

# 社会的エージェントのバランス理論

中澤 諭<sup>†</sup>, 中西 英之<sup>†, ††</sup>, 石田 亨<sup>†, ††</sup>, 高梨 克也<sup>†††</sup>

<sup>†</sup> 京都大学大学院情報学研究科社会情報学専攻

<sup>††</sup> 科学技術振興事業団(CREST)

<sup>†††</sup> 通信総合研究所

社会的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た。エージェントが人間の意見に同意や反意を表明することで、エージェントに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。その結果、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。会話分析の結果、人間同士の会話による同調がエージェントの同意や反意の効果を失わせることが分かった。

## Balance Theory with Social Agents

Satoshi Nakazawa<sup>†</sup>, Hideyuki Nakanishi<sup>†, ††</sup>, Toru Ishida<sup>†, ††</sup>, Katsuya Takanashi<sup>†††</sup>

<sup>†</sup> Department of Social Informatics, Kyoto University

<sup>††</sup> Japan Science and Technology Corporation (CREST)

<sup>†††</sup> Communications Research Laboratory

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, 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もの[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きている。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。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。

社会的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オ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。

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。我々は以前、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、エージェントの振る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．この研究ではこれをさらに進め，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同意や反意を表明することで，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オブジェクトに同じ心情を抱くかどうかに影響される[11]．この理論に従って，人間関係の鍵となるこのオブジェクト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，一人には好意を持たれ，もう一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力は，人間同士のコミュニケーションチャンネルが太くなくても働くか．

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．人間がエージェントに抱いた心情が弱ければ，エージェントの影響は人間同士の会話の影響によって打ち消される．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しても，エージェントの影響が残るかを調べる．

## 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．自分がXに対して正か負の心情(好意や敵意)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．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も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

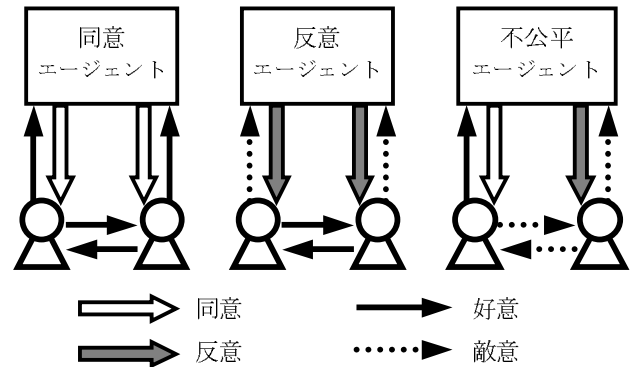


図1. 各エージェントと被験者間のバランス理論

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．2人の人間が物や事実などの対象に抱く心情を制御して，同じにしたり異なるものにしたりすることは容易ではないが，エージェントはそのような対象になる可能性がある．

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．我々は，エージェントに同意や反意の態度をとらせて，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．これは，エージェントが同意するときの方が反意するときよりも，被験者がより正の心情を抱くかで検証する．仮説を以下に示す．

**H1**： エージェントに同意されると好意を持ち，反意されると敵意を持つ．

被験者は，相手も同様にエージェントに反応すると推測する，と我々は予想した．仮説を以下に示す．

**H2**： 相手も，エージェントに同意されると好意を持ち，反意されると敵意を持つと推測する．

上記2つの仮説が成り立つことを前提に，両被験者に同意の態度をとる同意エージェント，反意の態度をとる反意エージェント，一人に同意もう一人に反意する不公平エージェントの3つの異なる性格の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．

**H3**： H1,H2の成立下でバランス理論が成立する．

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(同意か反意)をとるときの方が，不公平な態度をとるときよりも，相手に抱く心情がより正になるかで検証する．

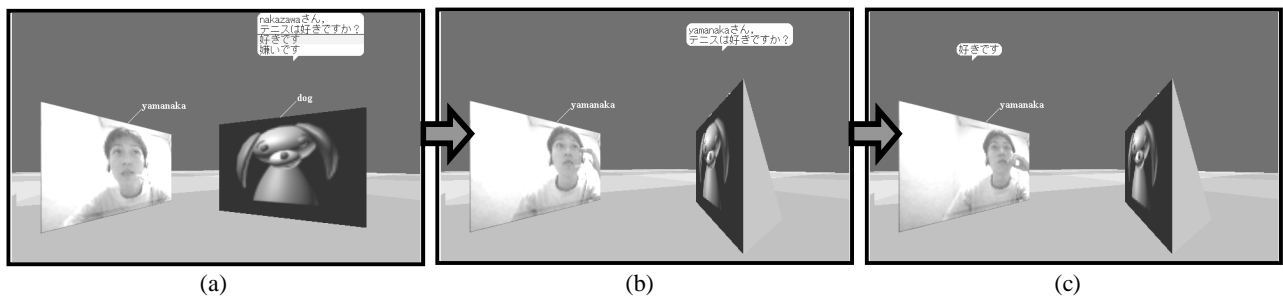


図 2. 同時環境： (a)エージェントの質問に答える. (b)エージェントが相手に質問する. (c)相手がエージェントの質問に答える.

### 3. 実験設定

上記の仮説を検証するために, 3D ビデオチャット FreeWalk[12]上で 15 分から 20 分間, 初対面の相手及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 人の被験者及びエージェントとやり取りをする, 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ン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同時環境と, 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逐次環境を試した. また, エージェントとの会話の間, 被験者同士で会話のできる会話環境と, 会話できない無会話環境を試した. この 4 つの実験環境を表 1 にま

とめる.

#### 3.1 同時環境

同時会話環境及び同時無会話環境では 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, 同意エージェントはその答えに同意し, 反意エージェントはその答えに反意し, 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, 同意及び反意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, 片方の答えに同意し, もう片方の答えに反意する. 同時会話環境では, 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く, 20 問すべて質問できないこともあった. 同時無会話環境では, 被験者はエージェントとの質疑応答のみを行った.

#### 3.2 逐次環境

逐次会話環境及び逐次無会話環境では, 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

|    | 会話  | 無会話                                    |
|----|---|--|
| 同時 | <p>同時会話環境：</p> <p>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.</p> <p>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.</p> <p>会話：エージェントとのやり取りに並行して, 被験者同士が会話する.</p> <p>相手の心情の推測：エージェントに対する相手の答えにもとづく.</p> | <p>同時無会話環境：</p> <p>会話がないこと以外は左記と同様</p> |
| 逐次 | <p>逐次会話環境：</p> <p>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.</p> <p>質問：12 問の社会問題に関する質問を出す.</p> <p>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.</p> <p>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もとづく.</p>               | <p>逐次無会話環境：</p> <p>会話がないこと以外は左記と同様</p> |

表 1. 4 つの実験環境

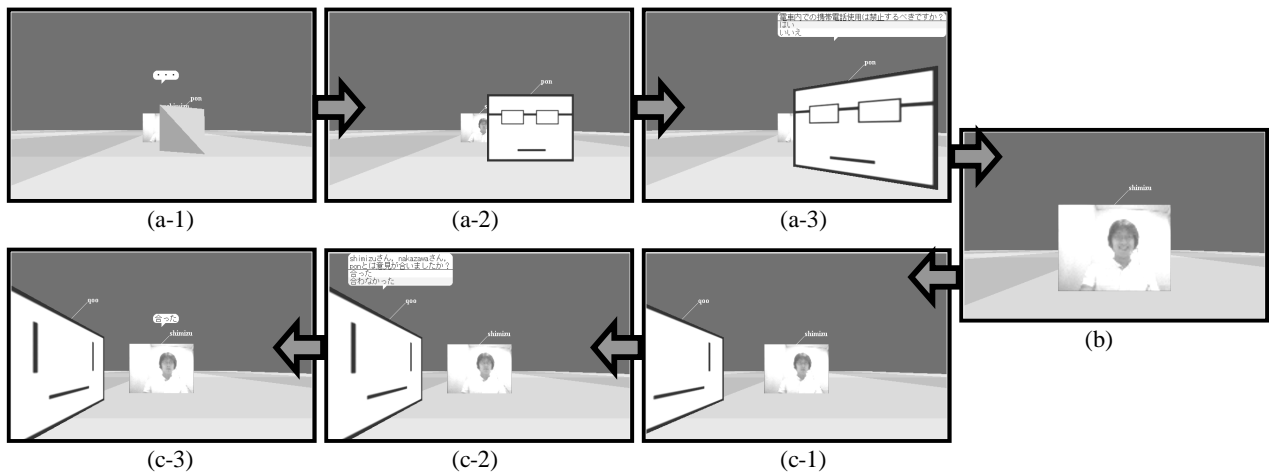


図 3. 逐次環境： (a-1) エージェントが相手と会話した後，(a-2) こちらに接近し，(a-3) 質問をする．(b) 相手と 1 対 1 で会話する．(c-1) エージェント No.2 が登場 (c-2) 両者に質問し，(c-3) 両者が答える．

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．同意エージェントは両方の被験者の答えに対し，10 問同意して 2 問だけ反意した．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取った．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．逐次会話環境では，エージェントの各フェーズの間に 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をした(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．逐次無会話環境では被験者同士は会話をしなかった．逐次環境では，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 被験者数及びアンケート内容

被験者は大学生である．同意エージェントが登場する条件を同意条件，反意エージェントが登場する条件を反意条件，不公平エージェントに同意される条件を不公平同意条件，不公平エージェン

トに反意される条件を不公平反意条件とする．同意条件，反意条件，不公平同意条件，不公平反意条件の被験者数はそれぞれ，同時会話環境で 10，10，8，8，同時無会話環境で 8，10，9，9，逐次会話環境及び逐次無会話環境では 8，8，8，8 である．合計 136 名の被験者が実験に参加した．同時環境の被験者数が各条件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全て 8 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推測，の 3 方向について質問した．同時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．逐次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分散分析で比較した結果，有意差の出た項目を表にまとめた(表 2 から表 5)．a, b, c, d, の欄はそれぞれ，同意条件，反意条件，不公平同意条件，不公平反意条件におけるアンケート結果の平均値である．1 が最も当てはまらない場合で，8 が最も当てはまる場合である．F 値の欄で，5% 有意には\*，1% 有意には\*\*，0.1%

| 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            | 5.0 | 2.4 | 5.8 | 2.1 | 23.2*** | <b>a&gt;b a&gt;d b&lt;c c&gt;d</b>        |
| 共感できる            | 4.9 | 2.1 | 4.6 | 3.9 | 5.4**   | <b>a&gt;b b&lt;c b&lt;d</b>               |
| 似ている             | 4.8 | 1.9 | 4.4 | 2.9 | 4.1*    | <b>a&gt;b b&lt;c</b>                      |
| 好感               | 5.1 | 2.4 | 4.3 | 3.1 | 7.6***  | <b>a&gt;b a&gt;d b&lt;c</b>               |
| 好き               | 4.9 | 2.8 | 4.5 | 3.6 | 5.6**   | <b>a&gt;b a&gt;d b&lt;c</b>               |
| 相手からみたエージェントについて |     |     |     |     |         |   |
| 意見が合う            | 5.0 | 2.6 | 2.4 | 6.0 | 12.3*** | <b>a&gt;b a&gt;c b&lt;d c&lt;d</b>        |
| 共感できる            | 4.3 | 2.6 | 2.6 | 5.5 | 6.5**   | <b>a&gt;b a&gt;c b&lt;d c&lt;d</b>        |
| 似ている             | 3.3 | 1.8 | 2.3 | 4.5 | 5.0**   | <b>b&lt;d c&lt;d</b>                      |
| 好感               | 5.0 | 2.9 | 2.9 | 5.6 | 9.2***  | <b>a&gt;b a&gt;c b&lt;d c&lt;d</b>        |
| 魅力的              | 4.4 | 2.9 | 3.4 | 4.9 | 3.6*    | <b>a&gt;b b&lt;d c&lt;d</b>               |
| 相手について           |     |     |     |     |         |   |
| 意見が合う            | 5.1 | 6.1 | 3.1 | 4.5 | 9.1***  | <b>a&gt;c b&gt;c b&gt;d c&lt;d</b>        |
| 共感できる            | 5.0 | 6.5 | 3.0 | 4.8 | 8.9***  | <b>a&lt;b a&gt;c b&gt;c b&gt;d c&lt;d</b> |
| 好感               | 5.8 | 5.9 | 4.1 | 4.1 | 5.1**   | <b>a&gt;c a&gt;d b&gt;c b&gt;d</b>        |

表 2. 逐次無会話環境・有意差項目

\* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$  ( $df=3,28$ )

a: 同意, b: 反意, c: 不公平同意, d: 不公平反意

有意には\*\*\*を付加した。全表において、自由度は (3, 28) 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% で有意差の出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出たかで確認できる。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出たかで確認できる。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出たかで確認できる。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 逐次無会話環境(表 2 参照)

バランス理論の成立がほぼみられた。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。エージェント及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 項目中 5 項目で有意差がみられ、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 3 項目で有意差がみられ、特に好意に関する質問「好感が持

| 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           | 3.3 | 1.6 | 4.8 | 1.6 | 10.6*** | <b>a&gt;b a&lt;c a&gt;d b&lt;c c&gt;d</b> |
| 関心が似ている          | 3.4 | 1.4 | 6.4 | 1.8 | 42.8*** | <b>a&gt;b a&lt;c a&gt;d b&lt;c c&gt;d</b> |
| 同じグループ           | 3.3 | 1.8 | 4.4 | 1.6 | 9.9***  | <b>a&gt;b a&gt;d b&lt;c c&gt;d</b>        |
| 相手からみたエージェントについて |     |     |     |     |         |   |
| 似通っている           | 3.4 | 1.9 | 1.5 | 6.6 | 31.6*** | <b>a&gt;b a&gt;c a&lt;d b&lt;d c&lt;d</b> |
| 関心が似ている          | 3.4 | 1.6 | 1.6 | 6.9 | 33.9*** | <b>a&gt;b a&gt;c a&lt;d b&lt;d c&lt;d</b> |
| 同じグループ           | 3.1 | 1.8 | 1.5 | 6.1 | 24.5*** | <b>a&gt;b a&gt;c a&lt;d b&lt;d c&lt;d</b> |
| 好ましい人物           | 3.8 | 2.5 | 2.1 | 4.0 | 3.4*    | <b>a&gt;c b&lt;d c&lt;d</b>               |
| 好意               | 3.4 | 3.0 | 2.1 | 5.0 | 5.5**   | <b>a&lt;d b&lt;d c&lt;d</b>               |
| 相手について           |     |     |     |     |         |   |
| 似通っている           | 3.9 | 4.1 | 2.4 | 3.5 | 3.2*    | <b>a&gt;c b&gt;c</b>                      |
| 関心が似ている          | 4.4 | 5.0 | 2.8 | 3.4 | 4.6**   | <b>a&gt;c b&gt;c b&gt;d</b>               |
| 同じグループ           | 3.6 | 3.8 | 2.1 | 3.4 | 3.7*    | <b>a&gt;c b&gt;c c&lt;d</b>               |
| 好意               | 5.0 | 5.6 | 3.5 | 4.9 | 5.2**   | <b>a&gt;c b&gt;c c&lt;d</b>               |
| 会話を続けたい          | 5.0 | 5.4 | 3.0 | 4.6 | 3.3*    | <b>a&gt;c b&gt;c</b>                      |

表 3. 同時無会話環境・有意差項目

てる」で H3 が受理された。残りの 2 つの類似に関する質問では、同意条件(a)と不公平反意条件(d)の間に **a>d** の有意な関係がみられないこと以外は、H3 に従う有意な関係がみられた。H3 はほぼ受理された。

2 つの質問で **a>d** がみられなかった理由は次の 2 点である。まず、同意条件(a)における相手への正の心情は反意条件(b)に比べて弱かった。これは、同じオブジェクトに共に好意を抱く人よりも、共に敵意を抱く人により類似性を感じることを示す。次に、不公平反意条件(d)は不公平同意条件(c)に比べ、相手により正の心情を抱いた。これは、自分が敵意を抱くオブジェクトに好意を抱く人よりも、自分が好意を抱くオブジェクトに敵意を抱く人により非類似性を感じることを示す。

#### 4.2 同時無会話環境(表 3 参照)

バランス理論が部分的に成立した。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 5 項目の有意差がみられたが、いずれも同意条件(a)と不公平反意条件(d)の間に **a>d** の有意な関係がみられなかった。H3 は部分的に受理された。この理由は逐次無会話環境と同様である。

| 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 | 2.5 | 5.8 | 2.1 | 25.5*** | a>b a>d b<c c>d     |
| 共感できる            | 5.3 | 3.0 | 5.4 | 2.3 | 14.4*** | a>b a>d b<c c>d     |
| 似ている             | 4.5 | 3.1 | 4.8 | 2.1 | 4.2*    | a>d c>d             |
| 相手からみたエージェントについて |     |     |     |     |         |                     |
| 意見が合う            | 4.8 | 2.1 | 2.8 | 6.4 | 14.0*** | a>b a>c a<d b<d c<d |
| 共感できる            | 5.1 | 3.4 | 2.8 | 6.3 | 6.9***  | a>c b<d c<d         |
| 似ている             | 4.0 | 2.1 | 2.4 | 4.6 | 5.3**   | a>b a>c b<d c<d     |
| 相手について           |     |     |     |     |         |                     |
| 意見が合う            | 4.5 | 5.9 | 5.1 | 3.8 | 4.2*    | a<b b>d c>d         |

表 4. 逐次会話環境・有意差項目

#### 4.3 逐次会話環境(表 4 参照)

バランス理論は成立しなかった。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。エージェントについて及び相手からみたエージェントについての質問で、類似に関する質問のみで有意差がみられた。H1 と H2 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出た項目が一つあるが、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。コミュニケーションチャネルが太くなると、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 同時会話環境(表 5 参照)

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。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質問で有意差がみられたが、質問「同じグループに感じる」の多重比較結果は有意な関係が一つのみであり、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。コミュニケーションチャ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は、エージェントが人間関係に及ぼす影響が若干弱まる。人間エージェント間の関係を先に形成

| 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                 | 3.8 | 1.3 | 4.5 | 2.4 | 7.2*** | a>b b<c c>d     |
| 相手からみたエージェントについて        |     |     |     |     |        |                 |
| 似ている                    | 2.5 | 1.6 | 2.4 | 4.5 | 7.4*** | a<d b<d c<d     |
| 関心が似ている                 | 3.3 | 1.8 | 2.3 | 5.3 | 9.9*** | a>b a<d b<d c<d |
| 同じグループ                  | 2.4 | 1.5 | 2.5 | 3.5 | 3.6*   | b<d             |
| 相手について、有意差がみられた項目はなかった。 |     |     |     |     |        |                 |

表 5. 同時会話環境・有意差項目

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、同時会話環境での会話分析[13]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ネルで、人間同士の会話が音声チャネルで行われる。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る舞いを行い、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ネルと被験者エージェント間のテキストチャ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。

第一に、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ネルで選択する(→1)と同時に、この応答内容をもう一人の被験者に対して音声チャネルを用いて伝達している(→2)。テキストチャネルにおける応答はもう一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、音声チャネルでの被験者間の情報交換はこのチャネルの確保のためのものであることが分かる(例 1)。

(例 1)<sup>1</sup> 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取りが割り込んでいる(→2)ことは、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掛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取りに対する言及・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: かなり突然だよな(笑)

→2 B→A: かなり突然ですね. 犬のくせに気を使ってますね(笑)

会話の流れとかに気を使ってますね.

→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取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る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. 例えば、不公平エージェントの場合、WOZ エージェントは、1)ターンを取る、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. 意見の違いが見られなければ、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. 例えば、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. エージェントの反意に対して、B はその理由を尋ねている(→1)が、エージェントはこれに答えず退出しており(→2)、被験者はこの振る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 <B> 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.

<B> なんだったんだ・・・

<B> Aさんは外国旅行とかは？

→3 <A> いきなり割り込んで・・・

→3 <B> ねえ.

さらに顕著な問題として、エージェントによる反意(→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 <B> 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か...

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、被験者同士の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

<sup>1</sup>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イタリックで示す.

ヤ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る舞いに対する評価が話し合われることが多い。従って、被験者間の音声チャンネルが追加されると、被験者間の肯定的関係は強化され、エージェントの人間関係への影響は打ち消される結果になる。

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に少ないものの、会話能力の限界によるエージェントの不適切な振る舞いに対する反発から、被験者間に同調が起こる場合があるため、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ため、エージェントの社会性を欠いた振る舞いは被験者間の同調を引き起こす契機となる。そして、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うるかを調べたものである。その第一歩として、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同意または反意を表明することで、人間に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。統計的分析及び会話分析の結果、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同意・反意の効果を失わせた。

この実験は、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共に、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。また、エージェントの会話能力が今日の技術よりも飛躍的に高まった時点において、再度実験を行う必要があることを示している。

**謝辞** 実験の運営に携わって下さいました(株)インターグループの大原裕子氏に感謝致します。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.L., 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] Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T.: FreeWalk: Supporting Casual Meetings in a Network, *CSCW-96*, pp. 308-314 (1996).
- [13] Levinson, S. C.: *Pragmatics*, Cambridge University Press (1983).