

社会的影響に関する Web 調査に基づく 人-エージェント共存社会の意見ダイナミクスシミュレーション

Opinion Dynamics in Human-Agent Society Based on Web Surveys Regarding Social Influence

布施 陽太郎¹ *

Yotaro Fuse¹

¹ 富山県立大学

¹ Toyama Prefectural University

Abstract: 対話型エージェントは人の意見形成に社会的影響を与え得る。エージェントが社会的影響源として混在する社会での意見ダイナミクスをシミュレートし、その社会でのマクロ現象を探究すべきである。本研究では、架空の状況において人またはエージェントから提示される意見が個人の判断に与える影響をオンライン調査で取得し、その結果に基づいてシミュレーション内のエージェントの意見更新則を設計する。設計したエージェントがもたらす合意形成への影響をシミュレーションにより探究する。

1 はじめに

人とロボットの相互作用に関する研究は、人とロボットが同じ生活空間や社会制度の中で共存する未来を見据え、ロボットが人に受容されうる振舞いを明らかにすることを目的の1つとして進められてきた [1, 2]。さらに、ロボットは大規模言語モデルの発展によって対話能力を飛躍的に高まる可能性も指摘されている [8]。加えて、実態を持つロボットだけではなく、画面内の仮想エージェントや AI ボットもまた、人々と自然に対話可能となる。そのような状況では、ロボットやエージェントが単なる道具としてではなく社会的存在として扱われ得る場面が増え、人の判断や意見形成に対してロボットが影響を及ぼしうることで、重要な論点になりつつある。

人が計算機やメディアを社会的存在として扱う傾向は、CASA 理論 (Computers Are Social Actors) や Media Equation として整理されてきた [5, 6]。実際、人々がロボットに精神を帰属するけいこうがあることも様々な研究によって報告されている [7]。こうした背景のもと、ロボットが人の行動や判断に影響を与える可能性を示す実証研究も蓄積されてきた [10, 11, 12, 13, 14, 15, 16]。しかし、ロボットの影響は外見や能力だけで一意に定まるのではなく、相互作用の状況、ロボットへの態度など複数要因によって現れるため、相互作用において観察された人への社会的影響が人間社会全体におい

てどのような帰結を引き起すのか、という問いへ接近することは容易ではない。

この点は、AI 倫理や ELSI の観点からも重要である。ロボットが人の意思決定に影響を与えることを前提に、ロボットが社会規範に準拠して振舞う際の問題が定性的に議論されている [17]。特に、Coggins らはロボットの振舞いが人々が共有する社会規範の変容を引き起す可能性を示唆している [17]。一方で、社会的影響は個々の相互作用にとどまらず、個人が複数の集団に属しながら相互に影響し合うことで、合意形成や分極、多数派の推移といったマクロな現象として現れうる [18, 9]。したがって、人-ロボット相互作用における「個人レベルの影響の受け方」が、社会全体の意見ダイナミクスとしてどのような帰結をもたらすのかを、定量的に検討できる枠組みが求められる。

しかし、多人数の人とロボットが同時に関与する状況を実験として設計・実施することは、時間や費用、再現性の観点から現状困難である。さらに、社会全体に十分普及した高度な社会ロボットを前提とする実データの収集も難しい。この制約を補う方法として、意見伝播や合意形成を扱う社会シミュレーション (オピニオンダイナミクス [22, 21]) があり、個人間の相互作用の規則からマクロな現象が創発する過程を検討できる。一方で、先行研究では、更新規則やパラメータをモデル設計者が設定し、その値の変化に対する社会的現象の発生を探索することに主眼が置かれてきたため、実証データによるグラウンディングが十分でないという問題がある [4]。結果として、モデルが示す帰結を現実

*連絡先：富山県立大学情報工学部データサイエンス学科
〒939-0398 富山県射水市黒河 5180
E-mail: fuse@pu-toyama.ac.jp

の人-ロボット相互作用へとどの程度適応できるのかに関して、説明可能性の根拠が弱くなりやすい。

そこで本研究では、抽象的な意見ダイナミクスモデル [22, 21] が持つ探索可能性を保持しつつ、その自由度の一部を実データによって制限する方法を提案する。具体的には、対話相手が人間である場合とロボットである場合とで、提示された意見と自分の意見の隔たりに応じて意見を更新し得るかどうかを Web 調査結果から推定し、推定結果をシミュレーションに組み込む。仮想社会のシミュレーションでは、人と人、人とロボットが相互作用するが、その相互作用の仕方に Web 調査結果を関連付ける。このことによって本研究の方法は、従来の意見ダイナミクスモデルにおいて、意見更新が生じるかどうかを決定する規則や確率が、主としてモデル設計者の仮定に基づいて設定されてきた点に着目し、そのパラメタ設定の自由度の一部を実データによって制約する。その結果、人とロボットが混在する集団における意見形成の帰結が、意見間距離や対話相手の種類に応じて更新が生じるかどうかを決定するミクロな生成規則を、仮定的に設定した場合と、Web 調査から推定された判断傾向に基づいて設定した場合とで、どの程度変化し得るのかを検討可能にする。これにより、社会レベルで観測される意見形成の差異を、モデル上の仮定の違いとしてではなく、Web 調査という個人レベルで観測された判断傾向の違いに基づいて解釈することが可能となる。

よって本研究は、HAI 分野の対話実験や Web 調査といった実証データを、オピニオンダイナミクスの社会シミュレーションに接続するための橋渡しとして位置づけられる。すなわち、個人レベルで観測される人同士と人エージェント間の社会的影響の発生をデータから推定し、それを仮想社会内の相互作用の意見更新規則として組み込む。それによって、HAI の実験知見を、個人レベルの相互作用にとどめるのではなく、人ロボット社会全体における意見形成の帰結へと拡張して説明可能とする枠組みを提示することを目的とする。

2 方法

2.1 概要

本研究では、エージェントの振舞いに関する Web 調査によって得られた知見を、オピニオンダイナミクスに関する社会シミュレーションの更新ルールの一部として組み込む。Web 調査の目的は、参加者が自らの判断を示した直後に、他者（知人または対話ロボット）から意見を提示された場合に、その意見を返答として取り入れて表明できる可能性がどの程度あるのかを、抵抗感の評定に基づいて把握することにある。得られた

結果を基に、社会シミュレーションでは、意見の隔たりの大きさや情報源の違いに応じて意見更新の有無の傾向が変化するような更新規則を設定し、そのような個人レベルの更新傾向が社会全体の意見分布や多数派意見や合意形成や分極といったマクロな人エージェント共存社会のダイナミクスにどのように影響するかを検討する。

本研究は富山県立大学「人を対象とする研究」倫理審査部会の承認を得て実験は実施された（受付番号 R7-22）。

2.2 Web 調査

本研究では、オンライン調査として、トロッコ問題を題材としたヴィネット課題を 500 名の Yahoo クラウドソーシングのユーザに対して実施した。参加者はまず研究目的・匿名性・正誤のない判断課題であること等の説明を読み、同意したうえで年齢と性別を回答した。次に、暴走するトロッコが本線上の 5 人に衝突する状況と、レバーを操作すると別線路上の 1 人に衝突する状況が提示され、参加者は「レバーを引かない（左端）/引く（右端）」を両端とする 0-100 の連続スライダーにより、自身の初期判断 x_0 を回答した。

図 1 は Web 調査でのブラウザ上での回答インタフェースを示す。参加者はトロッコ問題への初期判断を回答後、参加者には図 1 の画面が提示され、「自分の判断を伝えた直後に、相手と短い意見交換があった」という想定が提示された。その架空の相手は参加者の知人（同僚や友人）または人工知能を搭載した対話ロボットのいずれかとして参加者に提示された。相手の意見は 0-100 スライダー上の位置 x_s と短い理由文によって示され、理由文は x_s が引く側と引かない側のどちらにあるかに応じて、功利主義的（レバーを引くことによる人的被害の最小化）または義務論的（レバーを引かないことによる意図的加害の回避）のテンプレートが自動的に選択される形式とした。なお、架空の相手とその意見 x_s は参加者の初期判断送信後にランダムに決定され、参加者画面には固定スライダーとして提示された。

相手の意見提示を確認した後、参加者が返答として相手に口頭で伝えると仮定する意見（候補意見 x_λ ）に対する抵抗感を 4 件法（抵抗感が強い/抵抗感がある/抵抗感はない/抵抗感はない）で回答した。候補意見は、参加者の初期判断 x_0 と相手意見 x_s の間の補間点 $x_\lambda = x_0 + \lambda(x_s - x_0)$ として提示し、まず $\lambda = 0.5$ を呈示した。参加者の回答において、抵抗感が低い（「抵抗感はない/抵抗感はない」）を選択する場合、次のステップで $\lambda = 0.75$ の候補意見 x_λ をインタフェースは提示する。その時当該参加者にとって、抵抗感が高い場合は $\lambda = 0.25$ をさらに追加で呈示した。その上、

抵抗感の質問

あなたは最初に、自分の判断を相手に伝えました（灰色のスライダー）。その後、相手（対話ロボット）はあなたの判断を見ただうえで意見を述べました（オレンジのスライダー）。

灰色：あなたが最初に相手へ伝えた意見
左端：レバーを引かない 右端：レバーを引く
強く左 どちらともいえない 強く右

オレンジ：相手が述べた意見
左端：レバーを引かない 右端：レバーを引く
強く左 どちらともいえない 強く右

次に、あなたが相手への返答として、下に示された意見（候補のスライダー）を口頭で相手に伝えたとしたら、どの程度抵抗感がありますか？（迷う場合は、より抵抗感が強い方を選んでください。）

青色：あなたが返答として相手へ伝えると仮定する意見（候補）
左端：レバーを引かない 右端：レバーを引く
強く左 どちらともいえない 強く右

この候補を伝えることへの抵抗感

抵抗感が強い
 抵抗感がある
 抵抗感はありません
 抵抗感はない

次へ

図 1: オンライン調査で用いられたアンケート回答用のインタフェース。トロッコ問題におけるレバー操作に関する参加者自身の意見 x_0 を回答した後、架空の対話相手が x_0 とは異なる意見 x_s を表明する。次に、その意見に少し近づいた候補意見 x_λ をインタフェースは出力し、その候補意見を参加者自身の意見として表明することに関する抵抗感を調査する。

候補意見 $x_{\lambda=0.75}$ でも抵抗感が低い場合に限り、架空の相手の意見そのものである候補意見 $x_{\lambda=0.75}$ を呈示した。これらの場合分けによって、参加者が架空の相手の意見に対して少なくとも一部を取り入れて返答できると判断したかどうかを明らかにする。本実験において、もし参加者が「抵抗感はありません/抵抗感はない」を選択する場合、そのことは対話相手から社会的影響を受けたとみなす。加えて、品質管理として注意確認項目を設け、特定の指示に従っているかを 0-100 スライダーで検査した。

2.3 シミュレーション

本研究でのシミュレーションの方法は先行研究 [4] を参考にした。

2.3.1 モデル

本研究では、Watts–Strogatz モデル [19] によって生成されたスモールワールド性を持つ社会ネットワークに配置された人エージェントとロボットエージェントのダイナミクスを観察する。スモールワールド性とは、非常に多くの個体（ノード）が存在する大規模なネットワークにおいても、ほとんどのノード同士がごく短い経路（リンクの数）で繋がっているという性質を指す [20]。現実の人々の社会ネットワークはスモールワールド性を持つとされており、その性質を再現するネットワークを生成する Watts–Strogatz モデルを本研究では採用する。

加えて、ロボットはスモールワールド性を持つネットワークに組み込まれるというよりも、そのネットワークに新たにノードとして追加される存在であると仮定する。すなわち、ある人がロボットを手に入れ、そのロボットがその人とその周囲の人々とつながりを新たに形成することを想定する。自然発生的な社会ネットワークに新たにロボットを導入することが、現実の人口ロボットネットワークに即したシミュレーション用のネットワーク生成につながると考えられる。

図 2 とアルゴリズム 1 にスモールワールド性を持つ人エージェントの社会ネットワークにロボットエージェントを追加するための過程を示す。アルゴリズム 1 に従って生成された人社会ネットワークにロボットを追加することによって、人口ロボット社会ネットワークを生成する。このアルゴリズムによって、人社会ネットワーク N_H に n_R 体のロボットエージェントとそのエージェントとのつながりが追加された人口ロボット社会ネットワーク N_{HR} が生成される。

Algorithm 1 人口ロボットネットワーク N_{HR} 生成手順

- 1: 人エージェント H のみによって構成されるネットワーク N_H を生成 (図 2 左参照)
- 2: **for** $i = 1$ to n_R **do**
- 3: N を構成するある人エージェント H_n をランダムに選択 (図 2 中央参照)
- 4: H_n と H_n の近傍のエージェントで構成される集団 G_n を導出
- 5: 集団 G_n の各エージェントにロボットエージェント R_n を接続 (図 2 右参照)
- 6: **end for**

2.3.2 エージェント

人間エージェントは Deffuant モデル [22] に基づく意見の更新方法を採用する。具体的には、自身の意見を連続値 $x \in [0, 1]$ として保持し、ランダム逐次更新の 1

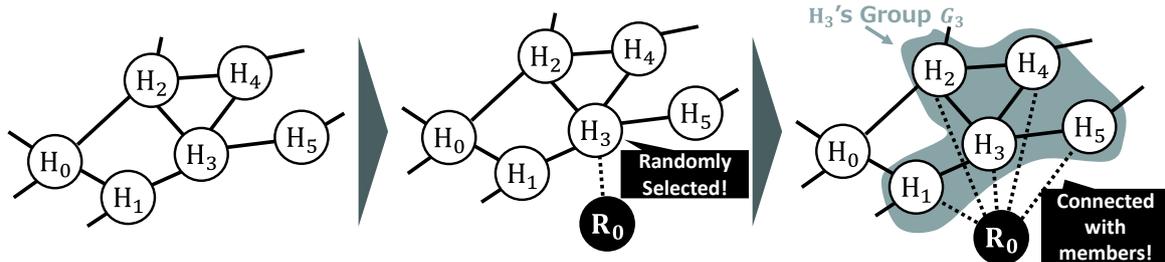


図 2: 形成されていた人エージェントのみで構成されるネットワーク (左) に 1 体のロボットエージェントを追加する流れ. ランダムに選ばれた人エージェント H_3 がロボットエージェント R_0 の所有者と見なされ (中央), H_3 の近傍の人エージェントと R_0 はつながりを生成する (右). このことは, 所有者の H_3 とその周辺の人エージェントの集団メンバとして R_0 が参加することを意味する.

ステップごとに近傍の他のエージェントとのペアを形成した時にその意見 x を更新する. 各ステップでは, まず自分 i の格子位置の近傍セルに存在する相手 (人間エージェントまたはロボットエージェント) の中で意見差 $d = |x_j - x_i|$ が信念許容幅 ϵ 以下の相手のみを相互作用候補集合に加える. これは, 人間エージェントが自身より ϵ 以上離れた意見を持つエージェントのことを無視することによる, そのエージェントの個性やバイアスを表現する. 候補が存在しない場合は意見を据え置く. 候補が存在する場合は候補から相手 j を 1 体ランダムに選択し, その相手が人間かロボットかを判定したうえで, 意見差 d に依存する更新確率 $p_{\text{upd}}^{\text{source}}(d)$ (source = Human or Robot) に基づいて更新の実行可否を決める. 更新が実行される場合に限り, 収束率 μ を用いて $x_i \leftarrow x_i + \mu(x_j - x_i)$ で自己意見を相手方向へ移動させる. 以上により, 本モデルは ϵ による相互作用相手の選別条件, 相手種別 (人もしくはロボット) 条件, 意見距離に依存する更新確率条件の 3 条件を組み合わせた確率的意見更新を, 人間エージェントの逐次更新規則として実装している.

ロボットエージェントは先行研究 [3] で提案されている意見更新法則を採用する. ロボットエージェントはステップ t において, 環境を観察し, その観察を基に内部状態を更新し, その内部状態に基づいて行動を出力する. この設計によってロボットエージェントは自身と接続している各エージェントが持っている意見 x を観察し, その意見の集合の中で多数派意見と推定される意見を次のステップでの自分の意見として保持する.

以上のような設計にすることによって, 仮想社会内での人エージェントは意見 x_i を持ち, 観測可能なエージェントが保持する意見 x_j の内, 閾値 ϵ よりも距離が近い ($d = |x_i - x_j| < \epsilon$) エージェント j のみと相互作用して, 意見 x_i を更新する. これは人間が好みやバイアスを持つことを表現している. 一方で, ロボットエージェントは周囲の全てのエージェントの意見を観察し,

その意見の集合の中でその集合内での多数派の意見と見なされる x_{majority} を自身の意見として採用する. これは, ロボットが周囲の人々から受容されるために社会的振舞うことを目論んだ結果, 好みやバイアスを持つ人とは異なり, 戦略的に

2.3.3 パラメータ

表 1 には固定されたパラメータが示されている. 各エピソードにおける初期意見の分布は seed 値に基づいて統一されており, 制御パラメータの変化によって発生する変化のみを原因としたマクロへの影響を観察する.

表 2 に基づき, 制御パラメータとして n_R , ϵ , $e_{R \rightarrow H}$, SBWS を変化させた混合モデルを用いて, 複数のシナリオをシミュレーションする. 1 度のシミュレーションの実施において, n_R , ϵ , SBWS, seed 値に個別の値が代入される.

意見閾値の大小は人エージェント社会全体での合意形成のしやすさに影響する. ϵ を 0.5 もしくは 0.25 に設定することで, 人エージェントが全体的な合意形成を起こしやすい人工社会と起こしにくい社会でのシミュレーションを実施する. ロボットエージェントの数 n_R は人ロボット社会ネットワークに参入するロボットの数を意味する. 特に, 本研究においてロボットは人に受容されるために他者への同調を目的とする. 観測範囲内の他者が共有する多数派意見に同調するロボットの数が社会全体で増えていくことが社会での合意形成にどのような影響を及ぼすのかを評価する. 加えて, パラメータ SBWS を制御することで, 人エージェントの意見更新規則が Web 調査により経験的に推定された更新確率分布 $p^{\text{upd}}(d)$ に基づくか否かを切り替えることができる. 具体的には, Ture のとき, 各エージェントは刺激意見 x_s と自己意見 x_0 の距離 d に応じて, Web 調査から得られた距離依存の更新傾向を参照し, 更新の生起確率が経験データに基づく更新を行う. 一

方で False のとき、エージェントは Web 調査分布を参照せず、相互作用相手 j が人エージェントであってもロボットエージェントであっても必ず意見更新を行うよう、 $p^{\text{upd}}(d) = 1$ となる。

したがってそのパラメータは、社会シミュレーションにおける、人同士の相互作用と人ロボット同士の相互作用が実データに紐づけられた更新過程を採用しているか否か、という設計上の前提を明示的に表すパラメータである。この切替により、経験的根拠に基づく挙動と抽象モデルに基づく挙動の差が多数派意見の変動などのマクロな意見動態に与える影響を比較できる。

表 1: 設定パラメータ.

Model Parameter	
Opinion x	[0.0, 1.0]
Number of Episodes (seed value)	100
Max Step	1000
Watts–Strogatz Model Parameter	
Number of Humans n	100
Number of nearest neighbors k	4
Rewiring Probability p	0.1
Robot Agent Parameter	
Increment of Opinion δ	0.01
number of set of actions $ X_{\text{self}} $	101
Learning rate α	0.9
Discount factor γ	0.1
Gaussian parameter kurtosis	0.01

表 2: 制御パラメータ.

Number of Robots n_R	0, 10, 25, 50
Tolerance Threshold ϵ	0.25, 0.5
Simulation Based on Web Survey	
SBWS	True/ False
Seed Value	0–99

2.4 データ分析

2.4.1 アンケート結果の処理

Web 調査データから、刺激意見と初期意見の隔たりに応じて人が意見を更新する確率を、情報源（人間の知人/対話ロボット）別に推定した。まず、各回答について初期意見 x_0 と刺激意見 x_s の距離 $d = |x_0 - x_s|$ を算出し、意見更新の有無は更新強度パラメータ λ^* に基づいて二値化した。具体的には、 $\lambda^* > 0$ を「更新あり」、 $\lambda^* = 0$ を「更新なし」とみなし、距離 d に条件づけた更新確率 $p^{\text{upd}}(d)$ を推定対象とした。

推定にあたっては、距離 d の各値における更新率をそのまま集計するのではなく、サンプル数が限られる距離帯でも極端な振れが生じにくいよう、距離を連続変数として扱い、カーネル密度推定 (KDE) に基づく平滑化を用いて $p^{\text{upd}}(d)$ の曲線を得た。この処理を知人 (Human) と対話ロボット (Robot) の各条件で独立に実行することで、刺激意見を出力する架空の相手の属性の違いによる距離依存の更新傾向を比較可能にした。なお、距離が大きい領域では観測数が少なく推定が不安定になるため、可視化ではサンプルが十分に存在する距離範囲に表示を限定した。

2.4.2 シミュレーション結果の分析

本研究のシミュレーションは、人社会にロボットが参入することによって、その社会のオピニオンダイナミクスの帰結がどのように変容するかを探究することを目的とする。特に、セクション 2.4.1 で述べられたようにアンケートの結果に基づいて、知人もしくはロボットといった他者から意見の提供を受けた人がその人自身の意見を他者の意見に合わせた意見表明を実施するかどうかの確率をカーネル密度推定によって算出する。その確率関数に基づいて人エージェントが自身の意見を更新する場合のシミュレーションが実施される。また、本研究では最終ステップにおいて社会の中の多数のエージェントが共有して保持している意見値を多数派意見と定義し、その多数派意見が人社会と人ロボット社会でのシミュレーションの帰結において、どの程度異なるのかを調査する。

人社会と人-ロボット社会における多数派意見の変化量を定量化するために、シミュレーションの各試行 (seed) ごとに、意見分布の最頻値の差の絶対値を指標として構成した。具体的には、まず各条件 $\{\epsilon, \text{SBWS}\}$ および各試行に対して、人社会 (ロボット数 $n_R=0$) で得られた意見分布の最頻値を argmaxKDE_H として取得する。次に、同一の条件と試行のままロボット数 $n_R > 0$ の人-ロボット社会で得られた最頻値を argmaxKDE_{HR} として取得し、両者の差分の絶対値 $|\text{argmaxKDE}_{HR} - \text{argmaxKDE}_H|$ を算出する。したがって、その指標は「ロボット導入によって、多数派意見がどれだけ移動したか」を表す。値が 0 に近いほど、多数派意見の位置がロボット導入前後でほとんど変化していないことを意味し、値が大きいほど、多数派意見がより大きくシフトしたことを意味する。

この指標を算出する意義は、ロボット導入の影響を社会全体の多数派意見の変化として直接解釈可能な形で捉えられる点にある。社会の中で共有される多数派意見が社会規範としてみなされる場合、その社会規範

の変容とは技術に関する倫理の文脈で言及される将来的に考慮されるべき現象である [17]. 平均値や個人レベルの意見更新量は, 少数の極端な意見や分散の増減の影響を受けやすい一方で, 上記指標は分布の山 (多数派が集中する領域) の移動に着目するため, 「社会的に支配的な意見がどの程度入れ替わったか」という観点から複数のロボットの社会全体への介入効果を評価できる. とりわけ, 本研究の関心が多数派の形成や変容にあることを踏まえると, 多数派の位置の変化を単一のスカラー量として比較可能にする当該指標は, ロボット数や更新確率の参照方法の違いが, 多数派意見の変動を抑制もしくは促進するかを明らかにするうえで有効な要約指標となる.

3 結果

3.1 Web 調査結果

図 3 は参加者自身の意見と架空の対話相手の意見の間の距離 d に条件づけた意見の更新確率 $p^{\text{upd}}(d)$ を示す. 横軸が距離を示し, 縦軸は各距離において参加者が更新を行う確率である. 加えて, 凡例にある通り, 架空の対話相手が知人であれば Human, 対話ロボットであれば Robot が付与されている. 各曲線は, Web 調査データに基づき KDE で平滑化した推定値であり, 距離に応じて更新確率がどのように変化するかを連続的に表す. なお, 距離が大きい領域では回答数が少なく推定が不安定になり得るため, 図 3 ではでは可視化範囲を $d < 55$ に制限している.

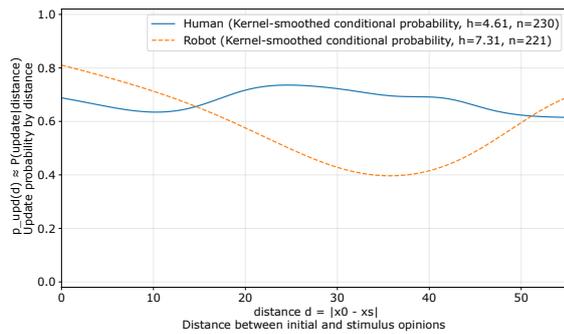


図 3: Web 調査参加者自身の意見と架空の対話相手の意見の間の距離 d に条件づけた意見の更新確率 $p^{\text{upd}}(d)$

3.2 シミュレーション結果

図 4, 5 はそれぞれ, ϵ 条件ごとの人社会と人ロボット社会の意見分布の最頻値の差の絶対値をグラフ化している. 人社会 $n_R = 0$ に対する人-ロボット社会

$n_R > 0$ での多数派意見の移動量を, $|\text{argmaxKDE}_{HR} - \text{argmaxKDE}_H|$ として可視化している. 横軸は人-ロボット社会に投入したロボット数 n_R ($n_R \neq 0$ 条件) を, 縦軸はロボット導入前後で最終ステップでの多数派意見がどれだけ移動したかを表す. 各箱ひげ図は試行間のばらつきを示している. あわせて, 各点は個々の試行における $|\text{argmaxKDE}_{HR} - \text{argmaxKDE}_H|$ の値を表している. さらに, 色分けにより Web 調査に基づく更新確率分布を逐次参照する条件 (True) と, 参照しない条件 (False) を比較している. したがって, 図 4, 5 はロボット数の増加および更新確率参照の有無が, 多数派意見の変容の大きさに与える影響を, ϵ 条件ごとに概観するためのものである.

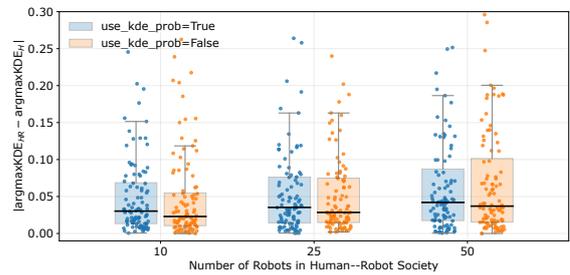


図 4: $\epsilon = 0.25$ 条件における人社会と人ロボット社会での多数派意見の差の絶対値.

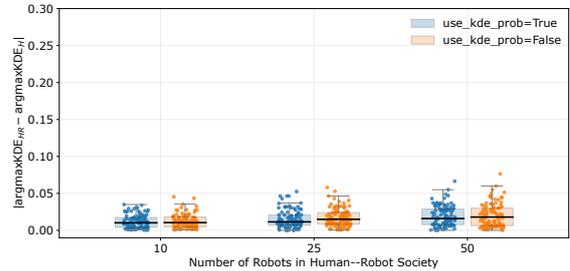


図 5: $\epsilon = 0.50$ 条件における人社会と人ロボット社会での多数派意見の差の絶対値.

4 考察

図 3 では, 対話相手が人であった場合の更新確率は一定して 0.6 を超えている. 興味深いことに, 対話相手がロボットであった時, 更新確率は距離が極端に近い場合は人よりも更新確率が高いものの, 距離が 15 以上となる時には更新確率が減少する. このような全体傾向は, 参加者に対してエージェントが人に与える社会的影響は人が人に与える社会的影響ほど大きいものではないという先行研究の実験結果とも一致する. したがって, 同じ距離であっても誰の意見として提示さ

れたかによって更新確率の水準や低下の仕方が異なり得ることが示唆される。

図4, 5は, ロボットの数という要因よりも人々が他者の意見との違いに寛容であるか不寛容であるかという要因が評価指標により大きく影響を与えることが分かった。意見値の幅は1であることを考慮すると, 0.25の変動は比較的大きい。また, Web調査の結果は図3が示す通り, 人に対する同調しやすさ, ロボットに対する同調しやすさには傾向の違いがあった。しかしながらその違いは, 社会全体の帰結として形成される多数派意見の変容という評価指標に影響を与えなかった(図4, 5の凡例の $use_{kdep_{rob}} = \text{True/False}$ 参照)。このことは, ロボットへの同調の可能性が低いという人とロボットの相互作用というミクロの状況に関する実証データはシミュレーションが導く社会全体での帰結に影響を与えないことが示唆された。よって, 実証データに基づかない人ロボット社会のシミュレーションに関する先行研究[4]と同様, ロボットというよりは人の特性が, 人ロボット共存社会で共有される多数派意見や社会規範の変容に影響を与えることが示された。

5 おわりに

本研究では周囲の多数派意見に同調してその周囲への適応を試みるロボットエージェントが社会に多数参入し, 人々と相互作用することによって引き起こされる影響を, オピニオンダイナミクスに関するシミュレーションによって評価した。特に, 人とエージェントの相互作用の架空の状況に関するWeb調査の結果によって, シミュレーションの一部パラメータを固定した上で, 社会シミュレーションの帰結の解釈を試みた。シミュレーション結果から, 調査された人とロボットが持つ社会的影響は仮想社会における多数派意見形成に比較的影響しないことが明らかになった。ロボットの社会的影響力や振舞い方というよりは, 人々が他者の持つ異なる意見にどの程度寛容かどうかが, その社会における多数派意見の形成に影響することが示された。将来的には, 異なる状況におけるWeb調査の実施や, 実際のHAI実験の結果に基づいた社会シミュレーションを実施していくことが考えられる。

参考文献

- [1] Leite, I. et al.: Social robots for long-term interaction: A survey, *International Journal of Social Robotics*, Vol. 5. No. 2, pp. 291–308 (2013)
- [2] Sheridan, T. B.: A review of recent research in social robotics, *Current Opinion in Psychology*, Vol. 36, pp. 7–12 (2020).
- [3] Fuse, Y. and Tokumaru, M.: Social influence of group norms developed by human-robot groups, *IEEE Access*, Vol. 8, pp. 56081–56091 (2020).
- [4] 布施陽太郎:同調するロボットが参加するスモールワールドネットワークの意見伝播に関する社会シミュレーション, *HAIシンポジウム 2025*, P2-23, 2025.
- [5] Nass, C., Steuer, J., and Tauber, E. R.: Computers are social actors, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 72–78 (1994).
- [6] Reeves, B. and Nass, C.: The media equation: how people treat computers, television, and new media like real people and places, Cambridge University Press, USA (1996).
- [7] Thellman, S., de Graaf, M., and Ziemke, T.: Mental state attribution to robots: A systematic review of conceptions, methods, and findings, *ACM Transactions on Human-Robot Interaction (THRI)*, Vol. 11, No. 4, Article 41, pp. 1–51 (2022).
- [8] Zhang, C., Chen, J., Li, J., Peng, Y., and Mao, Z.: Large language models for human-robot interaction: A review, *Biomimetic Intelligence and Robotics*, Vol. 3, No. 4, Article 100131 (2023).
- [9] Spears, R.: Social influence and group identity, *Annual Review of Psychology*, Vol. 72, pp. 367–390 (2021).
- [10] Powers, A. and Kiesler, S.: The advisor robot: Tracing people’s mental model from a robot’s physical attributes, *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 218–225 (2006).
- [11] Ham, J. and Midden, C. J. H.: A persuasive robot to stimulate energy conservation: The influence of positive and negative social feedback and task similarity on energy-consumption behavior, *International Journal of Social Robotics*, Vol. 6, No. 2, pp. 163–171 (2014).
- [12] Chidambaram, V., Chiang, Y.-H., and Mutlu, B.: Designing persuasive robots: How robots might

- persuade people using vocal and nonverbal cues, *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, pp. 293–300 (2012).
- [13] Salomons, N., van der Linden, M., Strohkorb Sebo, S., and Scassellati, B.: Humans conform to robots: Disambiguating trust, truth, and conformity, *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 187–195 (2018).
- [14] Salomons, N. and Scassellati, B.: Trust and conformity when interacting with a group of robots, *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 315–316 (2018).
- [15] Qin, X., Chen, C., Yam, K. C., Cao, L., Li, W., Guan, J., Zhao, P., Dong, X., and Lin, Y.: Adults still can't resist: A social robot can induce normative conformity, *Computers in Human Behavior*, Vol. 127, Article 107041 (2022).
- [16] Driver, T. B., Simecek, M., Clarke, C., You, Y., and Fraune, M. R.: Who is a better salesperson? People conform more to a human than robot facilitator in the wild, *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 427–431 (2024).
- [17] Coggins, T. N. and Steinert, S.: The seven troubles with norm-compliant robots, *Ethics and Information Technology*, Vol. 25, No. 2, Article 29 (2023).
- [18] Sunstein, C. R.: *Conformity*, NYU Press (2019).
- [19] Watts, D. J. and Strogatz, S. H.: Collective dynamics of 'small-world' networks, *Nature*, Vol. 393, No. 6684, pp. 440–442 (1998).
- [20] Watts, D. J., *Six Degrees: The Science Of A Connected Age*, W. W. Norton & Company (2004). (ダンカン・ワッツ, 辻竜平 (訳), 友知政樹 (訳): スモールワールド・ネットワーク [増補改訂版]: 世界をつなぐ「6次」の科学, 筑摩書房, (2016).)
- [21] Hegselmann, R. and Krause, U.: Opinion dynamics and bounded confidence: Models, analysis and simulation, *Journal of Artificial Societies and Social Simulation*, Vol. 5, No. 3 (2002).
- [22] Deffuant, G. et al.: Mixing beliefs among interacting agents, *Advances in Complex Systems*, Vol. 3, Issue 01n04, pp. 87–98 (2000)