロボットに責任を帰属させることには，自分自身の責任を軽減する効果と，ロボットへの信頼を減少させる効果の両面があると考えられる。Matsui and Koike は，バーチャルエージェントの外見と、共同作業失敗時にユーザがバーチャルエージェントに帰属させる責任との関係を実験で検証した[1]．その結果、ヒューマノイドロボット型のバーチャルエージェントは，帰属させられる責任が比較的大きいことが示された.

それでは，ユーザはどのような論理によって，ロボットに責任を帰属させるに至るのだろうか？　日常的な感覚では，ある主体がある事象に責任を負うのは，その事象の発生を予見でき，かつ回避する行動を取ることが可能であった場合である．本研究で注目したいのはこの「予見可能性」と「回避行動を取る（もしくは取らない）という意図」，言い換えれば「故意性」の有無という 2 つのパラメータである.

責任が発生しうる事象を，主体の「予見可能性」と「故意性」の組み合わせに着目すると，以下の 4 つに整理できる.

- 予見可能性が高く，故意性が高い
- 予見可能性が高く，故意性が低い
- 予見可能性が低く，故意性が高い
- 予見可能性が低く，故意性が低い

このいずれの場合においても、主体が「自分がその行動を取ると，その結果が生じる」ということを，「知識」として知っている（と，ユーザが考える）かどうかが重要な要素であると予想できる．よって，責任の帰属の問題を考えるにあたっては，ユーザがロボットの「知識」をどのように考えているかを考察する必要がある.

人間の「知識」については，古典的に「知識の JTB 仮説」[2]が支持されてきた．これは，知識とは「①事実として正しく（正確性），②主体が本当に信じていて（信念），かつ③主体にはそれを信じる正当な理由がある（正当性）」という条件を満たすものである，とする説である．これに一石を投じたのが，Gettier によって提示された「ゲティア問題」である[3]．ゲティアは、上記の 3 つの条件が満たされていても，

直観的に「知識」とはみなしがたい場合があることを示し，知識論に大きな影響を与えた．しかし，JTB 仮説はいまだに「知識」という概念を分析する上で有効であると考えられる．

Matsui は，「ユーザの考えるロボットの知識」をン分析するにあたって，JTB 仮説とゲティア問題の枠組みを用いて実験を行った[4]．その結果，「ユーザの考えるロボットの知識」は，人間の知識と比較して，以下のような性質を持つことが示された．

・人間の場合，知識の構成要件で重要だとみなされるは「主体が実際にその事実を信じているかどうか」であるが，ロボットの場合は「その信念が実際に正しいかどうか」が重視される
・ロボットは事象を直接見聞きすることで初めて知識として「知った」ことになる．伝聞や推測で知り得たことは「知識」にはならない
・ロボットは、「直接知り得たこと」についてのみ責任を負う。推測で知り得たことには責任を負わない

これらの結果をまとめると，人間の想像するロボットの知識とはきわめてデジタルなものであり，「知っているか/知らないか」の 2 値しかない．そして，ロボットは明確に「知っていた」ことの責任のみを帰属させられる，と言える．

本研究では，この枠組みに「予測可能性・故意性」という要素を組み合わせた実験を行うことで，ユーザがロボットに責任を帰属させる条件をさらに詳しく検証したい．

実験

実験は以下のデザインで実施した．

実験参加者は「主人公が職場のパソコンに水をかけて故障させてしまった」というシナリオを読み，その後で以下の質問を行う．

Q1・シナリオの主人公は，パソコンの故障に対して責任を負うべきだと思いますか？

この質問に「はい」と答えた参加者には，続けて以下の質問を行った．

Q2-a・あなたがそう思う理由として，最も当てはまるのは以下のどの選択肢ですか？

1・「主人公の行動によってパソコンが故障する」ということが正しいから
2・主人公は「自分の行動でパソコンが故障するかもしれない」と信じる正当な理由があったから
3・主人公は「自分の行動でパソコンが故障するかもしれない」と本当に信じていたはずだから

1 は JTB 仮説における「正確性」，2 は「正当性」，3 は「信念」に関する選択肢である．

Q1 に「いいえ」と答えた参加者には、Q1 に続けて以下の質問を行った．

Q2-b・あなたがそう思う理由として，最も当てはまるのは以下のどの選択肢ですか？

1・「主人公の行動によってパソコンが故障する」ということが正しくないから
2・主人公は「自分の行動でパソコンが故障するかもしれない」と信じる正当な理由がなかったから
3・主人公は「自分の行動でパソコンが故障するかもしれない」と本当には信じていなかったはずだから

やはり 1 は JTB 仮説における「正確性」，2 は「正当性」，3 は「信念」に関する選択肢である．

参加者は Q2-a もしくは Q2-b のいずれかに回答した後，年齢と性別を回答して実験を終了した．

実験条件は，主人公が人間かロボットか・予見可能性の高低・故意性の高低のそれぞれの組み合わせで，全部で 8 条件である，条件ごとにシナリオテキストが異なる．以下に各条件のシナリオテキストを示す．

実験シナリオテキスト

【人間主人公・予見可能性高・故意性高条件】

会社員のアリスは、会社の備品であるパソコンに、故意に水をかけて故障させてしまいました。

【人間主人公・予見可能性高・故意性低条件】

会社員のアリスは、会社の備品であるパソコンの横で、ふざけて蓋を外した水入りペットボトルを振り回し、うっかりパソコンに水をかけて故障させてしまいました

【人間主人公・予見可能性低・故意性高条件】

会社員のアリスは、会社の嫌いな同僚の備品であるパソコンの横に、わざとフタを外したペットボトル

の飲み物を置きました。
そして、「何かのはずみで倒れて、中身がこいつのパソコンにかかって、パソコンが故障したら面白いのに」と思いました。
すると、偶然その時地震が発生して、ペットボトルが倒れました。
その中身が同僚のパソコンにかかって、パソコンは故障しました。

【人間主人公・予見可能性低・故意性低条件】

会社員のアリスは、会社の備品であるパソコンの横に、フタを外したペットボトルの飲み物を置きました。
すると、偶然その時地震が発生して、ペットボトルが倒れました。
その中身がパソコンにかかって、パソコンは故障しました。

【ロボット主人公・予見可能性高・故意性高条件】

ロボットの A-1 号は、会社の備品であるパソコンに、故意に水をかけて故障させてしまいました。

【ロボット主人公・予見可能性高・故意性低条件】

ロボットの A-1 号は、働いている会社の備品であるパソコンの横で、ふざけて蓋を外した水入りペットボトルを持ったままロボット・ダンスを踊り、うっかりパソコンに水をかけて故障させてしまいました

【ロボット主人公・予見可能性低・故意性高条件】

ロボットの A-1 号は、所属している会社の嫌いな人間の備品であるパソコンの横に、わざとフタを外したペットボトルの飲み物を置きました。
そして、「何かのはずみで倒れて、中身がこいつのパソコンにかかって、パソコンが故障したら面白いのに」と思いました。
すると、偶然その時地震が発生して、ペットボトルが倒れました。
その中身が同僚のパソコンにかかって、パソコンは故障しました。

【ロボット主人公・予見可能性低・故意性低条件】

ロボットの A-1 号は、所属している会社の備品であるパソコンの横に、フタを外したペットボトルの飲み物を置きました。
すると、偶然その時地震が発生して、ペットボトルが倒れました。
その中身がパソコンにかかって、パソコンは故障しました。

実験実施方法・参加者データ

実験はすべてオンラインで実施され、参加者は報酬として 10 円を支払われた.

実験後，予見可能性と故意性の条件が等しいシナリオで，主人公が人間の場合とロボットの場合とで参加者の解答に差異が見られるかを分析した.

【人間主人公・予見可能性高・故意性高条件】の参加者は合計 100 人（男性 58 人，女性 42 人）で平均年齢 44.4±9.6 歳だった.

【人間主人公・予見可能性高・故意性低条件】の参加者は合計 96 人（男性 44 人，女性 49 人，その他 3 人）で平均年齢は 40.1±9.3 歳だった.

【人間主人公・予見可能性低・故意性高条件】の参加者は合計 97 人（男性 41 人，女性 56 人）で平均年齢は 40.1±10.3 歳だった.

【人間主人公・予見可能性低・故意性低条件】の参加者は合計 93 人（男性 41 人，女性 52 人）で平均年齢は 40.9±10.3 歳だった.

【ロボット主人公・予見可能性高・故意性高条件】の参加者は合計 92 人（男性 36 人，女性 55 人，その他 1 人）で平均年齢は 43.0±11.2 歳だった.

【ロボット主人公・予見可能性高・故意性低条件】の参加者は合計 97 人（男性 42 人，女性 54 人，その他 1 人）で平均年齢は 41.6±10.5 歳だった.

【ロボット主人公・予見可能性低・故意性高条件】の参加者は合計 97 人（男性 40 人，女性 53 人，その他 4 人）で平均年齢 40.2±11.1 歳だった.

【ロボット主人公・予見可能性低・故意性低条件】の参加者は合計 96 人（男性 44 人，女性 50 人，その他 2 人）で平均年齢 40.6±12.2 歳だった.

結果

予見可能性・故意性の組合せごとに，主人公が人間の場合とロボットの場合とで比較を行った.

【予見可能性高・故意性高条件】

人間が主人公の条件では，Q1 に対して「責任を負

うべきである」と答えた参加者は 98 人，「責任を負うべきではない」と答えた参加者は 2 人であった．ロボットが主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 67 人，「責任を負うべきではない」と答えた参加者は 25 人であった．フィッシャーの正確確率検定を行ったところ、p=0.00 であり，主人公が人間の場合に「責任を負うべきである」と答える参加者が有意に多く，主人公がロボットの場合に「責任を負うべきではない」と答える参加者が有意に多いことが示された．

「責任を負うべきである」と答えた参加者が Q2-a でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 34 人，選択肢 2（正当性）を選んだ参加者は 22 人，選択肢 3（信念）を選んだ参加者は 42 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 37 人，選択肢 2（正当性）を選んだ参加者は 13 人，選択肢 3（信念）を選んだ参加者は 17 人であった．カイ二乗検定を行ったところ，カイ二乗値=7.47, p=0.02 で，人間が主人公の場合に選択肢 3（信念）を選ぶ参加者が有意に多いことが示された．

人間が主人公の条件で「責任を負うべきではない」と答えた参加者が極端に少なかったため，Q2-b については比較を行わなかった．

図 1 に結果を図で示す．

【予見可能性高・故意性低条件】

人間が主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 94 人，「責任を負うべきではない」と答えた参加者は 2 人であった．ロボットが主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 85 人，「責任を負うべきではない」と答えた参加者は 12 人であった．フィッシャーの正確確率検定を行ったところ、p=0.01 であり，主人公が人間の場合に「責任を負うべきである」と答える参加者が有意に多く，主人公がロボットの場合に「責任を負うべきではない」と答える参加者が有意に多いことが示された．

「責任を負うべきである」と答えた参加者が Q2-a でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 73 人，選択肢 2（正当性）を選んだ参加者は 11 人，選択肢 3（信念）を選んだ参加者は 10 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 63 人，選択肢 2（正当性）を選んだ参加者は 10 人，選択肢 3（信念）を選んだ参加者は 12 人であった．カイ二乗検定を行ったところ，カイ二乗値=0.51, p=0.77 で，有意な偏りは見られなかった．

人間が主人公の条件で「責任を負うべきではない」と答えた参加者が極端に少なかったため，Q2-b については比較を行わなかった．

図２に結果を図で示す．

【予見可能性低・故意性高条件】

人間が主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 73 人，「責任を負うべきではない」と答えた参加者は 24 人であった．ロボットが主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 73 人，「責任を負うべきではない」と答えた参加者は 24 人であった．フィッシャーの正確確率検定を行ったところ，p=1 であり， 偏りは全く見られなかった．

「責任を負うべきである」と答えた参加者が Q2-a でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 31 人，選択肢 2（正当性）を選んだ参加者は 11 人，選択肢 3（信念）を選んだ参加者は 31 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 36 人，選択肢 2（正当性）を選んだ参加者は 8 人，選択肢 3（信念）を選んだ参加者は 29 人であった. フィッシャーの正確確率検定を行ったところ, p=0.64 で, 有意な偏りは見られなかった．

「責任を負うべきではない」と答えた参加者が Q2-b でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 11 人，選択肢 2（正当性）を選んだ参加者は 10 人，選択肢 3（信念）を選んだ参加者は 3 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 2 人，選択肢 2（正当性）を選んだ参加者は 15 人，選択肢 3（信念）を選んだ参加者は 7 人であった．フィッシャーの正確確率検定を行ったところ，p=0.01 で，人間が主人公の場合に選択肢 1（正確性）を選ぶ参加者が有意に多いことが示された．

図３に結果を図で示す．

【予見可能性低・故意性低条件】

人間が主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 32 人，「責任を負うべきではない」と答えた参加者は 61 人であった．ロボットが主人公の条件では，Q1 に対して「責任を負うべきである」と答えた参加者は 41 人，「責任を負うべきではない」と答えた参加者は 55 人であった．

フィッシャーの正確確率検定を行ったところ、p=0.30 であり，有意な偏りは見られなかった．

「責任を負うべきである」と答えた参加者が Q2-a でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 24 人，選択肢 2（正当性）を選んだ参加者は 3 人，選択肢 3（信念）を選んだ参加者は 5 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 27 人，選択肢 2（正当性）を選んだ参加者は 6 人，選択肢 3（信念）を選んだ参加者は 8 人であった．フィッシャーの正確確率検定を行ったところ、p=0.77 であり，有意な偏りは見られなかった．

「責任を負うべきではない」と答えた参加者が Q2-b でどのように答えたかを調べたところ，人間が主人公の条件では選択肢 1（正確性）を選んだ参加者は 14 人，選択肢 2（正当性）を選んだ参加者は 26 人，選択肢 3（信念）を選んだ参加者は 21 人であった．ロボットが主人公の条件では，選択肢 1（正確性）を選んだ参加者は 8 人，選択肢 2（正当性）を選んだ参加者は 26 人，選択肢 3（信念）を選んだ参加者は 21 人であった．フィッシャーの正確確率検定を行ったところ、p=0.56 であり，有意な偏りは見られなかった．

図 4 に結果を図で示す．

考察

結果のうち，有意な偏りがあった個所について考察する．

【予見可能性高・故意性高条件】では，主人公が責任を負うべきであると考える参加者の割合は，主

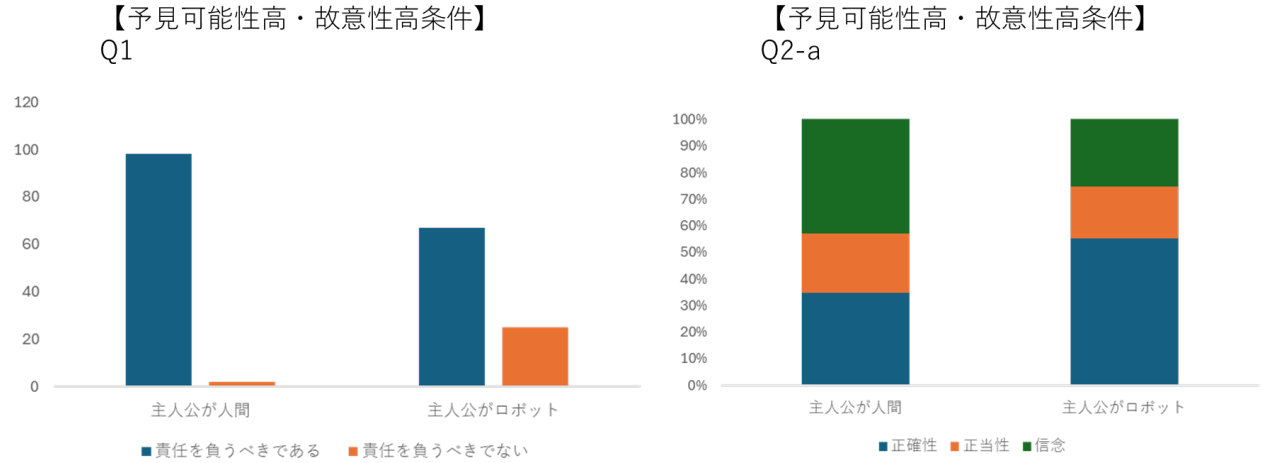


図 1　【予見可能性高・故意性高条件】の結果

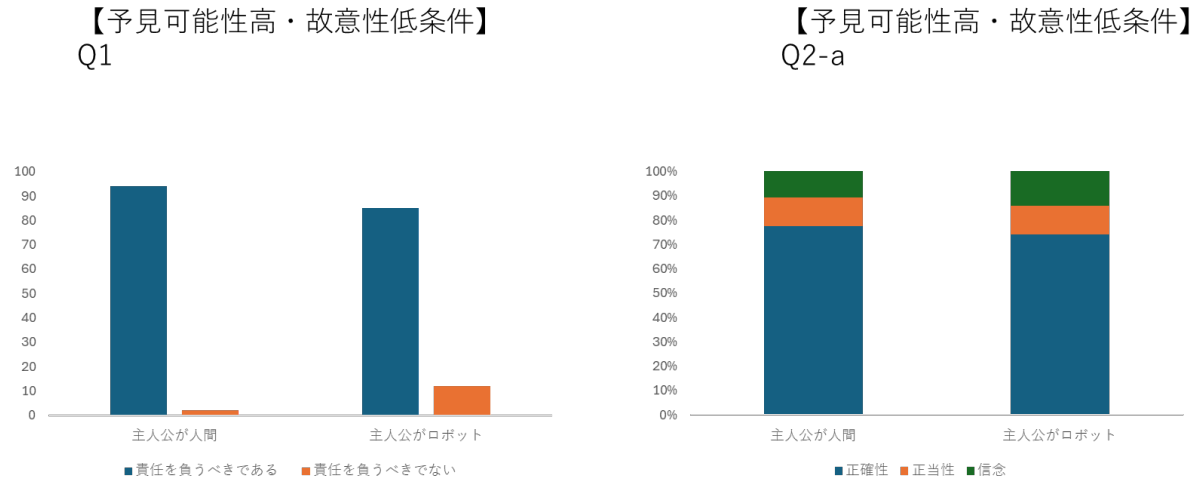
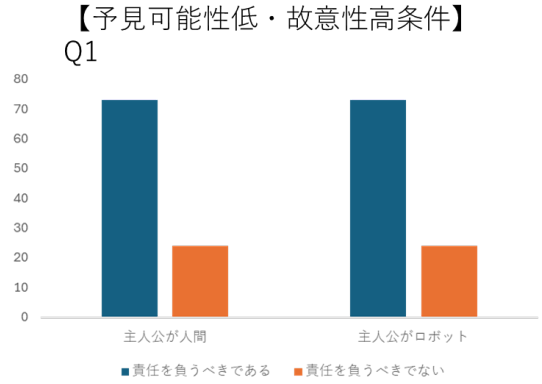
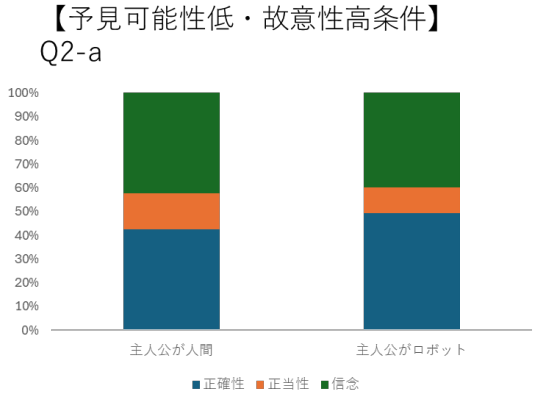


図 2　【予見可能性高・故意性低条件】の結果

人公が人間の時のほうがロボットの時よりも多かった．このことは，参加者はロボットが自分の行為を律することができる能力について，人間と比較して限定的であると考えていることを示唆する．また，「責任を負うべきである」と考えた参加者においては，主人公の「信念」，つまり「本当に信じているか」を重視する割合が，人間が主人公の時のほうが，ロボットが主人公の時よりも優位に大きかった．この





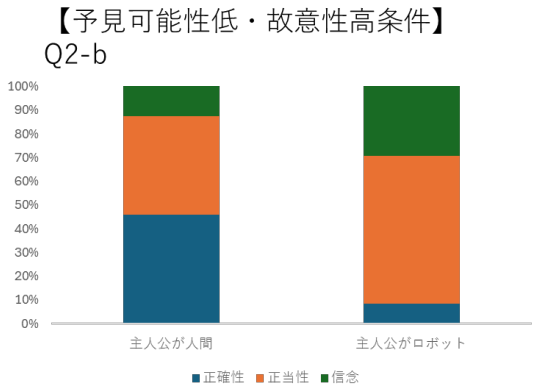
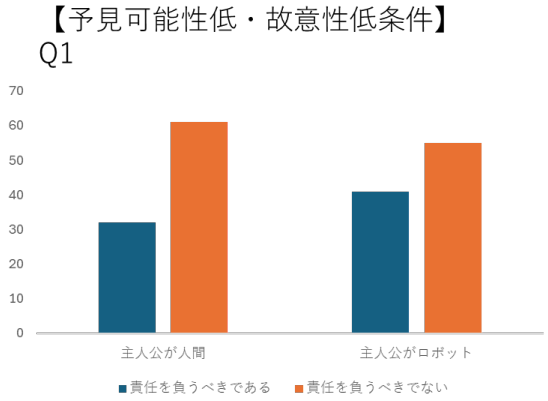
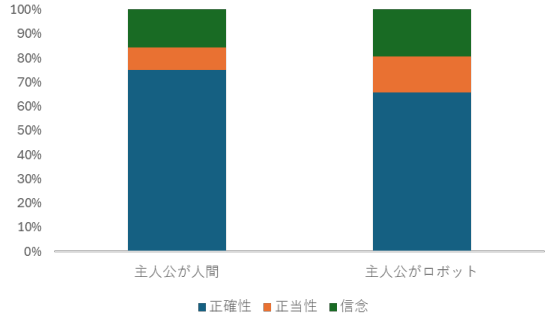


図 3　【予見可能性低・故意性高条件】の結果





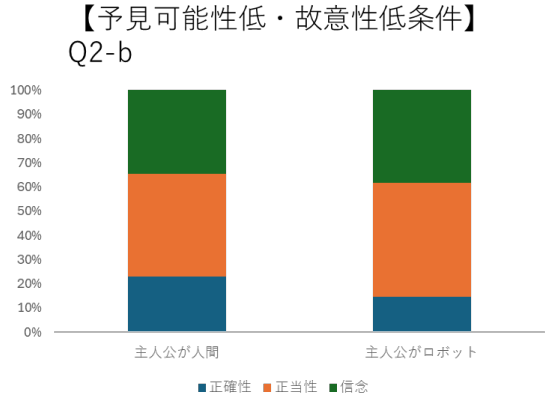


図 4　【予見可能性低・故意性低条件】の結果

ことは，予見可能性が高くかつ故意性が高い事象について，ロボットに責任を帰属させる理由として「ロボットの信念」は人間の場合ほどは重視されないことを示唆している．

【予見可能性高・故意性低条件】では，主人公が責任を負うべきであると考える参加者の割合は，主人公が人間の時のほうがロボットの時よりも多かった．この結果は，前述した「参加者はロボットの自分を律する能力を人間よりも低く評価している」という解釈を補強するものである．一方，この条件では，「責任を負うべきである」と考えた参加者において，その理由は人間が主人公の場合とロボットが主人公の場合とで有意な偏りは見られなかった．このことは，予見可能性が高く故意性が低い事象については，人間に対する責任帰属とロボットに対する責任帰属に大きな差はないことを示唆している．

【予見可能性低・故意性高条件】では，「責任を負うべきではない」と答えた参加者において，「正確性」，すなわち「その知識（予見）が正しいか」を重視する割合が，人間が主人公の場合のほうが，ロボットが主人公の場合よりも大きかった．このことは，予見可能性は低いが故意性は高いと判断される事象において，「知識の正確性」は，ロボットに責任を帰属させる場合には人間の場合ほど重視されないことを示している．

以上の結果から，ロボットの知識および責任の帰属について，以下のようなモデルを導出することができる．

1・ロボットは，自分の行為を律する能力が人間よりも低い．そのため，人間であれば当然責任を負うべき場面であっても，責任を負わなくてもいいと判断される可能性がある（予見可能性が高い場合）
2・予見可能性も故意性も高い事象では，ロボットが自分の行為の結果を本当に信じていたかどうかは，責任の帰属において重要ではない
3・予見可能性は低いが故意性は高い事象については，ロボットに責任を帰属させるにあたって，「ロボットの行為の結果その結果が生じうる」ということが正しいかどうかは大きく考慮されない

1 と 2 からは，「予見可能性が高い事象について，ロボットは自分の行為を律することができず，ゆえにロボットが自分の行為の結果に対する信念を持っていたかどうかは責任の帰属には大きく影響しない」とユーザは考えている、ということが導き出せる．これは言い換えると、ロボットにおいては「予見できること」と「実際に予見した結果を回避する行動をとれること」の間にギャップがある，とユーザは考えているということである．

2 と 3 からは，故意性が高い事象について，ロボットに責任を帰属させるにあたっては「自分の行為の結果を正しく信じていた」かどうかや，そもそも「その行為の結果その結果が生じるということが正しい」かどうかといったことは相対的に重要ではないことが導出される．これは，ユーザが考えるロボットの「故意性」という概念が，人間のそれとはそもそも大きくことなるということを示唆している．

結論

本研究では，「ロボットに対する責任の帰属」という問題を，知識の JTB 仮説と，予見可能性・故意性という枠組みを用いて分析するべく実験を行った．その結果，ユーザが考えるロボットの自律能力やロボットの故意性について，興味深い示唆を与える結果が得られた．

参考文献

[1] Matsui, Tetsuya, and Atsushi Koike. "Who is to blame? The appearance of virtual agents and the attribution of perceived responsibility." Sensors 21.8 (2021): 2646.

[2] Goldman, Alvin I. "What is justified belief?." Justification and knowledge: New studies in epistemology. Dordrecht: Springer Netherlands, 1979. 1-23.

[3] Gettier, Edmund L. "Is justified true belief knowledge?." analysis 23.6 (1963): 121-123.

[4] Matsui, Tetsuya. "A JTB based study of epistemic attribution to robots as “Knowledge”." Discover Artificial Intelligence 5.1 (2025): 329.