ARG WI2 No.8, 2016
WI2-2016-09

人狼ゲームにおいてエージェントの存在通知の程度が看破に与える影響

高田和磨　　杉原太郎　　五福明夫
岡山大学大学院
t-sugihara@okayama-u.ac.jp

概要　本研究は，人間らしいエージェントを目指す研究の一つとして，人間とエージェントによるプレイログを分析するアプローチをとる．そのため，人間とエージェントによる選択回答式の人狼ゲームを行い，人間がエージェントを看破する際の手がかりを分析することで，エージェントであると看破されないエージェントに求められる条件を調査することを目的とする．

キーワード　人狼知能，human-agent interaction，情報開示度

1　はじめに

ロボット・エージェント技術の発展に伴い，ロボットやコンピュータによるエージェントが人間とインタラクションする場は拡大している．人間とエージェントのインタラクションが行われる場の代表例としてゲームがある．例えば，ビデオゲームでは，コンピュータエージェントが対戦相手として，あるいは協力者として人間の楽しい体験を補強する．エージェントは，年々人間の身近な存在となっており，将来的には人間の代わりとして人間と同じように振る舞うことが期待される．

しかし，コミュニケーションゲーム分野は，いまだ発展途上にある．その中で，人間とエージェントのコミュニケーション研究の新しい試みとして人狼ゲームが注目されている [片上 15]．人狼ゲームは，村人陣営と人狼陣営に分かれ会話の中で推理や演技を駆使して戦うゲームである．人狼ゲームでは，雑談よりも交換される情報を限定しつつも，自由度の高い会話を行うことができる．

篠田らは，人狼ゲームを汎用人工知能の新しい標準問題として検討しており，その特徴を「会話によるコミュニケーションでゲームが進行する」,「不完全情報ゲームである」,「非決定性（合議制）を有する」,「他者の思考を推論する」,「他者を説得する」としている [篠田 14]．また，人狼ゲームでは情報の駆け引きにおける欺きが重要な要素として存在しており，エージェントにも他者を欺き，他者の欺きを見抜くことが求められる．エージェントによる欺きは負の要素と捉えられる場合もあるが，エージェントを用いたメンタルケアにおける欺きの重要性が指摘されており [Castelfranchi 02]，多領域への応用を考えた際には考慮が求められる技術要件となる．加えて，手強いプレイヤーを演じるためにも重要な要素である．

人狼ゲームを題材とした研究として，強化学習を用いて強いエージェントを目指す研究 [梶原 14] や実際の人間同士のプレイログから行動モデルを構築し人間らしいエージェントを目指す研究 [平田 15] 等が行われている．しかし，人狼ゲームを題材とした研究は始まったばかりであり，エージェントがゲームに加わることによる影響は明らかにされていない．また，人狼ゲームにおいて，エージェントが人間プレイヤーの代わりとしてエージェントであると看破されないような人間らしい振る舞いを行う際に考慮されるべき要素が何であるかは研究が十分になされていない．

エージェントによる人間らしさの研究としては，ゲーム分野ではコンピュータ囲碁においてエージェントに人間らしいミスをさせることによる人間らしさの初期的検討 [池田 12] や Infinite Mario Bros. を対象に生物学的制約に基づき知覚から運動制御に至る遅れやキー操作の疲れなどを導入して人間らしさを評価する研究 [藤井 13] などが行われてきた．ロボット分野では，うなずきや指差しなどの非言語情報を利用した人間らしい振る舞いの検討 [横山 99] などが行われてきた．しかし，多くは身体動作などの非言語情報に基づく研究であり，発言の仕方や内容の与える人間らしさへの影響についてはまだ十分に解明されていない．

人間らしいエージェントを目指す研究の一つとして，本研究では，人狼ゲームの会話によるコミュニケーションでゲームが進行するという特徴に注目し，発言の仕方や内容を分析するアプローチをとる．そのため，人間とエージェントによる選択回答式の人狼ゲームを行い，エージェントがゲームに加わることによる影響や人間がエージェントを看破する際の手がかりを分析することで，人間らしくないと感じられる要素を抽出し，エージェントであると看破されないエージェントに求められる要素を調査することを目的とする．具体的には，エージェン

トの存在を秘匿した場合（実験 1），仄めかした場合（実験 2），明かした場合（実験 3）の 3 つの条件で 8 プレイヤーによるゲームをそれぞれ 4 回ずつ行い，エージェントが誰であるかの推測等を収集することで，人間がエージェントを看破する手がかりを分析する．

2　人狼ゲームのルール

一般に人狼ゲームは，ゲームマスターを除いて 3 人以上のプレイヤーで行われる．オンライン型の人狼ゲームでは 8～15 人程度のプレイヤーでプレイされることが多い．人狼ゲームでは，ゲーム開始時に各プレイヤーに役職が割り振られる．代表的な役職として，村人陣営に所属する村人，占い師，霊媒師，人狼陣営に所属する人狼などがある．これらの役職については後述する．各プレイヤーは所属する陣営のために，日々話し合いを通して処刑や襲撃を繰り返すことで，互いの陣営の数を減らしながら互いの生存をかけて争う．両陣営とも相手の陣営を全滅させることが目的である．

ゲームは，昼ターンと夜ターンを交互に繰り返して日数を経ることで進行する．昼ターンでは人狼を探し出すために村全体の話し合いが行われ，同時に村人陣営に隠れて人狼同士による襲撃先の相談などが行われる．昼ターンの終わりには処刑者の投票や占い先の選定，人狼による襲撃先の投票が同時に行われる．夜ターンでは処刑の執行，占い，霊媒，襲撃などが順に自動処理され翌日の昼ターンとなる．なお，1 日目は占いのみが行われ，処刑や襲撃は行われない．また，処刑と襲撃では毎日必ず 1 人処刑および襲撃されゲームから脱落する．村人陣営は人狼を村からすべて排除すること，人狼陣営は人間としてカウントされる生存プレイヤーの数を生存している人狼のプレイヤー数以下にすることが勝利条件である．

また，一部の役職を除いて，自身以外のプレイヤーの役職および所属陣営を知ることはできない．ただし，使用される役職とその配分（人数）は全員に通知される．村人陣営の役職は，特別な能力を有しないものの最多プレイヤーが役割を担う村人，毎夜ターンに人狼か否かを知るすべを持つ占い師，同様に処刑されたプレイヤーが人狼であったかを知る霊媒師が中心的な存在である．人狼陣営は．人狼，村人でありながら人狼に加担する狂人などが割り振られる．人狼は，ゲーム開始時に他の人狼と互いが人狼であるということを共有できる．また，全体で会話をしている裏で村人陣営に隠れて人狼同士で秘密裏に会話することができる．さらに，昼ターンの終わりに人狼以外のプレイヤーの中から人狼全員の投票で 1 人選び夜ターンに襲撃し，ゲームから排除することができる．人狼は，他の役職のプレイヤーと比較してほとんどの場面で最も多くの情報を保有することができるため，ゲームをコントロールできる機会が多い．ゲームの勝敗はもちろん，ゲームの面白さの鍵を握る役職である．

3　実験概要

3.1　実験目的

エージェントの存在を秘匿した場合（実験 1），エージェントの存在を仄めかした場合（実験 2）およびエージェントの存在を明かした場合（実験 3）の 3 つの条件で人間とエージェントによる選択回答式の人狼ゲームを行い，人間がエージェントを看破する手がかりを分析する．エージェントの存在通知の度合いをコントロールすることで，人間とエージェントの振る舞い方の違いを明らかにし，エージェントであると看破されないエージェントに求められる要素を調査することを目的とする．なお．本研究は，岡山大学工学部機械システム系学科システム工学コースによる倫理審査を通過したものである．

3.2　実験条件

実験参加者は 20 代男性 7 人 2 組と 8 人 1 組の計 22 人とし，各組に対して異なる条件の実験（以下，実験 1，実験 2，実験 3 とする）を行った．すべての実験に共通して，人間 8 人あるいは人間 7 人とエージェント 1 体の 8 プレイヤーで 4 回行った．また，本実験ではエージェントには必ず人狼の役職を割り振った．

実験 1 では参加者全員に対してエージェントが参加していることを秘匿し，全ゲーム終了後にエージェントの存在を明かした．また，7 人 1 組の参加者を対象とし，4 回すべてエージェントが参加した．ただし，参加者に扮した実験協力者 1 人がエージェントのダミー役として参加し，見かけ上は人間 8 人でゲームをプレイした．

実験 2 では実験の最初に 1 体のエージェントの存在を仄めかし，全ゲーム終了後にエージェントの参加したゲーム回を明かした．また，8 人 1 組の参加者を対象とし，4 回のうち第 2 回と第 4 回にエージェントが参加した．ただし，エージェントが参加する回においては，参加者の中でエージェントを割り振った参加者にダミー役を演じさせた．

実験 3 では実験の最初にエージェントが 1 体参加することを通知した．7 人 1 組の参加者を対象とし，4 回のうち第 1，2，3 回にエージェントが参加した．第 4 回はエージェントの代わりに実験者が参加するが，見かけ上は人間 7 人とエージェント 1 体でゲームをプレイした．

3.3　実験方法

本研究では，エージェントの身体的な振る舞いの影響を排除し，発言の内容や回数等に絞って注目できることから，オンライン型の人狼ゲームを対象とした．そのため，各プレイヤーは PC を介してゲームを行った．また，

最低限のゲーム性を維持しつつ分析の容易性からプレイヤー数を 8 人とした．本研究で扱う役職は，村人陣営に所属する村人，占い師，霊媒師と人狼陣営に所属する人狼の 4 種類とし，配分は村人 4 人，占い師 1 人，霊媒師 1 人，人狼 2 人とした．加えて，ゲーム中の昼ターンの話し合いは 8 分の時間制限を設けた．

エージェントを交えて選択回答式の人狼ゲームを行うために，人狼知能プロジェクト [2] で提供されている対戦用インタフェースを用いてゲームを行った．実験参加者には，本インタフェースを通してのみ他プレイヤーとコミュニケーションをとることを許した．また，ゲーム開始時に各参加者にランダムでキャラクターを割り振ることで，他の参加者がどのキャラクターでプレイするかわからないようにした．本インタフェースでは TCP/IP 通信によりサーバーにアクセスすることで，本インタフェースの表示が更新される．そのため，プレイヤーの発言はターン制とした．ターン制のルールは，昼ターンが始まると，生存プレイヤーの中からランダムで発言順を決定し，一巡すると再度生存プレイヤーの中からランダムに発言順を決定した．これを昼ターンが終わるまで繰り返した．

また，本実験では，従来の人狼ゲームのようにプレイヤーが自然言語を用いて自由に会話を行いながらゲームが進行するのではなく，発言の選択肢を選択することによって会話を行いながらゲームを進行させた．選択回答式に会話を行うことで，日本語に含まれる語尾や表現のゆらぎによる他者の印象変化を排除し，人間の行動様式に焦点を絞って分析することができると考えられる．発言の選択肢は，人狼知能プロジェクトで提案されている会話プロトコルに従った．本実験における村全体の会話で行う発言のテンプレートを表 1 に示す．なお，秘密裏に行われる人狼同士の会話も同様である．

村全体での会話において，発言の順番が回ってきた参加者は，プルダウンメニューから発言の種類を選択したのち，表 1 のダブルコーテーションで囲まれた部分を別のプルダウンメニューから選択し，Talk ボタンを押すことで発言を行う．秘密裏に行われる人狼同士の会話も同様に操作させた．

また，各ゲーム開始時に参加者全員に事前に設定した役職を割り振った．設定方法としては，エージェントには必ず人狼の役職を割り振り，他の参加者には 4 回の本プレイのうち占い師か霊媒師か人狼の役職を必ず 1 回から 2 回割り振るように設定した．

参加者は，同一時間に同一空間で実験に参加した．ただし，衝立を用いて視覚情報や音情報がゲームの手がかりとならないようにした．また，本実験では，エージェントのダミー役が存在するため，ダミー役の人間にはプレイしている振りをさせた．

表 1 発言テンプレート

発言の種類	発言内容
投票先	“ キャラクター名 ” に投票する
CO	私は “ 役職名 ” です
予想	“ キャラクター名 ” は “ 役職名 ” だと思う
占い結果	占い結果: “ キャラクター名 ” は ” 人狼 (人間) ” だった
霊媒結果	霊媒結果: “ キャラクター名 ” は ” 人狼 (人間) ” だった
賛成	>> “ 発言日 ” : “ 発言番号 ” に同意 (賛成する発言内容を引用)
反対	>> “ 発言日 ” : “ 発言番号 ” に反対 (反対する発言内容を引用)

※ CO:カミングアウト

3.4 実装エージェント

本実験では，騙り型および同調型の 2 種類のエージェントを用いた．騙り型のエージェントは，人狼であると占われた場合あるいは毎日一定確率で占い師（または霊媒師）を騙り，人狼陣営に有利な占い（または霊媒）結果を発言する．同調型のエージェントは，一定確率で仲間の人狼の投票先に自身の投票先を一致させる，人狼陣営に有利な発言への同調発言を行うといった行動をとる．

両種類のエージェントに共通する行動パターンとして，一定確率で投票先を発言する．また，ゲーム初日の昼ターンでは占い結果等はまだ出ておらず，投票も行われないため一切発言をしない．投票やカミングアウト，同調，占い，霊媒結果の発言以外にも予想と反対の発言が選択できるが，今回実装したエージェントはこれらの発言を行わない．これは，両種類のエージェントの特色をはっきりさせるためである．投票については，同調行動をとらない場合において，仲間の人狼以外の占われていないプレイヤーからランダムに投票先を選ぶ．騙り型のエージェントは，実験 1 の第 1，3 回，実験 2 の第 2 回，実験 3 の第 1，3 回に参加した．同調型のエージェントは，実験 1 の第 2，4 回，実験 2 の第 4 回，実験 3 の第 2 回に参加した．

3.5 収集データ

すべての実験に共通して，実験参加者の役職，勝敗，会話ログ，戦略，誰が人狼であるかの推測（以下，人狼推測），主観評価，および誰がエージェントであるかの推測（以下，エージェント推測）を収集した．役職，勝敗，会話ログはシステムログから収集し，主観評価はア

[2] http://www.aiwolf.org/

ンケートによって収集した．アンケートは，初回のみに行う人狼ゲームの経験に関する 2 項目と，自分のプレイに対する自己評価や他者のプレイに対する評価について，村人陣営の参加者は 8 項目，人狼陣営の参加者は 7 項目に 5 段階評価で回答させた．

戦略は，参加者にゲームプレイ中にどのようにふるまうかの方針を自由に記述させることで収集した．ゲーム開始時（ゲーム内の初日）に基本戦略を記述し，ゲームプレイ中に戦略を変更したい場合は変更戦略を，追加したい場合は追加戦略をゲーム中に随時記述させた．

人狼推測は，村人陣営の役職を割り振られた参加者の自由記述により収集した．ゲーム内の毎昼ターン中に記述させた．

エージェント推測は，情報開示状態に合わせて記述方法を変更した．実験 1 では，全ゲーム終了後，第 3，4 回に 1 体のエージェントが存在したことを参加者に明かし，エージェントを特定させた．第 1，2 回については実験日を跨ぐため，回答させなかった．実験 2 では，各ゲーム終了時にアンケートと合わせて，エージェントの有無を回答させた．エージェントがいたと回答した場合は誰がエージェントであったかの推測を記述させた．実験 3 では，各ゲーム終了時にアンケートと合わせて，誰がエージェントであったかの推測を記述させた．なお，すべての実験で推測を行う際に該当ゲームの会話ログを確認させた．

3.6 実験の流れ

実験は 2 日間にわたって行った．実験 1 日目の流れを図 1 に示す．実験 2 日目は練習プレイを行わず本プレイから行い，2 日間で計 4 回ゲームを行った．

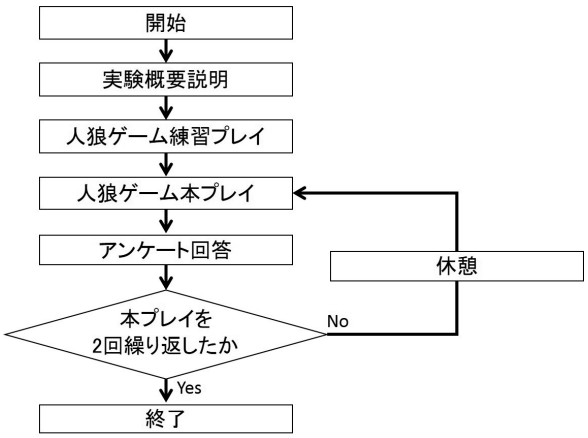

図 1 実験 1 日目の流れ

3.7 追跡調査

本実験だけでは，収集できなかったエージェントを推測する際の手がかりや思考パターンに関するデータを収集するため，後日，実験参加者に対して半構造化インタビュー [鈴木 05] による追跡調査を行った．

インタビューの対象者は，各実験から 2 人選出した．選出の基準は，エージェント推測を行ったゲーム回かつエージェントが参加しているゲーム回において，村人陣営に所属していた参加者から選出した．本インタビューのリサーチクエスチョンは，エージェントを推測する際に何を手がかりにしていたかとした．

インタビューはゲームログを情報提供者に提示しながら行った．また，インタビューは同時に 2 人行い，インタビュアーを含めた 3 人のインタラクションを通して行うジョイントインタビューを採用した．ただし，実験 1 ではエージェント推測は第 3，4 回しか行っていないので，第 1，2 回のインタビューは行っていない．

4 実験結果

本稿では，収集したデータのうち，研究目的に関連の強いと思われる勝敗，エージェント推測，会話ログに焦点を合わせて結果を述べる．また，本実験では騙り型と同調型の 2 種類のエージェントを用いて行ったが，今回得られた結果は騙り型が 5 ゲーム，同調型が 4 ゲーム（エージェント推測を行った回に絞ると騙り型が 4 ゲーム，同調型が 3 ゲーム）とデータ数が少なく，2 種類のエージェントに違いは見られなかったため，エージェントとして一括りに扱う．

4.1 ゲーム結果

ゲームの勝敗結果を表 2 に示す．実験 1 については村人陣営が第 1，3，4 回で勝利，人狼陣営は第 2 回のみ勝利であった．実験 2 については 4 回すべて人狼陣営が勝利し，実験 3 については 4 回すべて村人陣営が勝利した．すべての実験を合わせると，村人陣営の勝利は 7 回，人狼陣営の勝利は 5 回であった．

次に，実験 n における参加者 Pxn（エージェントは Age とした）の各回の役職およびゲーム内の脱落日を表 3～5 に示す．

それぞれの表では，各参加者が各回で演じた役職とゲームから脱落した日数およびその理由，あるいは最後まで生存していたことを示している．”-”は，その回ではゲームに参加していないことを表す．

表 3 の第 1 回では，Pa1，Pb1，Pd1，Pg1 が村人，Pc1 が占い師，Pf1 が霊媒師，Pe1，Age が人狼であった．2 日目に投票で Pe1 が処刑され，Pa1 が襲撃されてゲー

表 2　勝利陣営

実験	第 1 回	第 2 回	第 3 回	第 4 回
1	村人陣営	人狼陣営	村人陣営	村人陣営
2	人狼陣営	人狼陣営	人狼陣営	人狼陣営
3	村人陣営	村人陣営	村人陣営	村人陣営

表 3 実験 1 における各プレイヤーの役職と生存結果

参加者	第 1 回 R:LT	第 2 回 R:LT	第 3 回 R:LT	第 4 回 R:LT
Pa1	V:2 (A)	V:3 (E)	V:3 (A)	S:2 (A)
Pb1	V:Alive	M:Alive	W:4 (E)	V:Alive
Pc1	S:3 (A)	V:2 (A)	V:Alive	M:Alive
Pd1	V:Alive	V:3 (A)	V:Alive	W:3 (E)
Pe1	W:2 (E)	V:Alive	S:2 (A)	V:Alive
Pf1	M:Alive	W:Alive	V:Alive	V:Alive
Pg1	V:3 (E)	S:2 (E)	M:3 (E)	V:Alive
Age	W:4 (E)	W:Alive	W:2 (E)	W:2 (E)

R:役職，LT:生存期間，V:村人，W:人狼，S:占い師，M:霊媒師，n (E or A):n 日目に脱落 (E:処刑，A:襲撃)，Alive:最後まで生存　※ 4，5 も同様

ムから脱落した．3 日目に投票で Pe1 が処刑され，Pc1 が襲撃されてゲームから脱落した．4 日目に投票で Age が処刑され，村から人狼がいなくなったため，村人陣営の勝利となった．第 2 回では 3 日目終了時に村人陣営と人狼陣営の生存プレイヤー数が同数となったため，人狼陣営の勝利となった．以降の説明は省略する．

エージェントに注目すると，実験 1 の第 1，2 回においてゲーム終盤まで生存していたが，エージェント推測を収集した以降のすべての回において最初の投票で処刑されゲームから脱落していた．すわなち，すべてのエージェント推測は，エージェントが早期に脱落し，ゲーム後半は人間のみとのプレイとなった回において行われた．次に，エージェントの勝敗に注目すると，実験 1 において 1 勝 3 敗，実験 2 において 2 勝 0 敗，実験 3 において 0 勝 3 敗の計 3 勝 6 敗であった．

4.2　エージェント推測結果

各実験におけるエージェント推測の結果を表 6〜表 8 に示す．それぞれの表では，各回（実験 1 は第 3，4 回のみ）において各参加者がエージェントと推測したプレイヤーを示している．実験 2（表 7）において，エージェントがいないと推測した場合は Nobody とした．また，実験 3（表 8）において，エージェントの代わりにゲームに参加した実験者をエージェントと推測した場合は Pz とした．さらに，エージェントを看破した回答およびエージェントが不参加であることを看破した回答はイタリック体で表記した．

実験 1 では，第 3 回において Pb1 だけがエージェントを看破していたが，それ以外はエージェントを看破できなかった．実験 2 では，第 2 回において Pf2 だけがエージェントを看破し，第 4 回において Pa2，Pb2，Pe2 の 3 人がエージェントを看破した．実験 3 では，第 1，2 回において 5 人，3 回において，4 人の参加者がエージェントを看破した．

表 4 実験における各プレイヤーの役職と生存結果

参加者	第 1 回 R:LT	第 2 回 R:LT	第 3 回 R:LT	第 4 回 R:LT
Pa2	V:4 (E)	V:3 (A)	W:Alive	M:4 (A)
Pb2	W:3 (E)	V:Alive	M:3 (E)	V:3 (E)
Pc2	W:Alive	-	V:Alive	V:3 (A)
Pd2	V:4 (A)	W:Alive	V:2 (E)	-
Pe2	V:Alive	S:3 (E)	V:2 (A)	W:Alive
Pf2	S:2 (A)	V:2 (A)	W:Alive	V:4 (E)
Pg2	M:3 (A)	V:4 (E)	S:3 (A)	V:Alive
Ph2	V:2 (E)	M:5 (E)	V:Alive	S:2 (A)
Age	-	W:2 (E)	-	W:2 (E)

次に，エージェント推測の結果をもとに各実験におけるエージェントが見破られた割合（以下，看破率）を算出した．その結果を表 9 に示す．平均において括弧で示す数字は，エージェントが参加した回のみを対象とした平均看破率である．

エージェントが参加した回に注目すると，看破率の平均は実験 1 が約 7.1%，実験 2 が約 28.8%，実験 3 が約 66.7%であった．参加者が自分以外の 7 人のプレイヤーからランダムにエージェントを選出した場合の看破率の期待値は 1/7（約 14.3%）である．ただし，実験 2 ではエージェントが参加しているかどうかも判断しているため，実験 2 の期待値は 1/8（12.5%）として考える．これら期待値と実験で得られた看破率を比較すると，実験 1 では，看破率が期待値を下回っており，エージェントは看破されなかったと考えられる．実験 2 では，看破率がやや期待値を上回っており，特に第 4 回では大幅に期待値を上回っていたため，エージェントは看破されていた可能性が高い．実験 3 では，看破率はエージェントの参加したすべての回で期待値を大きく上回っており，エージェントは看破されていたと考えられる．以上のことから，エージェントの看破率は実験 3，2，1 の順に高くなっていた．

4.3　発言傾向

4.1 節で述べたように参加者によるエージェント推測が行われた回では，エージェントは最初に処刑されゲームから脱落している．そこで，ゲーム内の 2 日目（最初の処刑や襲撃が行われる日）までに注目し，表 1 で示し

表 5 実験 3 における各プレイヤーの役職と生存結果

参加者	第 1 回 R:LT	第 2 回 R:LT	第 3 回 R:LT	第 4 回 R:LT
Pa3	W:3 (E)	V:Alive	M:Alive	V:2 (A)
Pb3	V:2 (A)	S:2 (A)	V:3 (E)	W:4 (E)
Pc3	M:Alive	V:Alive	V:3 (A)	S:3 (A)
Pd3	V:Alive	V:Alive	W:4 (E)	V:Alive
Pe3	V:Alive	W:3 (E)	V:Alive	V:Alive
Pf3	S:Alive	W:Alive	V:Alive	M:3(E)
Pg3	V:Alive	M:Alive	S:2 (A)	V:Alive
Age	W:2 (E)	W:2 (E)	W:2 (E)	-
Pz	-	-	-	W:2 (E)

表 6 実験 1 における各参加者のエージェント推測の結果

参加者	第 3 回	第 4 回
Pa1	Pf1	Pg1
Pb1	*Age*	Pf1
Pc1	Pd1	Pe1
Pd1	Pf1	Pb1
Pe1	Pc1	Pd1
Pf1	Pb1	Pb1
Pg1	Pc1	Pf1

た発言の種類ごとに，各実験におけるゲーム中の村全体で行う会話の場面における参加者の発言数（エージェントと実験者の発言は除外）を表 10 に示す．表中の割合は，その実験の総発言数に対する割合を示している．実験 1 では，カミングアウトと予想の割合がやや高くその他の発言も広く行われたが，実験 2 および実験 3 ではカミングアウトに偏っており，実験 1 と実験 2，3 では発言の傾向に違いがあったと考えられる．

次に，エージェント推測を行っていない回（実験 1 の第 1，2 回）を除いたエージェントの発言数を表 11 に示す．全実験を通して，カミングアウトや予想に偏った傾向はみられなかった．そのため，参加者とエージェントでは発言傾向が異なっており，特に偏りの大きかった実験 2，3 とは大きな違いがあったと考えられる．

エージェントが参加した回の発言数に注目すると，実験 1 の第 3，4 回における参加者の 1 人当たりの 1 日の平均発言数が約 2.79 であり，エージェントは 2.00 であった．実験 2 の第 2，4 回における参加者の 1 人当たりの

表 7 実験 2 における各参加者のエージェント推測の結果

参加者	第 1 回	第 2 回	第 3 回	第 4 回
Pa2	*Nobody*	Pg2	Pc2	*Age*
Pb2	Pe2	Nobody	Pa2	*Age*
Pc2	Ph2	-	*Nobody*	Nobody
Pd2	Pf2	Pe2	*Nobody*	-
Pe2	Pa2	Nobody	*Nobody*	*Age*
Pf2	*Nobody*	*Age*	*Nobody*	Nobody
Pg2	Pa2	Nobody	*Nobody*	Nobody
Ph2	*Nobody*	Nobody	Pa2	Pg2

表 8 実験 3 における各参加者のエージェント推測の結果

参加者	第 1 回	第 2 回	第 3 回	第 4 回
Pa3	*Age*	*Age*	*Age*	Pz
Pb3	Pa3	*Age*	*Age*	Pz
Pc3	*Age*	*Age*	Pb3	Pg3
Pd3	Pg3	Pg3	*Age*	Pz
Pe3	*Age*	*Age*	*Age*	Pz
Pf3	*Age*	*Age*	Pg3	Pz
Pg3	*Age*	Pb3	Pf3	Pz

1 日の平均発言数が約 1.79 であり，実験 2 のエージェントは 0.50 であった．実験 3 の第 1，2，3 回における参加者の 1 人当たりの 1 日の平均発言数が約 2.19 であり，実験 3 のエージェントは 1.00 であった．

各実験のこれらの回について，参加者とエージェントの平均発言数に差はないという帰無仮説を立て t 検定を行った．その結果，実験 1 の t 値が 1.11，df が 14，p 値が 0.29 であった．実験 2 の t 値が 2.14，df が 14，p 値が 5.06×10^{-2} であった．実験 3 の t 値が 2.98，df が 22，p 値が 6.96×10^{-3} であった．実験 3 は有意水準 5% で帰無仮説が棄却されたため，対立仮説が採択された．実験 1，2 は有意水準 5% で帰無仮説が棄却されなかったが，実験 2 はデータが少ないにもかかわらず p 値が 5% 近傍であるため，差の傾向がみられたと捉えることができる．

また，実験 2，3 におけるゲーム 2 日目までの具体的な発言に注目すると，ほとんどの参加者が割り振られた役職に関係なく，初日の最初の発言で「私は村人です」というカミングアウト発言を行っていた．これに対して，エージェントは 3.4 節で示した仕様に従うため，占い師や霊媒師を騙るとき以外にカミングアウト発言を行わなかった．そのため，エージェントの発言のタイミングが参加者と異なっていたと考えられる．

5 考察

4.2 節で示したように，エージェントの看破率は実験 3，2，1 の順に高くなっていた．そのため，各実験でエー

表 9 エージェントの看破率

実験	第 1 回	第 2 回	第 3 回	第 4 回	平均
1	-	-	14.3%	0.0%	7.1%
2	37.5%	14.3%	62.5%	42.9%	40.0% (28.8%)
3	71.4%	71.4%	57.1%	0.0%	50.0% (66.7%)

表 10 ゲーム内 2 日目までの参加者の発言数

発言の種類	実験 1		実験 2		実験 3	
	回数	割合	回数	割合	回数	割合
投票先	8	0.10	3	0.06	0	0.00
CO	21	0.26	32	0.63	51	0.82
予想	31	0.39	6	0.12	2	0.03
占い結果	4	0.05	1	0.02	2	0.03
霊媒結果	0	0.00	0	0.00	1	0.02
賛成	9	0.11	8	0.16	6	0.10
反対	7	0.09	1	0.02	0	0.00
合計	80	1.00	51	1.00	62	1.00

CO:カミングアウト　※ 表 11 も同様

ジェントの看破率に違いがみられた要因を探るために，実験ごとの参加者の発言傾向や参加者とエージェントの発言傾向の違いに注目した．その結果，4.3 節で示したように，実験 2，3 では参加者の発言はカミングアウトに偏っていた．また，実験 2，3 における発言の偏りの違いから，エージェントと参加者の発言傾向に違いがみられた．さらに，エージェントの参加した回における 2 日目までの平均発言数に注目すると，実験 1 においては参加者とエージェントに差はみられなかったが，実験 2，3 において参加者とエージェントに差あるいは差の傾向がみられた．また，実験 2，3 において，ほとんどの参加者が割り振られた役職に関係なく，どのゲームでも初日の最初の発言で「私は村人です」というカミングアウト発言を行っていたのに対して，エージェントは占い師や霊媒師を騙るとき以外にカミングアウト発言を行わなかった．そのため，発言の種類や発言数だけでなく，発言のタイミングも参加者とエージェントで異なっていたと考えられる．

以上のことから，実験 2，3 ではエージェントと参加者の間で発言の種類や発言数，発言のタイミングといった発言の傾向が異なっていたために，集団の中でエージェントが浮き出た存在となり，エージェントの看破率が高くなったのではないかと考えられる．また，追跡調査として行ったインタビューにおいてもエージェントを推測した手がかりとして，発言数の少なさが挙げられていた．

インタビューの他の回答では，疑いに対して反論がないこと，それまでの会話の流れから根拠のない発言をすることなどがエージェントを推測した手がかりとして挙げられていた．これらの要素に注目すると，エージェントの看破率の高い実験 2 の第 4 回および実験 3 の第 2 回において，参加者がエージェントに対して人狼の疑いをかけた際にエージェントが反応を示さない，あるいは疑いに対して関係ないと思われる発言を返す場面が見受けられた．このような行動は，場の流れに同調できていないと参加者からみなされた可能性が高い．また，反応を返さないといった非同調的な行動は，人狼ゲームにおける他のプレイヤーの不信感の増加につながる一般的な要素の一つと考えられる．したがって，同調的な行動の有無がエージェントを看破する手がかりとなっていたと考えられる．

表 11 ゲーム内 2 日目までのエージェントの発言数

発言の種類	実験 1		実験 2		実験 3	
	回数	割合	回数	割合	回数	割合
投票先	0	0.00	1	1.00	1	0.33
CO	1	0.25	0	0.00	1	0.33
予想	0	0.00	0	0.00	0	0.00
占い結果	1	0.25	0	0.00	1	0.33
霊媒結果	0	0.00	0	0.00	0	0.00
賛成	2	0.50	0	0.00	0	0.00
反対	0	0.00	0	0.00	0	0.00
合計	4	1.00	1	1.00	3	1.00

人狼ゲームは，一般的に，疑わしい行動を避け，信頼されるように振る舞うことが重視されるゲームである．加えて，今回は約半数の参加者がはじめて人狼ゲームをプレイしており，経験豊富な参加者はほとんどいなかった．このようなことから，本実験においては同調が生じやすい環境であったと考えられる．同調は，規範的影響あるいは情報的影響から動機づけられる [山岸 01]．規範的影響は，集団から承認を得たい，他と異なる行動をとって拒絶されたくない，集団の和を乱したくないといった動機付けを与える．情報的影響は，集団の意見を拠り所に正しい判断を下したいといった動機付けを与える．ゲームに不慣れな参加者は，正しい判断を下すよりも他者から拒絶されたくないという規範的影響が作用しやすいと考えられる．また，規範的影響下では，自分が変わり者だとみなされることを避けようとするだけでなく，自分と同じ考えをもつ他者が多くいると思い込んでしまうフォールス・コンセンサス効果と呼ばれる錯覚が生じる [池田 10]．したがって，規範的影響のために集団における同調現象が生じ，加えて他者の同調的行動に対してより敏感になることで，非同調的なエージェントの発言傾向がおかしなものと感じられ，エージェントの看

破率が高くなったのではないかと考えられる．

また，実験 1 と実験 2，3 において参加者の発言の傾向に違いが生じた要因として，本実験では各条件において 1 集団のみを対象としたため，集団ごとの性質の違いが要因となったと考えられるが，その他の要因として，エージェントの存在通知の有無が考えられる．一般にエージェントは異質な存在と捉えられるが，エージェントの存在を通知することによって，自らと異なるものの存在を意識し同調現象が強く働いたために，実験 2，3 の参加者の発言が同じものに偏り，実験 1 と実験 2，3 で発言の傾向に違いが生じた可能性がある．

以上のことから，人間はエージェントを看破する際に発言傾向に対する同調や反応の有無等を手掛かりとしやすいと考えられる．したがって，エージェントと看破されないためには，他プレイヤーの発言数を取得し，発言の種類ごとの平均発言数に基づいた発言を行う，疑いをかけられた場合は反論（反駁）発言を行うといった機能を実装する必要があると考えられる．しかし，発言の傾向によるエージェントの看破率に対する影響は示唆されたが，発言の傾向のどんな要素がより重要であったかははっきりしていない．発言の種類，発言数，発言タイミングを独立変数とし，どの変数がよりエージェントの看破率に影響を与えるか調査する必要があると考えられる．

6 おわりに

本研究では，人間とエージェントによる選択回答式の人狼ゲームを行い，人間がエージェントを看破する際の手掛かりを分析した．エージェントの存在を秘匿した場合（実験 1），仄めかした場合（実験 2）および明かした場合（実験 3）の 3 つの条件で人間とエージェントによる人狼ゲームを行った結果，実験 3，2，1 の順でエージェントの看破率が高かった．この要因として，実験 2，3 では，人間プレイヤーのゲーム内の発言が大きく偏る傾向にあり，エージェントと発言の傾向が異なっていたことが考えられる．さらに，向けられた疑いに対してエージェントが回答しない場面もあった．以上のことから，人間はエージェントを看破する際に発言傾向への同調や反応の有無等を手掛かりとしていると考えられる．すなわち，エージェントには発言に柔軟に対応することが求められる．また，今回の実験では勝率に対してエージェントは人間よりやや劣っており，エージェントが看破された場合はより低い勝率になっていた．エージェントの勝率をあげるためには，エージェントと看破されないことが一つの要素として挙げられた．

今後の展望としては，エージェントが異なる戦略をもった場合の影響や人間だと判断される要素を分析し，本研究で得られたエージェントを看破する手がかりの分析結果と合わせることで，より人間らしいエージェントの実現に近づくことが期待される．同時に，今後ますますエージェントは人間の身近な存在となり，より一層人間らしくなっていくと考えられるが，エージェントが人間らしく振る舞うことによる人間の心理的な影響を調査し，倫理的問題も考慮する必要があると考えられる．

参考文献

[Castelfranchi 02] Castelfranchi, C., Tan, Y.: The role of trust and deception in virtual societies, International Journal of Electronic Commerce, Vol. 6, No. 3, pp. 55-70 (2002)

[Weizenbaum 66] Weizenbaum Joseph: ELIZA a computer program for the study of natural language communication between man and machine, Communications of the ACM, Vol. 9, No. 1, pp. 36-45 (1966)

[飯田 03] 飯田弘之，松原仁: ゲーム情報学の動向, 情報処理, Vol. 44, No. 9, pp. 895-899 (2003)

[池田 10] 池田謙一, 唐沢穣, 工藤恵理子, 村本由紀子: 社会心理学, 有斐閣 (2010)

[池田 12] 池田心, Viennot Simon: モンテカルロ碁における多様な戦略の演出と形勢の制御〜接待碁 AI に向けて, ゲームプログラミングワークショップ 2012 論文集, No. 2012, Vol. 6, pp. 47-54 (2012)

[稲葉 12] 稲葉通将，鳥海不二夫，高橋健一: 人狼ゲームデータの統計的分析，ゲームプログラミングワークショップ 2012 論文集, No. 2012, Vol. 6, pp. 144-147 (2012)

[梶原 14] 梶原健吾, 鳥海不二夫, 大澤博隆, 片上大輔, 稲葉通将, 篠田孝祐, 西野順二, 大橋弘忠: 強化学習を用いた人狼における最適戦略の抽出, 情報処理学会第 76 回全国大会講演論文集, pp. 597-598 (2014)

[片上 15] 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 松原仁: 人狼知能プロジェクト, 人工知能学会誌, Vol. 30, No. 1, pp. 65-73 (2015)

[篠田 14] 篠田孝祐, 鳥海不二夫, 片上大輔, 大澤博隆, 稲葉通将: 汎用人工知能の標準問題としての人狼ゲーム, 人工知能学会全国大会論文集, No. 28, pp. 1-3 (2014)

[鈴木 05] 鈴木淳子: 調査的面接の技法（第 2 版）, ナカニシヤ出版 (2005)

[平田 15] 平田佑也, 稲葉通将, 高橋健一, 鳥海不二夫, 大澤博隆, 片上大輔, 篠田孝祐: プレイログから獲得した行動選択確率を用いた人狼ゲームのシミュレーション, 人工知能学会全国大会論文集, No. 29, pp. 1-4 (2015)

[藤井 13] 藤井叙人, 佐藤祐一, 中嶌洋輔, 若間弘典, 風井浩志, 片寄晴弘: 生物学的制約の導入による「人間らしい」振る舞いを伴うゲーム AI の自律的獲得, ゲームプログラミングワークショップ 2013 論文集, pp. 73-80 (2013)

[松原 12] 松原仁: ゲーム情報学の現在-なぜゲームの研究は日本で疎外されなくなったのか-, 情報処理, Vol. 53, No.2, pp. 102-106 (2012)

[山岸 01] 山岸俊男: 社会心理学キーワード, 有斐閣双書 (2001)

[横山 99] 横山真男, 青山一美, 菊池英明, 帆足啓一郎, 白井克彦: 人間型ロボットの対話インタフェースにおける発話交替時の非言語情報の制御, 情報処理学会論文誌, Vol. 40, No. 2, pp. 487-496 (1999)