

推薦論文

バランス理論を用いた社会的エージェントの分析

中澤 諭[†] 中西 英之^{†,††}
石田 亨^{†,††} 高梨 克也^{†††}

社会的エージェントが人間関係に与える影響を調べる第1歩として、エージェントがバランス理論に従って人間関係に影響を与えうるかを調べる実験を行った。エージェントが人間の意見に対して賛成や反対を表明することで、人間がエージェントに対して好意や敵意を抱くかを確かめた。次に、人間エージェント間の関係と人間関係の間にバランス理論が働くかを実験した。実験結果から、人間同士が会話をしない状況ではエージェントが人間関係に影響を与えることが分かった。しかし、人間同士で少しでも会話をすると、エージェントは人間関係に影響を与えられなくなった。会話分析の結果から、人間同士の会話による同調がエージェントの賛成や反対の効果を失わせることが分かった。

Analysis of Social Agents Using Balance Theory

SATOSHI NAKAZAWA,[†] HIDEYUKI NAKANISHI,^{†,††} TORU ISHIDA^{†,††}
and KATSUYA TAKANASHI^{†††}

We conducted an experiment to determine whether agents could influence human relations along the lines of the balance theory. We confirmed that our agent could win favorable feelings from subjects by showing an agreeing attitude to them and, conversely, unfavorable feelings by showing a disagreeing attitude. Next, we validated the balance theory as a rule governing both agent-human relations and human relations. In our experiment, we found that the agent could influence subjects' relations providing that they did not have a conversation. As a result of conversation analysis, we conclude that the subjects' tendency to sympathize with others through conversation eliminated the influence of the agent's agreeing or disagreeing attitude.

1. はじめに

仮想世界の中で人間とコミュニケーションを行うソフトウェアエージェントが現れてきている。テキストチャットの世界で会話も含めたサービスを提供するエージェント^{1),2)}は、オンラインコミュニティの形成に大きな役割を果たす仮想世界³⁾に必要な住人となっている。我々はこのようなエージェントを社会的エージェントと呼んでいる。インタフェースエージェントが人間とコンピュータのインタラクションを支援するのに対して、社会的エージェントは人間同士のインタラクションを支援する。実在する都市と同じ構造の仮

想空間であるデジタルシティの建設も始まっており⁴⁾、社会的エージェントの活動の場はますます広がっている。この研究は仮想世界に住む社会的エージェントの設計に関する一連の研究の文脈に沿ったものである。

ソフトウェアエージェントが社会性を持つことを証明する研究がある。人間はエージェントの表情、視線、頭の動きに対して、人間のそれであるかのように反応する⁵⁾。また、人間はコンピュータの性格や性別の違いに対して、人間の場合と同じように反応する⁶⁾。さらに、囚人のジレンマゲームにおいて、人間は人間の顔をしたエージェントともビデオ会議の相手の人間とも同じくらい協力した⁷⁾。これらの研究は、社会的インタラクションや社会的関係といった、人間同士の間だけで見られる現象が人間とエージェントの間でも起こったことを報告している。

[†] 京都大学大学院情報学専攻

Department of Social Informatics, Kyoto University

^{††} 科学技術振興事業団 (CREST)

Japan Science and Technology Corporation (CREST)

^{†††} 通信総合研究所

Communications Research Laboratory

本論文の内容は2002年3月のインタラクション2002シンポジウムにて報告され、シンポジウムプログラム委員長により情報処理学会論文誌への掲載が推薦された論文である。

我々の研究の目的も上記の研究と類似しているが、人間とエージェントが1対1で接する場合⁸⁾ではなく、エージェントが仮想世界のような人間のコミュニティに入っていく場合に注目した点で異なっている。以前、我々は3Dビデオチャットの仮想世界で、エージェントが人間同士のコミュニケーションを支援する実験を行った。その結果、エージェントの振舞いがエージェント、会話相手、さらに会話相手の国民性に関するステレオタイプに対する印象にまで大きく影響することが分かった⁹⁾。この結果は、エージェントがコミュニティで果たす役割に大きな期待を抱かせる一方、エージェントがコミュニティを制御する能力をどこまで持つてしまうのか不安を感じさせる。社会性の基本は他者と関係を結ぶことであり¹⁰⁾、社会性を持ったエージェントがコミュニティ内の人間関係を制御してしまうかもしれない。この研究では、エージェントが人間関係やコミュニティに与える影響をより詳しく調べるために、以下の2点を明らかにすることを目的としている。

(1) エージェントが人間の意見に賛成や反対を表明することで、人間関係に影響を与えるか。

バランス理論によると、2人の人間関係はそれぞれがある対象に同じ心情を抱くかどうかの影響される¹¹⁾。この理論に従って、人間関係の鍵となるこの対象の役割をエージェントが果たせるのかを調べる。任意の2人から好意を持たれる存在や、1人には好意を持たれ、もう1人には敵意を持たれる存在に、エージェントが会話を通じてなれるのかを調べる。

(2) エージェントの影響は、人間同士のコミュニケーションチャンネルの太さの影響を受けるか。

エージェントが人間に好意や敵意を持たれたとしても、バランス理論に従って人間関係が変化するとは限らない。なぜなら、人間関係は人間同士の会話に大きく影響されるからである。そこで、人間同士のコミュニケーションチャンネルを太くする、つまり人間同士の会話の量および質を増加させることで、エージェントの影響がどのように変化するかを調べる。

2. バランス理論

バランス理論は人間関係の相互作用で成立する。そこで、人間エージェント間の関係と人間関係の間に同様の相互作用が起こるかを検証した。バランス理論は、2人の人間と1つの対象X(人間, 物, 事実)の間に成立するものである^{12)~14)}。自分がXに対して正の心情(好意, 類似性, 賛成など)、あるいは負の心情(敵意, 非類似性, 反対など)を抱いているとき、相手もXに対して同じ心情を抱いていると推測すれば、

相手に正の心情を抱く。相手のXに対する心情が自分と異なると推測すれば、相手への心情は負となる。Xが人間の場合、この理論は3者関係に成立する。バランス理論はハイダーによって提唱され、さらに様々な研究者によって発展してきた¹⁵⁾。我々は、ハイダーによって提唱されたバランス理論に基づき、それをエージェントに適用した。我々の実験におけるXはエージェントであり、被験者がエージェントに抱く心情を制御することで、2人の被験者間の心情を制御できるかを観察した。Xは物にも事実にもなりうるので、バランス理論の成立はエージェントが意見対立を生む争点や好き嫌いの分かれる食べ物と同様の影響力を持つことを示すだけである。しかし、争点や食べ物と異なり、エージェントは人間と同様に他者との会話を通じて関係を形成する。物や事物などの対象が、2人の人間に働きかけを行って対象自身に抱く心情を制御し、同じにしたり異なるものにしたりすることは困難である。しかし、エージェントならば、2人の人間がエージェント自身に抱く心情を制御できる可能性は十分にある。また、対象Xが人間の場合、X自身が2人の他者に働きかけを行ってX自身に抱く心情を制御することは可能であろう。バランス理論では、対象Xとして人間と物や事実を区別していないので、2人の人間がエージェントに対して抱く心情をエージェント自身が制御できれば、人間と同様の影響力を持つことになる。しかし、人間の場合、働きかけの方法にエージェントよりも多様な方法を用いることができる。この働きかけの多様さの違いが心情を制御する際に影響する可能性がある。

この実験の課題は人間とエージェントの関係にバランス理論が成立することを示すことである。これを示すために、我々は以下に示す3つの仮説を立てた。

H1: 被験者はエージェントに賛成されるとエージェントに対して好意を持ち、反対されるとエージェントに敵意を持つ。

H2: 相手がエージェントに賛成されると、相手がエージェントに対して好意を持っていると被験者は推測する。相手がエージェントに反対されると、相手がエージェントに対して敵意を持っていると被験者は推測する。

H3: H1, H2の成立する条件の下でバランス理論が成立する。

仮説H1を検証するために、我々はエージェントに賛成や反対の態度をとらせて、被験者がエージェントに好意や敵意を持つかを観察した。仮説H1は、被験者がエージェントに賛成されるときの方が反対される

ときよりも、エージェントに対してより正の心情を抱くかで検証する。同様に、仮説 H2は、相手がエージェントに賛成されるときの方が反対されるときよりも、エージェントに対してより正の心情を抱くと被験者が推測するかで検証する。

仮説 H1, H2 が成り立つことを前提として、両被験者に賛成の態度をとる賛成エージェント、反対の態度をとる反対エージェント、1人に賛成し、もう1人に反対する不公平エージェントの3種類の性格の異なるエージェントを用いて、バランス理論が成立するかを観察した。図1はバランス理論が成立したときの関係を示したものである。バランス理論の成立は、エージェントが両者に同じ態度(賛成か反対)をとるときの方が、不公平な態度をとるときよりも、被験者が相手に抱く心情がより正になるかで検証する。

3. 実験設定

上記の仮説を検証するために、我々が開発した3Dビデオチャット FreeWalk¹⁶⁾上で15分から20分間、初対面の相手およびエージェントとミーティングを行い、それぞれへの関係を形成する実験を行った。被験者にはミーティング前に、もう1人の被験者およびエージェントとやりとりをする、とだけ教示した。エージェ

ントはあらかじめ発言内容と動作ルールを持っており、自動的に行動する。人間同士は音声と映像による会話が可能で、人間とエージェントは吹き出しで会話する。エージェントの吹き出しの中には質問文と選択肢が表示され、尋ねられた被験者は選択肢のどれかをマウスでクリックして答える(図2参照)。3Dビデオチャットを用いた理由は、コミュニケーションチャネルの太さの違いを試すために、次のような2×2の4つの設定で実験を行うためである。各被験者がお互いのエージェントへの心情形成過程を共有するH-A-H環境と、個別にエージェントへの心情を形成した後で、相手のエージェントへの心情を知るH-A環境を試した。ここで、Hは被験者を指し、Aはエージェントを指す。また、エージェントとの会話の間、被験者同士で会話できる場合(H-H会話あり)と、会話できない場合(H-H会話なし)を試した。この4つの実験環境を表1にまとめる。

3.1 H-A-H環境

H-A-H環境では2人の被験者と犬の絵のエージェントが輪になって3人で会話した(図2参照)。ミーティングの間、エージェントは様々な事物の好き嫌いに関する事前アンケートの結果に従って、2人の被験者の答えが一致する質問と一致しない質問を10問ずつ計20問を2人の被験者に出した。被験者の答えに対するエージェントの反応を次のように変えた。2人の答えが一致した場合、賛成エージェントはその答えに賛成し、反対エージェントはその答えに反対し、不公平エージェントは中立的な反応を返す。一致しなかった場合、賛成および反対エージェントは中立的な反応を返し、不公平エージェントは片方の答えに賛成し、もう片方の答えに反対する。

H-H会話ありの場合、被験者同士の会話を20分と定め、その間にエージェントが割り込んだ。そのため、被験者が質問に答えるのが遅い場合、20問すべて質

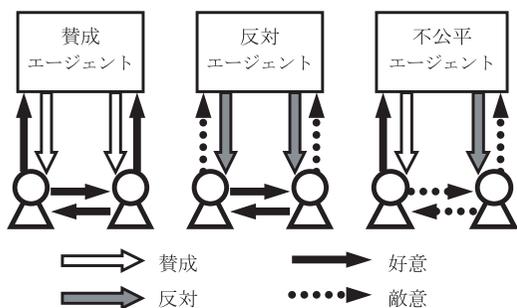


図1 各エージェントと被験者間のバランス理論

Fig. 1 Balance theory with agent and two humans.

表1 4つの実験環境

Table 1 Four environments in the experiment.

	H-H 会話あり		H-H 会話なし
H-A-H 環境	H-A-H 環境 (H-H 会話あり): 様式: 2人の被験者とエージェントが輪になって3人で会話する。 質問: 被験者の答えが一致する質問と一致しない質問を10問ずつ計20問出す。 会話: エージェントとのやりとりに並行して、被験者同士が会話する。 相手の心情の推測: エージェントに対する相手の答えに基づく。		H-A-H 環境 (H-H 会話なし): 被験者同士で会話できないこと以外は左記と同様。
H-A 環境	H-A 環境 (H-H 会話あり): 様式: エージェントがそれぞれの被験者と1対1で会話する。 質問: 12問の社会問題に関する質問を出す。 会話: 1分間の休憩を2回挟み、その間だけ被験者同士が会話する。 相手の心情の推測: 後から現れる別のエージェントへの相手の答えに基づく。		H-A 環境 (H-H 会話なし): 被験者同士で会話できないこと以外は左記と同様。

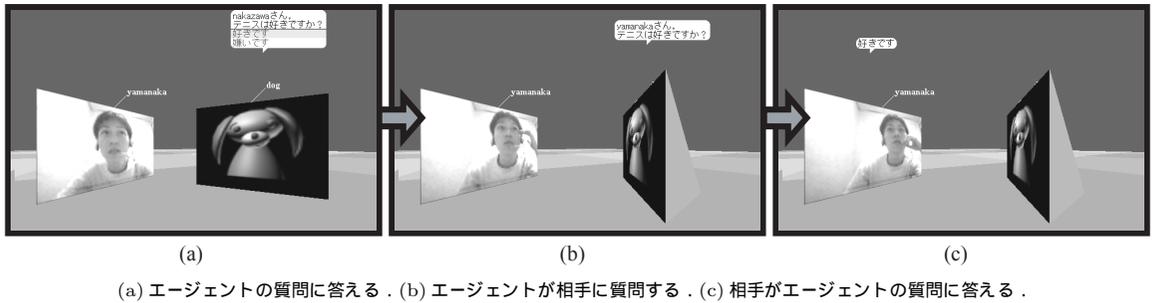


図2 H-A-H 環境

Fig.2 H-A-H environments.

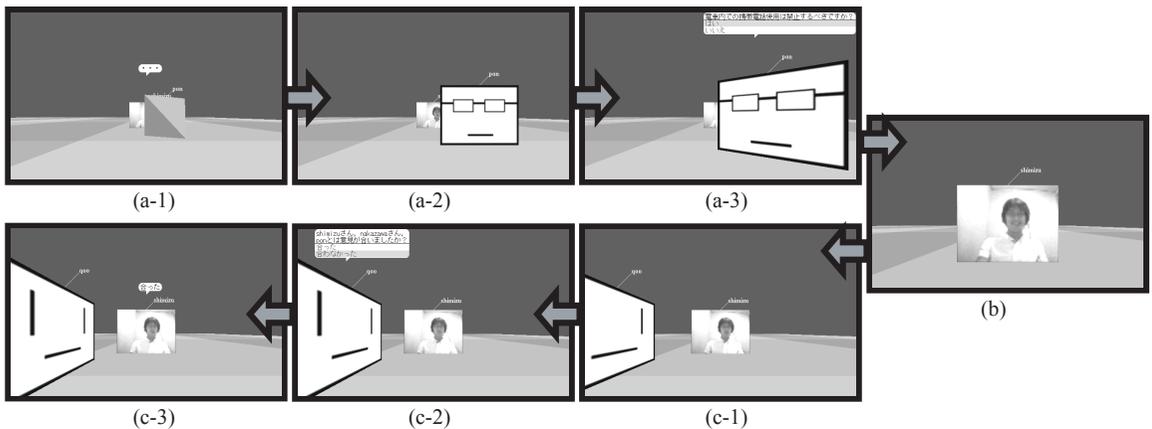


図3 H-A 環境

Fig.3 H-A environments.

問できないこともあった . H-H 会話なしの場合は , 被験者はエージェントとの質疑応答のみを行った .

3.2 H-A 環境

H-A 環境では , 仮想空間の中で離れて立っている 2 人の被験者の間をエージェントが往復して , 各被験者と 1 対 1 で会話した . このとき , 相手とエージェントとの会話内容は分からないようにした (図 3 (a-1) , (a-2) , (a-3) 参照) . エージェントは片方の被験者に 4 問質問し , 次にもう片方に 4 問質問するフェーズを 3 回繰り返し , 合計 12 問の社会問題に関する質問を各被験者に出した . 賛成エージェントは両方の被験者の答えに対し , 10 問賛成して 2 問だけ反対した . 他のエージェントも , 同様に 10 問は自分の性格に従う態度をとった . 2 問例外を混ぜたのは , その方がより態度が明確に伝わるのが予備実験で明らかになったからである . 質問し終わったエージェントが画面から消えた後 , 別の顔をしたエージェント No.2 が現れ , 2 人の被験者にエージェントへの好意に関する質問を 8

問出した (図 3 (c-1) , (c-2) , (c-3) 参照) . エージェント No.2 との会話は 3 人で行われ , このときに , 各被験者は相手がエージェントに抱く心情を知ることができた .

H-H 会話ありの場合 , エージェントの各フェーズの間で 2 回 1 分間の休憩を挟み , その間だけ被験者同士が近づいて会話した (図 3 (b)) . このとき , エージェントへの心情が相手に露呈されてしまわないように , 実験に関することは話さないよう教示した . H-H 会話なしの場合 , 休憩を挟まずに , 被験者はエージェントとの質疑応答のみを行った . H-A 環境では , 2 人のエージェントが登場するため , 区別しやすい線画の 2 つの絵を用いた .

3.3 被験者数とアンケート内容

被験者は大学生である . 賛成エージェントが登場する条件を賛成条件 , 反対エージェントが登場する条件を反対条件 , 不公平エージェントに賛成される条件を不公平賛成条件 , 不公平エージェントに反対される条

件を不公平反対条件とする。不公平賛成条件の場合、被験者は不公平エージェントに賛成されるので、相手はエージェントに反対される。同様に、不公平反対条件の場合、被験者は不公平エージェントに反対されるので、相手はエージェントに賛成される。賛成条件、反対条件、不公平賛成条件、不公平反対条件の被験者数はそれぞれ、H-A-H環境では12, 10, 11, 11, H-A環境(H-H 会話なし)では14, 12, 12, 12, H-A環境(H-H 会話あり)では12, 12, 12, 12である。合計186名の被験者が実験に参加した。各環境の被験者数が条件ごとで異なっていたので、スクリーニングを行ってサンプル数をすべて10に揃えた後分析した。

ミーティング後のアンケートでは、類似と好意に関する質問を3問ずつ、被験者からみた相手、被験者からみたエージェント、相手からみたエージェントについて、の3方向について質問した。H-A-H環境における質問項目は、類似については「似通っている」、「関心が似ている」、「同じグループに感じる」、の3つであり、好意については「好ましい人物に近い」、「好意を抱く」、「会話を続けたい」、の3つである。H-A環境における質問項目は、類似については「意見が合う」、「共感できる」、「似ている」、の3つであり、好意については「好感が持てる」、「魅力的に感じる」、「好き」、の3つである。

4. 統計的分析

4条件のアンケート結果を1要因分散分析で比較した結果、有意差の表れた項目を表にまとめた。a, b, c, d, の欄はそれぞれ、賛成条件、反対条件、不公平賛成条件、不公平反対条件におけるアンケート結果の平均値である。1が最もあてはまらない場合で、8が最もあてはまる場合である。F値の欄において、5%有意には*、1%有意には**、0.1%有意には***をそれぞれ付加した。全表において、自由度は(3, 36)である。その横の欄はLSDによる多重比較の結果である。少なくとも5%で有意差の表れた大小関係を示した。

バランス理論の成立は次のように検証する。仮説H1が成立していることは、エージェントについての質問で有意差が表れたかで確認できる。仮説H1によれば、エージェントに賛成される条件(賛成条件(a)、不公平賛成条件(c))の方が、反対される条件(反対条件(b)、不公平反対条件(d))よりもエージェントに対してより好意を持つはずである。よって、多重比較の結果は、a>b, a>d, b<c, c>dとなるはずである。仮説H2は、相手からみたエージェントについての質問で有意差が表れたかで確認できる。仮説H2によれば、

表2 H-A環境(H-H 会話なし)において、エージェントの違いを1要因分散分析で比較した結果

Table 2 Summary of ANOVA in H-A environment (H-H non-conversation).

項目	a	b	c	d	F	多重比較
エージェントについて						
意見が合う	4.8	2.3	5.8	2.3	36.3***	a>b a<c a>d b<c c>d
共感できる	4.3	2.1	4.9	3.4	11.1***	a>b b<c b<d c>d
似ている	4.0	1.9	4.6	2.4	8.2***	a>b a>d b<c c>d
好感	5.3	2.5	4.4	3.0	14.5***	a>b a>d b<c c>d
魅力的	5.1	3.7	4.4	3.5	4.3*	a>b a>d
好き	5.3	3.0	4.3	3.4	14.2***	a>b a>c a>d b<c c>d
相手からみたエージェントについて						
意見が合う	5.0	2.5	2.3	5.7	17.7***	a>b a>c b<d c<d
共感できる	4.1	2.4	2.4	5.5	13.4***	a>b a>c a<d b<d c<d
似ている	3.7	1.7	2.3	4.4	11.0***	a>b a>c b<d c<d
好感	5.2	2.7	2.6	5.6	23.8***	a>b a>c b<d c<d
魅力的	4.5	2.9	2.9	4.6	7.3**	a>b a>c b<d c<d
好き	5.3	3.1	3.2	4.8	9.7***	a>b a>c b<d c<d
相手について						
意見が合う	4.7	5.9	3.2	4.2	12.5***	a<b a>c b>c b>d c<d
共感できる	4.7	6.0	3.3	4.3	10.2***	a<b a>c b>c b>d
似ている	3.3	4.9	2.5	3.3	7.7***	a<b b>c b>d
好感	5.6	5.6	4.5	4.5	4.1*	a>c a>d b>c b>d

*p < .05, **p < .01, ***p < .001 (df=3,36)

a: 賛成, b: 反対, c: 不公平賛成(被験者に賛成, 相手に反対), d: 不公平反対(被験者に反対, 相手に賛成)

相手がエージェントに賛成される条件(賛成条件(a)、不公平反対条件(d))の方が、反対される条件(反対条件(b)、不公平賛成条件(c))よりも、相手がエージェントに対してより好意を持っていると推測するはずである。よって、多重比較の結果は、a>b, a>c, b<d, c<dとなるはずである。仮説H3は、相手についての質問で有意差が表れたかで確認できる。仮説H3によれば、エージェントが両者に同じ態度をとる条件(賛成条件(a)、反対条件(b))の方が、不公平な態度をとる条件(不公平賛成条件(c)、不公平反対条件(d))よりも、相手に対してより好意を持つはずである。よって、多重比較の結果は、a>c, a>d, b>c, b>dとなるはずである。多重比較の欄において、仮説に従ったものを太字で示した。厳密には仮説に従っていないものもあるが、仮説に反するものはなかった。4環境の結果をバランス理論により従った順に述べる。

4.1 H-A環境(H-H 会話なし)

バランス理論の成立がほぼみられた(表2参照)。まず、エージェントは被験者がエージェントに抱く心情を制御できた。エージェントおよび相手からみたエージェントについての質問で、それぞれ6項目すべてで

表3 H-A-H環境(H-H会話なし)において、エージェントの違いを1要因分散分析で比較した結果

Table 3 Summary of ANOVA in H-A-H environment (H-H non-conversation).

項目	a	b	c	d	F	多重比較
エージェントについて						
似通っている	3.2	2.4	5.3	1.7	9.1***	a<c a>d b<c c>d
関心が似ている	3.8	1.4	6.0	1.7	35.6***	a>b a<c a>d b<c c>d
同じグループ	3.5	2.1	4.3	1.6	8.2***	a>b a>d b<c c>d
相手からみたエージェントについて						
似通っている	3.4	2.0	1.5	6.4	36.9***	a>b a>c a<d b<d c<d
関心が似ている	3.7	1.7	1.6	6.7	48.1***	a>b a>c a<d b<d c<d
同じグループ	3.0	1.9	1.5	5.2	18.7***	a>c a<d b<d c<d
好ましい人物	3.7	2.6	2.1	3.7	3.6*	a>c c<d
好意	3.1	3.1	2.0	4.7	6.2**	a<d b<d c<d
相手について						
似通っている	4.0	4.3	2.7	4.1	3.1*	a>c b>c c<d
同じグループ	3.7	3.8	2.2	3.3	4.3*	a>c b>c c<d
好意	4.9	5.5	3.7	4.9	4.8**	a>c b>c c<d

有意差がみられ、H1とH2は受理された。次に、エージェントは人間関係に影響を与えることができた。相手についての質問では4項目で有意差がみられ、特に好意に関する質問「好感が持てる」でH3が受理された。残りの3つの類似に関する質問では、賛成条件(a)と不公平反対条件(d)の間にa>dの有意な関係がみられず、さらに、1つの質問項目では、賛成条件(a)と不公平賛成条件(c)の間にa>cの有意な関係もみられなかった。これらの点を除き、H3に従う有意な関係がみられた。H3はほぼ受理された。

相手についての類似に関する3つの質問すべてでa>dがみられなかった理由は次のように考えられる。まず、3質問すべてにおいて、正の心情を抱くと予想された賛成条件(a)と反対条件(b)の間に、a<bの有意な関係がみられている。これは、同じ対象にともに好意を抱く人よりも、ともに敵意を抱く人により類似性を感じることを示している。次に、1つの質問において、負の心情を抱くと予想された不公平賛成条件(c)と不公平反対条件(d)の間に、c<dの有意な関係がみられている。これは、自分が敵意を抱く対象に好意を抱く人よりも、自分が好意を抱く対象に敵意を抱く人により非類似性を感じることを示している。これら2点より、賛成条件(a)では、正の心情をあまり抱かなかった一方で、不公平反対条件(d)では、負の心情をあまり抱かなかったため、a>dがみられなかった

表4 H-A環境(H-H会話あり)において、エージェントの違いを1要因分散分析で比較した結果

Table 4 Summary of ANOVA in H-A environment (H-H conversation).

項目	a	b	c	d	F	多重比較
エージェントについて						
意見が合う	5.5	2.3	5.9	2.3	50.6***	a>b a>d b<c c>d
共感できる	5.2	2.9	5.4	2.6	15.4***	a>b a>d b<c c>d
似ている	4.3	3.0	5.2	2.3	9.5***	a>b a>d b<c c>d
相手からみたエージェントについて						
意見が合う	5.6	2.3	2.7	6.8	61.7***	a>b a>c a<d b<d c<d
共感できる	5.4	3.1	3.1	6.3	17.0***	a>b a>c b<d c<d
似ている	4.2	2.4	2.5	4.9	11.8***	a>b a>c b<d c<d
好感	5.4	3.6	4.3	5.4	3.6*	a>b b<d
相手について						
意見が合う	4.9	6.0	5.2	4.3	6.3**	a<b b>d c>d
共感できる	5.0	5.8	4.9	4.3	3.7*	b>d

と考えられる。

4.2 H-A-H環境(H-H会話なし)

バランス理論が部分的に成立した(表3参照)。まず、エージェントは被験者がエージェントに抱く心情をほぼ制御できた。相手からみたエージェントについての質問で、5項目の有意差がみられ、H2は受理された。エージェントについての質問で、類似に関する3質問で有意差がみられ、H1はほぼ受理された。次に、エージェントは人間関係に部分的に影響を与えることができた。相手についての質問で3項目の有意差がみられたが、いずれも賛成条件(a)および反対条件(b)と不公平反対条件(d)の間に、a>dおよびb>dの有意な関係がみられなかった。よって、バランス理論は不公平反対条件(d)では成立せず、残りの条件では成立した。H3は部分的に受理された。この理由は、H-A環境(H-H会話なし)と同様に、負の心情を抱くと予想された不公平賛成条件(c)と不公平反対条件(d)の間にc<dの有意な関係がみられていることから、不公平反対条件(d)では、ほとんど負の心情を抱かなかったためであると考えられる。

4.3 H-A環境(H-H会話あり)

バランス理論は成立しなかった(表4参照)。まず、エージェントは被験者がエージェントに抱く心情をほぼ制御できた。相手からみたエージェントについての質問で、4項目の有意差がみられ、H2は受理された。エージェントについての質問で、類似に関する3質問で有意差がみられ、H1はほぼ受理された。次に、エージェントは人間関係に影響を与えることができなかった。相手についての質問で有意差の表れた項目が2つあるが、各条件間の有意な関係がH3と合っていない。H3は棄却された。人間同士のコミュニケーションチャ

表5 H-A-H環境(H-H会話あり)において、エージェントの違
い1要因分散分析で比較した結果

Table 5 Summary of ANOVA in H-A-H environment
(H-H conversation).

項目	a	b	c	d	F	多重比較
エージェントについて						
関心が似ている	3.7	1.4	3.8	2.1	7.8***	a>b a>d b<c c>d
相手からみたエージェントについて						
似通っている	3.0	1.8	2.0	5.3	15.5***	a>b a<d b<d c<d
関心が似ている	3.8	2.0	2.1	5.5	12.8***	a>b a>c a<d b<d c<d
同じグループ	3.0	1.7	2.2	3.6	3.9*	a>b b<d c<d
相手について、有意差の表れた項目はなかった。						

ネルが太くなると、被験者がエージェントに抱く心情は制御できても、人間関係に影響を与えることはできなくなった。

4.4 H-A-H環境(H-H会話あり)

エージェントは被験者がエージェントに抱く心情をほぼ制御できなかった(表5参照)。エージェントについての質問で有意差がみられたのが1項目であり、H1はほぼ棄却された。相手からみたエージェントについての質問で類似に関する3質問で有意差がみられたが、多重比較の結果がH2と完全に一致しているのは、質問「関心が似ている」の1項目のみである。H2は部分的に受理された。エージェントは人間関係に影響を与えることができず、H3は棄却された。人間同士のコミュニケーションチャンネルが太くなりすぎると、人間関係さらには被験者がエージェントに抱く心情すら制御できなくなった。

4.5 結論

人間同士が会話できない状況で、人間エージェント間の関係を人間同士の関係よりも先に形成する場合、エージェントは人間関係に影響を与えることができる。それらの関係を同時に形成する場合、エージェントが人間関係に及ぼす影響が弱まる。人間エージェント間の関係を先に形成している最中に、人間同士で会話を行う機会が少しでもあると、エージェントは人間関係に影響を与えられなくなる。エージェントが人間同士の会話に参加して3者関係を同時に形成する場合、人間関係はおろか被験者がエージェントに抱く心情すら制御できなくなる。

5. 会話分析

なぜ、人間同士の会話が行われると、エージェントは人間関係に影響を与えられなくなるのだろうか。その原因を探るために、H-A-H環境での会話分析¹⁷⁾を

行った。この環境では、人間エージェント間の会話がテキストチャンネルで、人間同士の会話が音声チャンネルで行われる。ただし、統計的実験の自動的に動作するエージェントとは異なり、会話をよりもっともらしくするために Wizard of Oz (WOZ) を利用し、実験者が密かにエージェントを操作した。WOZ エージェントは話題を誘導する、質問の前に導入的な発言をするなどの振舞いを行い、自分を知的に見せるを試みた。

まず顕著に観察される現象は、被験者同士の音声チャンネルと被験者エージェント間のテキストチャンネルが目的に応じて同時並行的に使い分けられることである。チャンネルの同時並行性を裏づける以下のような特徴が観察された。第1に、被験者はエージェントの質問に対する応答をテキストチャンネルで選択する(⇒1)と同時に、この応答内容をもう1人の被験者に対して音声チャンネルを用いて伝達している(⇒2)。テキストチャンネルにおける応答はもう1人の被験者に見えるので、こうした二重の応答は冗長であり、音声チャンネルでの被験者間の情報交換はこのチャンネルの確保のためのものであることが分かる(例1)。

(例1) 並行チャンネルによる二重の応答

X → A: Aさん、突然ですが、日本料理と
中華料理ではどちらの方が好きですか?

A → ? : 日本料理...

B → A: ほんまに突然。

⇒2 A → B: (笑) まあ、日本料理ですかね。

⇒1 A → X: 日本料理

X → A: なるほど。

次に、テキストチャンネルにおけるエージェントの質問(⇒1)と被験者の応答(⇒3)からなる隣接ペアの間に被験者間でのやりとりが割り込んでいる(⇒2)。これは、音声チャンネルの相対的な利用しやすさを表すものであり、エージェントの質問に応答することよりももう一方の被験者に向けて話しかけることの方が優先されていることが分かる(例2)。内容的にも、被験者間の会話はエージェントとのやりとりに対する言及や注釈となっていることが多い。

(例2) 被験者間チャンネルの優先

⇒1 X → A: Aさん、かなり突然ですが、

音楽はよく聞きますか?

⇒2 A → B: かなり突然だよ(笑)

⇒2 B → A: かなり突然ですね。犬のくせに

転記の行形式: 発話者 → 受け手: 発話

Xはエージェント、A、Bは被験者のイニシャルを表す。

テキストチャンネルでの発話は太字で示す。

気を使っていますね(笑)

会話の流れとかに気を使っていますね。

⇒2 A → B: 賢いよ。ある意味(笑)

A → B: …ま、よく聞きますね。

⇒3 A → X: はい

X → A: なるほど

つまり、被験者間の音声チャンネルは、話題上関連しているものの、被験者エージェント間のやりとりに対してメタレベルに位置づけられるような、並行する別のコミュニケーションチャンネルとして機能している。同時に、このチャンネルでは被験者がエージェントの振舞いに関する評価を話し合い、合意に至っていることが多い。これらの特徴の背景には、音声チャンネルによる被験者同士の会話がエージェントには理解されていないという被験者間の想定がある。

それでは、並行するコミュニケーションチャンネルの発生を抑えれば、エージェントの影響を維持できるだろうか。それを確認するために、テキストチャットを用いた実験を行った。この実験でも、WOZ エージェントを用いた。たとえば、不公平エージェントの場合、WOZ エージェントは、1) ターンをとる、2) 質問をする、3) 意見の違いを確認する、4) ターンを渡す、というプロセスを基本単位とし、これを会話中に4回行う。意見の違いがみられなければ、2), 3) を繰り返す、基本単位ごとに被験者間の意見の違いを確認するようにした。

このようにすると、並行するチャンネルの発生は大幅に減少する。しかし、並行するチャンネルの発生を抑えることでは解決されない多くの課題が残されている。たとえば、次の(例3)にはエージェントの会話能力の限界が端的に表れている。エージェントの反対に対して、Bはその理由を尋ねている(⇒1)。しかし、エージェントはこれに答えず退出しており(⇒2)、被験者はこの振舞いを非難している(⇒3)。

(例3) 退出

<X> Bさん、私はあまり中南米(ブラジル・ペルー)の方に行きたくないのです。私は北アメリカ(カナダ・合衆国)の方に行きたいですね。

⇒1 どうして?

⇒2 <X> それでは、またしばらくご自由にお話し下さい。

 なんだったんだ…

 Aさんは外国旅行とかは?

⇒3 <A> いきなり割り込んで…

⇒3 ねえ。

さらに顕著な問題として、エージェントによる反対(⇒1)が被験者の反発を招き(⇒2)、その結果、被験者間で同調が起こる場合がある(例4)。

(例4) 被験者間の同調

⇒1 <X> ユニバーサルスタジオジャパンの方に行きたい

Aさんとは気が合いませんね…

<A> そうねー。

⇒2 てか、コンピューターさんちょっと失礼ね。

<A> このままずっと合わないんじゃない…

会話分析の結果をまとめると次のようになる。

まず、FreeWalkでは被験者間にエージェントには理解されていないと想定される音声チャンネルでのコミュニケーションが同時並行的に開かれる。このチャンネルではエージェントの振舞いに対する評価が話し合われることが多い。したがって、被験者間の音声チャンネルが追加されると、被験者間の好意的関係は強化され、エージェントの人間関係への影響は打ち消される結果となる。

他方、音声チャンネルを含まないテキストチャットではFreeWalkのような会話の分裂は比較的少ない。しかし、エージェントの会話能力の限界による不適切な振舞いに対する被験者の反発から、被験者間に同調が起こる場合がある。そのため、エージェントが目指していた人間関係の操作にとっては逆効果となる。

このように、被験者間で会話を行う場合、被験者間には同調への力が作用している。そのため、エージェントの社会性を欠いた振舞いは被験者間の同調を引き起こす契機となる。統計的分析の結果を考慮すると、この作用は被験者間のチャンネルが太くなるほど強まると考えられる。

6. おわりに

この研究は、近未来に生まれるであろう、エージェントと人間が共存するコミュニティで、エージェントが人間関係にどのような影響を与えるかを調べたものである。その第1歩として、エージェントがバランス理論に従って人間関係に影響を与えるかを調べる実験を行った。エージェントが人間の意見に賛成または反対を表明することで、エージェントに対して好意または敵意を抱かせることができるか、人間エージェント間の関係と人間関係の間にバランス理論が働くかを実験した。統計的分析および会話分析の結果、以下の知見を得た。

(1) 人間同士が会話しない状況では、エージェント

が人間関係に影響を与えることが分かった。エージェントと人間との会話内容を、もう1人の人間が知ることができない状況では、より強固に影響を与えることができた。

- (2) 人間同士で少しでも会話をすると、エージェントは人間関係にほとんど影響を与えることができなくなった。人間同士の会話による同調がエージェントの賛成あるいは反対の効果を失わせた。

この実験結果は、エージェントが人間関係に与える能力と限界を示唆している。人間同士に十分なコミュニケーションがあれば、エージェントをコミュニティの中で安全に利用することができることを示すとともに、人間同士に十分なコミュニケーションがなければ、エージェントは危険な存在にもなりうることを示している。そのため、エージェントの会話能力が現在の技術よりも飛躍的に高まった際には、エージェント開発者と利用者のための倫理規定が必要となるだろう。我々の研究は、エージェントの開発にとって、設計や実装の問題と並んで倫理的な問題が重要になることを示している。

我々は人間とエージェントの関係にバランス理論を適用した。我々はこの研究が、人間社会に成立する理論をエージェントを含む社会に適用しようと試みる他の研究の参考になると考えている。

謝辞 実験の運営に携わってくださいました(株)インターグループの大原裕子氏に感謝いたします。本研究は科学技術振興事業団CREST「デジタルシティのユニバーサルデザイン」と通信総合研究所の共同研究として行われました。

参 考 文 献

- 1) Isbell, C., Kearns, M., Kormann, D., Singh, S. and Stone, P.: Cobot in LambdaMoo: A Social Statistics Agent, *AAAI-2000*, pp.36-41 (2000).
- 2) Foner, L.: Entertaining Agents: A Sociological Case Study, *AGENTS-97*, pp.122-129 (1997).
- 3) Kim, A.J.: *Community Building on the Web*, Peachpit Press (2000).
- 4) Ishida, T. and Isbister, K. (Eds.): *Digital Cities: Experiences, Technologies and Future Perspectives*, Lecture Notes in Computer Science 1765, Springer-Verlag (2000).
- 5) Takeuchi, A. and Naito, T.: Situated Facial Displays: Towards Social Interaction, *CHI-95*, pp.450-455 (1995).
- 6) Reeves, B. and Nass, C.: *The Media Equation*, Cambridge University Press (1996).
- 7) Parise, S., Kiesler, S., Sproull, L. and Waters, K.: My Partner is a Real Dog: Cooperation with Social Agents, *CSCW-96*, pp.399-408 (1996).
- 8) Bickmore, T. and Cassell, J.: Relational Agents: A Model and Implementation of Building User Trust, *CHI-2001*, pp.396-403 (2001).
- 9) Isbister, K., Nakanishi, H., Ishida, T. and Nass, C.: Helper Agent: Designing an Assistant for Human-Human Interaction in a Virtual Meeting Space, *CHI-2000*, pp.57-64 (2000).
- 10) Damon, W.: *Social and Personality Development: Infancy Through Adolescence*, W.W. Norton & Company (1983).
- 11) Heider, F.: *The Psychology of Interpersonal Relations*, Wiley (1958).
- 12) Jordan, N.: Behavioral Force that are a Function of Attitudes and of Cognitive Organization, *Human Relations*, Vol.6, pp.273-287 (1953).
- 13) Horowitz, W.M., Lyons, J. and Perlmutter, H.V.: Induction of Forces in Discussion Groups, *Human Relations*, Vol.4, pp.57-76 (1951).
- 14) Kogan, N. and Tagiuri, R.: Interpersonal Preference and Cognitive Organization, *Journal of Abnormal and Social Psychology*, Vol.56, pp.113-116 (1958).
- 15) Insko, C.A.: Balance Theory, The Jordan Paradigm, and The Wiest Tetrahedron, *Advances in Experimental Social Psychology*, Berkowitz, L. (Ed.), Vol.18, pp.89-140, Academic Press (1984).
- 16) Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: FreeWalk: Supporting Casual Meetings in a Network, *CSCW-96*, pp.308-314 (1996).
- 17) 高梨克也: 社会的相互行為を「見る」方法, 人工知能学会誌, Vol.16, No.6, pp.799-805 (2001).

(平成 14 年 5 月 2 日受付)

(平成 14 年 10 月 7 日採録)

推 薦 文

本論文は、インタラクシオン 2002 シンポジウムに投稿された 30 件の論文の中で、査読スコアが他論文より明らかに高かった論文である。論文の内容は以下のとおりである。

本論文はソフトウェアエージェントが人間関係に及ぼす影響を調べた社会心理学実験について述べたものである。この実験を行った結果、人間同士のコミュニケーションが十分でない場合は、人間関係がエージェントによって大きな影響を受ける可能性が判明した。また、会話を通して生じる人間同士の共感がこのよう

なエージェントの影響を打ち消すことも分かった。

社会的エージェントという人間とコンピュータの新しいインタラクション様式に着目し、ソフトウェアが人間関係に与える影響という先進的なテーマを取り扱い、それを統計的実験によって分析している。また、会話分析によって、単なる定量的評価にとどまらずに、エージェントとのインタラクションが及ぼす影響のメカニズムを明らかにしている。これらの結果やその実験手法は、今後顕在化してくるであろうソフトウェアの倫理という課題への道標にもなったと考えられる。

以上の査読者評価、および内容の重要性から、推薦論文に値するものと考えられる。

(インタラクション 2002 シンポジウムプログラム
委員長 垂水 浩幸)



中澤 諭

2000 年京都大学工学部情報学科卒業，2002 年同大学院情報学研究科社会情報学専攻修士課程修了。



中西 英之(正会員)

1996 年京都大学工学部情報工学科卒業，1998 年同大学院工学研究科情報工学専攻修士課程修了。2001 年同大学院情報学研究科社会情報学専攻博士課程修了。同年同専攻助手となり現在に至る。博士(情報学)。仮想環境，社会的エージェント，デジタルシティ等に興味を持つ。



石田 亨(正会員)

1976 年京都大学工学部情報工学科卒業，1978 年同大学院修士課程修了。同年日本電信電話公社電気通信研究所入所。米国コロンビア大学計算機科学科客員研究員，ミュンヘン工科大学客員教授，パリ第六大学招聘教授等。現在，京都大学大学院情報学研究科社会情報学専攻教授。工学博士。人工知能，コミュニケーション，社会情報システムに興味を持つ。編著書に、『分散人工知能』(コロナ社)，Real-time Search for Learning Autonomous Agents(Kluwer Academic Publishers)，Community Computing(John Wiley and Sons)，Digital Cities (Springer-Verlag) 等。



高梨 克也

2000 年京都大学人間・環境学研究科文化・地域環境学専攻博士課程単位取得退学。現在独立行政法人通信総合研究所専攻研究員。コミュニケーションの組織化を支える認知的・社会的プロセスの解明に従事。人工知能学会，認知科学会，言語処理学会，社会言語科学会，語用論学会各会員。