解　説

ロボットへの原因と責任の帰属

Attribution of Causality and Responsibility to a Robot

河 合 祐 司*　*大阪大学先導的学際研究機構

Yuji Kawai*　*Institute for Open and Transdisciplinary Research Initiatives, Osaka University

1. は じ め に

何かの失敗や損害を「モノ」のせいにしたことはないだろうか．筆者の両親はしばしば，不調のパソコンを前にして「何もしていないのに壊れた」や「パソコンが不機嫌だ」という．大抵の場合，何かの設定が変更されていることから，本人が原因なのだろう．もちろんパソコンに機嫌という状態や機能はないはずである．それにもかかわらず，人は人工物の取り扱いの失敗の原因や責任†を人工物に帰属させることがある．1995 年の米国での調査 [1] では，計算機科学専攻の大学生の多くがコンピュータにエージェンシー（意図を有し，意思決定する能力）を感じ，医学的な放射線治療システムのエラーによる患者への過放射のケースなどにおいて，大学生の 21%がコンピュータシステム自体に非があると回答したことが報告されている．コンピュータが普及した現在でも少なからず同様の傾向はあるだろう．

将来，人工知能やロボットがより自律的に振る舞うようになり，人々の生活の身近なものになったとき，人が望ましくない事象の原因や責任を人工物に帰属する問題はより深刻になるだろう．高度な人工知能を搭載したロボットはユーザにとってもはや単なる道具ではなく，ある種の主体性を感じさせるエージェントとみなされる可能性が高い．そのロボットに関して事故が起きた場合，ユーザはそのロボットの使い方についての自身の過失だけでなく，たとえその事故とロボットの行動の因果関係が間接的であったとしても，そのロボットやそれを作ったメーカーに責任を問いたくなるかもしれない．そのようなユーザのロボットへの責任帰属は，ロボットの見た目や振る舞いからユーザがロボットへ抱く期待や信頼，エージェンシーが影響すると考えられる．ロボットを設計する際には，このような人が事故や損害の原因や責任を人工物へ主観的に帰属させる心理過程を考慮する必要があるだろう．

社会心理学において，対人の場合の原因や責任の帰属は 1950 年代から研究されている [2][3] が，対人工物の場合についてはあまり検討されてきていない．本稿では，まず，対人における原因と責任の帰属理論を概説する．そして，対ロボットの責任帰属を実験的に検討した研究と，エージェンシーに注目した筆者らの実験 [4] を紹介する．さらに，ロボットが事故を起こす架空の事例映像を用いて筆者らが調査した，その事故の関係者ら（ユーザ，メーカー，および，ロボット）への非専門家の人々による責任帰属について議論する [5].

原稿受付　2019 年 11 月 26 日
キーワード：Causal Attribution, Attribution Theory, Human-Robot Interaction, Autonomous Robot, Responsibility
*〒565-0871 吹田市山田丘 1-1
*Yamadaoka 1–1, Suita-Shi, Osaka 565-0871
†本稿での「責任」は，法的責任ではなく，非専門家の人々が素朴に判断する責任を指す．

2. 対人の帰属理論

社会心理学において「帰属」とは，ある状況で生じた事象（出来事や行為）の原因を推論する心理過程であり，Heider [6] がその帰属理論を初めて提唱した．ある人の行為の原因はその行為者の性格や意図（内的帰属）とその行為者のおかれていた環境や運（外的帰属）に分けられ，判断者はその行為を内的にまたは外的に帰属する．そして，多くの場合，この原因帰属に基づいて，その行為の結果に対する行為者の責任の大きさが判断される，すなわち，責任帰属される （例えば，文献 [3][6]）．責任帰属には，行為と結果の因果性だけでなく，行為者の意図性や予見可能性（行為に先立って，行為者が結果を知り得たかどうか）が考慮される [3]．すなわち，行為者の意図しなかった行為や，結果を予見できなかった行為の結果については，行為者は免責される．これに加え，Weiner [7] は行為者がその原因をコントロールできた（統御可能だった）場合に行為者の責任が大きく帰属されることを主張している．例えば，行為者の能力不足（統御不可能な原因）ではなく，努力不足（統御可能な原因）によって引き起こされた事象に対する行為者の責任は大きくなる．これらの因果性に基づく責任帰属だけでなく，部下の不祥事の責任を上司が問われるといったように，その人の役割や地位に応じた期待や規範からの逸脱（すべきことをしなかった，または，すべきでないことをした）に基づく責任帰属も考えられる [8].

本来，因果関係は曖昧であり，帰属には様々なバイアスがかかる．特に，自己防衛的なバイアスの存在が多く報告されている．自己奉仕バイアスは，成功を自分の能力や努力などに内的帰属させる一方で，失敗を課題の難しさなどに外的帰属させる傾向である [9]．また，ある行為の結果が深刻であるほど，判断者はその責任を行為者に帰属しやすくなるバイアスがあり，その背後には，そうすることで偶然に不幸な事態が発生したのではないから，同様の事態に自分は遭遇しないという判断者の自己防衛的な考えがあると

されている [10]．さらに，同様の事態に判断者自身が陥ったときに，非難が自身に向けられることを回避しようとする欲求から，判断者と行為者の態度や立場が類似していれば，行為者への責任帰属が小さくなるバイアスも報告されている [11]．

3. ロボットへの責任帰属

Hinds ら [12] は，対ロボットの非難／賞賛は対人のものと同程度であることを実験により示した．この実験では，実験参加者と人，または，参加者とロボットが協力して部品を集める課題において，その課題の失敗／成功に対する相手への非難／賞賛を参加者に尋ねた．その結果，人への責任帰属とロボットへの責任帰属の程度には有意な差はなく，ロボットも人と同程度に責任帰属されることが示された．さらに，人らしい見た目のロボットと機械らしい見た目のロボットで比較し，課題の失敗における人らしいロボットへの非難が，機械らしいロボットへの非難に比べて有意に大きいことが示された．

成功は自分のおかげで失敗は自分以外のせいという自己奉仕バイアスが人工物に対しても存在することが，多くの実験で示されている [13]〜[18]．Moon and Nass [13] は，実験参加者とコンピュータがテキストで対話し，意思決定する課題におけるコンピュータへの責任帰属を調査した．このとき，参加者を支配的か服従的かの性格傾向で 2 群に分け，コンピュータにも同様の二つの対話スタイルを設計した．その結果，課題の成功／失敗の帰属に自己奉仕バイアスがみられ，ただし，参加者とコンピュータの性格傾向が類似している場合にはこのバイアスが緩和される，すなわち，課題の成功はコンピュータのおかげでもあり，失敗は自身のせいでもあるという傾向が比較的強くなることが分かった．このことは，参加者が性格の類似するコンピュータを自身と同一視して，自己防衛的に非難を回避し，賞賛を増加させるという，対人でも観察されるバイアス [11] であると解釈できる．

人工物特有の要因である自律性（自律的だと知覚されること）の程度が大きいことで，課題の失敗時に人工物に帰属される責任が大きくなることが報告されている [14] [15] [19]．Kim and Hinds [14] は実験参加者とロボットが部品を集めて運搬する課題で，運搬を自動的に始める自律的なロボットと参加者の合図で出発する自律的でないロボットの条件を設けた．その結果，これらの自律性の違いは課題の成否に関係しないにもかかわらず，課題の失敗時において，自律的に振る舞うロボットのほうが大きく責任帰属されることが示された．

ロボットへの責任帰属におけるエージェンシーの影響に踏み込んだ実験として，van der Woerdt and Haselager [20] は，ロボット単体が課題を達成するだけの能力を有していながら，その課題に失敗する事態を設定した．それを観察した実験参加者は，その失敗の原因が能力不足ではなく，努力不足であると思われるロボットに，失敗の原因が能力不足であると思われるロボットよりも，大きなエージェンシー（自己制御や意思決定の能力）と大きな責任を帰属することが分かった．この結果は，その原因が統御可能である場合に大きな責任が帰属されるという Weiner [7] の対人の帰属モデルと一致する．

4. 心の知覚と原因帰属

上記のように，人工物に対する原因と責任帰属の多くは，対人の帰属理論で解釈可能であるといえる．ただし，人工物においては，その見かけの人らしさや知覚された自律性が重要な要因になる．その理由として，人工物への原因や責任の帰属には，判断者がエージェンシーといった人のような心や機能を推論する，いわゆる擬人化の過程があることが考えられる．

そこで，筆者らのグループでは，擬人化の種類と程度を定量化する「心の知覚尺度 [21] [22]」を用いて，人工物への心の知覚と原因帰属の関係を調査した [4]．心の知覚尺度は，人とロボットを含む多種のエージェントがどのような心を持っていると思うかを評価するものであり，Gray ら [21] は主に二次元の軸で心の知覚を説明できることを示した．一つは，思考や計画，自己制御の能力に関係する「エージェンシー」であり，もう一つは，快や痛み，怒りを感じることに関係する「エクスペリエンス」である†．

筆者らは，実験参加者とエージェントによる囚人のジレンマのような繰り返しゲームの成績によって，参加者の金銭報酬が決まる実験を行った．このゲームでは，お互いが同時に二つの選択肢（「たくさん欲しい」か「相手に譲る」）のどちらかを選ぶ．両者とも「たくさん欲しい」を選ぶと報酬がマイナスになり，両者とも「相手に譲る」を選ぶと報酬がもらえない．片方が「たくさん欲しい」を選び，もう片方が「相手に譲る」を選ぶと，「たくさん欲しい」を選んだほうが報酬を得られる．エージェントはランダムに手を選ぶため，すべての参加者が期待される額ほどの報酬は得られず，ゲーム後にそのゲームに失敗したことを告げられる．参加者はそのゲームの前後で，エージェントに対する心の知覚を評価し，ゲーム後にゲームの失敗が「自分のせい」と思う程度と「相手のせい」と思う程度を評価した．エージェントには，人，ロボット，および，コンピュータの 3 条件を設けた（図 1 下）．

図 1 に，自分よりも相手のせいだと思った相対的な原因帰属（「相手のせい」−「自分のせい」）の程度の平均を示す．この値が大きいほど，エージェントに失敗の原因を大きく帰属したことになる．この図より，今回の実験においては，人は人工物よりも大きく原因帰属されたことが分かる．一方で，ロボットとコンピュータの人工物間の見かけの違いの効果はみられなかった．なお，これらの原因帰属の程度と，実験参加者が獲得した金額や「たくさん欲しい」を選択した回数の間に有意な相関はなかった．

三つの条件での心の知覚スコアを合わせて次元圧縮した

†それぞれ「自律性」と「感情能力」という語訳を当てることもある [23]．

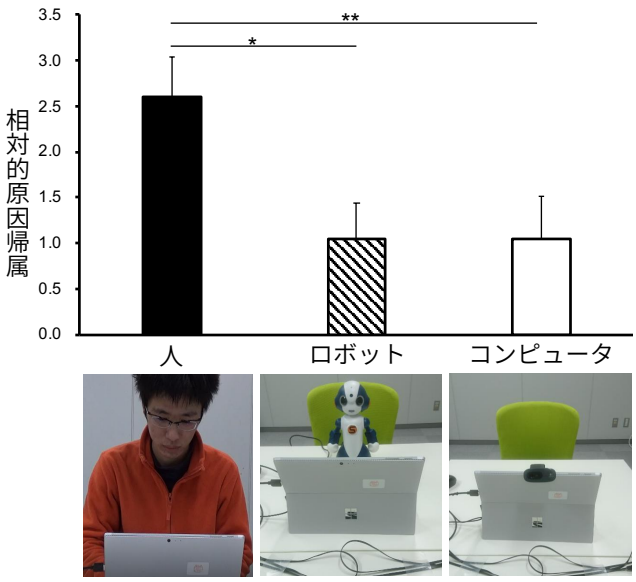


図 1 エージェント別の相対的原因帰属（*: $p < .05$, **: $p < .01$）

表 1 エージェンシー／エクスペリエンスと相対的原因帰属の相関

	エージェンシー		エクスペリエンス	
	ゲーム後	前後の差	ゲーム後	前後の差
すべての条件	−.54 **	−.31 **	.34 **	.05
人	−.61 **	−.67 **	.19	.18
ロボット	−.38 †	−.18	.20	.12
コンピュータ	−.51 **	−.07	.14	−.09

†: $p < .1$, *: $p < .05$, **: $p < .01$

結果，Gray ら [21] が報告したものと同様の「エージェンシー」と「エクスペリエンス」の軸が得られた．自分よりも相手のせいだと思った相対的な原因帰属の程度とゲーム後の心の知覚との相関係数，および，ゲーム前後の心の知覚の差との相関係数を表 1 にまとめる．ゲーム後については，すべてのエージェント条件においてエージェンシーと相対的原因帰属の間の負の相関（の傾向）がみられる．すなわち，エージェンシーを感じない相手に原因を帰属する傾向である．相対的原因帰属と有意な負の相関があった心の知覚の下位項目はエージェンシーに含まれる「自己制御」，「計画」，「思考」および，「記憶」であった．報酬を「相手に譲る」には「自己制御」が必要であり，過去の履歴に基づいて次の手を決めるには「計画」，「思考」，「記憶」の能力が必要である．したがって，課題の達成のための能力を有していないと感じられたエージェントには原因が帰属される傾向にあるといえる．

ゲーム前後のエージェンシーの差と相対的原因帰属の負の相関が，3 条件をまとめたときと人条件において有意であった．この相関は，ゲームを通してエージェンシーのスコアが低下したエージェントには，大きく原因帰属されたことを示す．人条件において有意な負の相関を示した下位項目は，エージェンシーに含まれる「思考」，「コミュニケーション」，「自己制御」，「モラル」，「感情認識」，「計画」，「記憶」に加えて，エクスペリエンスに含まれる「意識」，「痛み」，「恐れ」であった．すなわち，これらの能力についての参加者の事前の期待に反することが，原因帰属につながったと考えられる．

本実験において，課題遂行の能力が不足していると思われることと，事前に期待した能力に反して低い能力だと思われることが，帰属の要因であることが示された．前者はゲームの失敗の直接的な原因の帰属を反映していると考えられる．一方，後者について，Weiner [7] のモデルに基づくと，人にはゲームを成功させる能力があるにもかかわらず，その努力を怠ったように感じられたことが大きな帰属につながったと解釈できる．また，「モラル」や「コミュニケーション」，「痛み」の下位項目スコアの減少が関係していることから，規範を重視するモデル [8] に基づいて，人は相手のことを考慮して，「相手に譲る」選択をすべきであるという道徳的な規範から逸脱したことが大きな帰属につながったとも解釈できる．これらの統御可能性や規範の逸脱は責任判断において考慮される要因であるとされるが，「相手のせい」という言葉には責任帰属のニュアンスも含まれるため，これらの要因が影響したと考えられる．なお，事前の期待に反して低い能力であることがそのエージェントの悪い印象につながることは，ヒューマンエージェントインタラクション研究の分野では「適応ギャップ」として知られている [24]．

一方，今回の実験では，この期待に反することの影響は人工物条件ではみられなかった．このことは，人工物への期待や人工物に適用される規範が，対人のものと異なる可能性を示唆する．Malle ら [25] はトロッコ問題のようなモラルジレンマの状況において，実験参加者は人には道徳的な決定を，ロボットには合理的な決定を期待することを示した．今回においても，参加者は人工物には「相手に譲る」という道徳的な振る舞いを期待していなかったことが考えられ，これが小さな帰属につながった可能性がある．今後，より精緻な実験によってこれらの推測を検証し，人への帰属と人工物への帰属の類似性と差異を明らかにしたい．

5. ロボットの事故事例映像を用いた責任帰属の調査

ロボットが環境や人との相互作用や学習によって創発的に行動することは，ロボットが複雑な実世界に適応するための重要な機能であるといえる．一方で，その創発行動が設計者の事前の意図どおりではない，事故などの社会的に望ましくない結果をもたらす可能性を完全に否定できない．ロボットの用途が汎用的になるほど，すべての作用の可能性を事前に予見することは困難を極めると考えられる．このような場合，ロボットの事故の責任を人々はどのように関係者へ帰属するのだろうか．

筆者らのグループは，ロボットの故障や明らかな欠陥ではなく，社会的な相互作用の結果として発生したロボットの事故における責任帰属についての多様な意見を集めるためのワークショップを開催した [5] [27]．人とロボットの社会的な相互作用の状況を理解しやすくするために，筆者らは架空の事故事例の映像を制作した（巻頭カラーページ参

照）．この映像はウェブ上で一般公開されている [28]．ワークショップの参加者はこの映像を視聴した後に，関係者（ユーザ，メーカー，ロボット）それぞれの事故に対する責任を問うアンケートに答え，その理由を自由記述で回答した．合計 125 名以上の参加者のうち，本稿では，この映像を視聴した 18 歳以上の日本国籍の 47 人（女性 28 名，男性 19 名）から収集した，関係者への責任帰属の理由を筆者がまとめたものを以下に示す．

A さんに責任があると思う理由
- A さんは優先順位（薬かセキュリティか）をロボットに明確に伝えていないから．
- ロボットがどんな対応をするのか，A さんはしっかり理解できていないから．
- A さんはロボットを信用しすぎているから．
- 人が自己の判断と責任で機械を扱うべきだから．

A さんに責任がないと思う理由
- A さんのロボットの使い方には問題がないから．
- A さんはロボットに「知らない人が来たら鍵を開けないように」と指示していたから．
- ユーザに責任を求めるのは酷だから．

メーカーに責任があると思う理由
- ロボットのプログラムを作ったのはメーカーだから．
- ホームセキュリティロボットとして社会に出すまでのレベルに達していないから．
- メーカーはどんなことが起きてもよいように予想すべきだから．

メーカーに責任がないと思う理由
- ロボットは A さんの指示どおりに正当な判断をしたから．
- 完璧な技術はあり得ないので、メーカーに 100%の責任があると考えるのは非合理的だから．

ロボットに責任があると思う理由
- ロボットが A さんの指示に反して鍵を開けたから．
- ロボットが事故の原因の張本人だから．

ロボットに責任がないと思う理由
- ロボットはプログラムどおり人命を優先しただけだから．
- もしロボットが人であったら、人道的な判断だったと思うから．
- ロボットに責任能力はないから．

A さんに責任があると思う主な理由は，A さんがロボットに「ありがとう．でもダメよ」という曖昧な指示をしたことであった．この背後には，ロボットに対するユーザの過信や過剰な擬人化があると考えられる．この事例と同様に，オートパイロットシステムへの操縦者の過信が操縦者の誤った行動につながることが指摘されている [29] [30]．さらに，このようなロボットや人工知能への過信の問題は，それらの技術の発展と普及に伴ってより深刻になることが予想されている [31]．したがって，ロボットや人工知能の特性や性能の限界を適切に理解するリテラシーの発展がユーザに必要であるといえる [32]．

メーカーの責任に関して，メーカーはあらゆる可能性を予見し，事故を回避すべきであるという意見があった．事故を起こさせない作り手の努力は必須であり，そのためのガイドラインが国際的に策定されつつある（例えば，文献 [33] [34]）．一方で，このような事故の責任をメーカーが負うことが合理的でないという意見も少なからずあった．このことは，過剰な法的責任をメーカーに帰することで，新しい技術の開発が萎縮することを懸念する考え [35] に近いといえる．それに対して，民事法では保険や基金の運用が考えられており [36]，刑事法では追訴裁量の活用が考えられている [35] [37]．

6. お わ り に

本稿は，対人の帰属理論に基づいて，人工物への原因と責任の帰属を解説した．特に，自律性に関する能力であるエージェンシーを人工物に対して事前に期待したり，実際に知覚したりすることが，人工物への原因や責任の帰属に大きく影響することが複数の実験で示されている．このことは，ロボット設計に際して，その客観的な性能だけでなく，ユーザのエージェンシー知覚も考慮に入れる必要性を示唆する．しかし，人工物への原因と責任の帰属の要因について研究の余地が大いにある．例えば，ロボットや人工知能への肯定／否定的態度，ロボットに関する判断者の立場といった個人的傾向や，ロボットに対する期待や規範の文化差や世代差についても検討すべきだろう．また，「責任」の概念は多義的であるため，人工物やその製造者にどのような責任を考えているのかについて詳細化する必要もある．責任概念は出来事に関与したかどうかという意味での責任や，非難や制裁を負担する意味での責任，責務としての責任を含む [38]．ロボットやそのメーカーがどれだけ事故に関与したと思われるか，どのような刑罰や賠償を負うべきだと思われるか，どのような役割や義務を持つべきだと思われるかも重要な問いである．

ロボットの性能への（過剰な）期待や信頼の問題は，映像を用いた調査でも，ユーザの課題として議論した．ユーザのロボットへの過信は予期せぬ事故の原因になりえ，さらに，ロボットやメーカーへの過剰な責任帰属につながりえる．ユーザはロボットの性能を適切に理解し，メーカーはロボットの仕様や使用方法をユーザに説明するというお互いの努力が必要であろう．これに加え，ロボット開発者とメーカー，および，将来のユーザである非専門家の意見を考慮して，ロボットの事故に対して適切に帰責する社会制度設計も不可欠である．社会的に受容されるロボット開発に向けて，ロボット研究者と開発者といった作り手だけでなく，利用者や社会制度の専門家を含めた相互理解と協働が求められる．

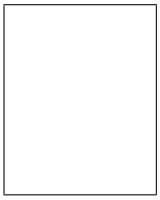
謝　辞　第 5 章のワークショップは日本科学未来館市民参加型実験「オープンラボ」事業として実施されたものである．その映像制作からワークショップ開催に至るまで，片平圭貴氏と宮田龍氏をはじめとした日本科学未来館の多くの科学コミュニケーターとスタッフにご協力いただいた．

この場を借りて改めて感謝を申し上げる．

本研究は，JST 戦略的創造研究推進事業 (RISTEX)「人と情報のエコシステム」研究開発領域「自律性の検討に基づくなじみ社会における人工知能の法的電子人格」 (JPMJRX17H4)，および，JSPS 課題設定による先導的人文学・社会科学研究推進事業「工学・脳科学をエビデンスとした社会的基盤概念と価値の創生」による研究成果の一部である．これらのプロジェクトメンバーに深く感謝する．

参 考 文 献

[1] B. Friedman: ""It's the computer's fault": Reasoning about computers as moral agents," Proc. of the ACM Conference on Human Factors in Computing Systems, pp.226–227,1995.
[2] 唐沢，松村，奥田（編著）：責任と法意識の人間科学．勁草書房，2018.
[3] K.G. Shaver: An Introduction to Attribution Process. Cambridge, MA: Winthrop, 1975. [K.G. シェーバー（著），稲松・生熊（訳）：帰属理論入門　対人行動の理解と予測．誠信書房，1981.]
[4] T. Miyake, Y. Kawai, J. Park, J. Shimaya, H. Takahashi and M. Asada: "Mind perception and causal attribution for failure in a game with a robot," Proc. of the 28th IEEE International Conference on Robot and Human Interactive Communication, TuCT1.2, 2019.
[5] Y. Kawai, T. Inatani, T. Yoshida and K. Matsuura: "Exploring future rules for AIs with citizens using a fictitious case video: A workshop report," Proc. of International Workshop on "Envision of Acceptable Human Agent Interaction based on Science Fiction," 2019.
[6] F. Heider: The Psychology of Interpersonal Relations. New York: Wiley, 1958. [フリッツ・ハイダー（著），大橋正夫（訳）：対人関係の心理学．誠信書房，1978.]
[7] B. Weiner: Judgments of Responsibility: A Foundation for a Theory of Social Conduct. New York/London: Guilford Press, 1995.
[8] V.L. Hamilton: "Who is responsible? Toward a social psychology of responsibility attribution," Social Psychology, vol.41, no.4, pp.316–328, 1978.
[9] D.T. Miller and M. Ross: "Self-serving biases in the attribution of causality: Fact or fiction?," Psychological Bulletin, vol.82, no.2, pp.213–225, 1975.
[10] E. Walster: "Assignment of responsibility for an accident," Journal of Personality and Social Psychology, vol.3, pp.73–79, 1966.
[11] K.G. Shaver: "Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident," Journal of Personality and Social Psychology, vol.14, pp.101–113, 1970.
[12] P.J. Hinds, T.L. Roberts and H. Jones: "Whose job is it anyway? A study of human-robot interaction in a collaborative task," Human-Computer Interaction, vol.19, no.1, pp.151–181, 2004.
[13] Y. Moon and C. Nass: "Are computers scapegoats? Attributions of responsibility in human-computer interaction," International Journal of Human-Computer Studies, vol.49, no.1, pp.79–94, 1998.
[14] T. Kim and P. Hinds: "Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction," Proc. of the 15th IEEE International Symposium on Robot and Human Interactive Communication, pp.80–85, 2006.
[15] A. Serenko: "Are interface agents scapegoats? Attributions of responsibility in human-agent interaction," Interacting with Computers, vol.19, no.2, pp.293–303, 2007.
[16] K.L. Koay, D.S. Syrdal, M.L. Walters and K. Dautenhahn: "Five weeks in the robot house: Exploratory human-robot interaction trials in a domestic setting," Proc. of the 2nd International Conferences on Advances in Computer-Human Interactions, pp.219–226, 2009.
[17] S. You, J. Nie, K. Suh and S.S. Sundar: "When the robot criticizes you...: Self-serving bias in human-robot interaction," Proc. of the ACM/IEEE 6th International Conference on Human-Robot Interaction, pp.295–296, 2011.
[18] G.N. Vilaza, W. Haselager, A. Campos and L. Vuurpijl: "Using games to investigate sense of agency and attribution of responsibility," Proc. of the XIII Brazilian Symposium on Computer Games and Digital Entertainment, pp.393–399, 2014.
[19] A. Waytz, J. Heafner and N. Epley: "The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle," Journal of Experimental Social Psychology, vol.52, pp.113–117, 2014.
[20] S. van der Woerdt and P. Haselager: "When robots appear to have a mind: The human perception of machine agency and responsibility," New Ideas in Psychology, vol.54, pp.93–100, 2019.
[21] H.M. Gray, K. Gray and D.M. Wegner: "Dimensions of mind perception," Science, vol.315, no.5812, p.619, 2007.
[22] 上出，高嶋，新井："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.
[23] 月本敬："実在ロボットに対する不気味の谷現象に関する完成評価研究 –心の知覚の観点から–"，日本感性工学会論文誌，vol.16, no.3, pp.293–298, 2017.
[24] T. Komatsu and S. Yamada: "Adaptation gap hypothesis: How differences between users' expected and perceived agent functions affect their subjective impression," Journal of Systemics, Cybernetics and Informatics, vol.9, no.1, pp.67–74, 2011.
[25] B.F. Malle, M. Scheutz, T. Arnold, J. Voiklis and C. Cusimano: "Sacrifice one for the good of many?: People apply different moral norms to human and robot agents," Proc. of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction, pp.117–124, 2015.
[26] J. Nadler: "Flouting the law," Texas Law Revew, vol.83, pp.1399–1441, 2005.
[27] 日本科学未来館オープンラボ「一緒にさがそう未来のルール　～ロボットの事故は誰かのせい？」開催レポート：http://www.ams.eng.osaka-u.ac.jp/ristex/index.php/openlab-report/
[28] ロボットの事故事例：https://youtu.be/8wO4sqZYUOI
[29] R. Parasuraman and V. Riley: "Humans and automation: Use, misuse, disuse, abuse," Human Factors, vol.39, no.2, pp.230–253, 1997.
[30] 稲垣敏之：人と機械の共生のデザイン．森北出版，2012.
[31] A.R. Wagner, J. Borenstein and A. Howard: "Overtrust in the robotic age," Communications of the ACM, vol.61, no.9, pp.22–24, 2018.
[32] M.H. Jarrahi: "Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making," Business Horizons, vol.61, no.4, pp.577–586, 2018.
[33] IEEE Global Initiatives on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design, First edition. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf, 2017.
[34] European Commission: Ethics Guidelines for Trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, 2019.
[35] 稲谷龍彦："技術の道徳化と刑事法規制"，松尾陽（編著），アーキテクチャと法　法学のアーキテクチュアルな回転？．pp.93–128，弘文堂，2017.
[36] 栗田昌裕："ロボット事故の民事責任"，日本ロボット学会誌，vol.38, no.1, pp.●–●, 2020.
[37] 稲谷龍彦："ロボット事故の刑事責任"，日本ロボット学会誌，vol.38, no.1, pp.●–●, 2020.
[38] 瀧川裕英：責任の意味と制度　負担から応答へ．勁草書房，2003.



河合祐司（Yuji Kawai）

2016 年大阪大学大学院工学研究科博士後期課程単位取得満期退学．博士（工学）．同大学院助教を経て，2019 年より大阪大学先導的学際研究機構特任講師，現在に至る．ヒューマンロボットインタラクションとニューラルネットワークの研究に従事．人工知能学会，日本神経回路学会，日本認知科学会の会員．（日本ロボット学会正会員）