

メッセージ付きジレンマゲームにおける コミュニケーションシステムの実験的検討

Experimental study of communication systems in dilemma games with messaging

井上 直紀[†], 森田 純哉[‡]

Naoki Inoue, Junya Morita

^{† ‡} 静岡大学

Shizuoka University

inoue.naoki.15@shizuoka.ac.jp[†], j-morita@inf.shizuoka.ac.jp[‡]

概要

本研究では、ジレンマ環境での新規なコミュニケーションシステムの成立と変化について、ゲームを用いて実験的に検討した。ゲームでは、利益を独占する行動と二者間でやり取りされるメッセージの曖昧さに相互関係があると想定された。ゲームを用いた予備的な実験の結果から、曖昧な意味付けを含むコミュニケーションの生起の可能性が示された。その上で集団実験を行い、生起したコミュニケーションに多様性があることが確認された。

キーワード：実験記号論, ジレンマ, コミュニケーションシステム

1. はじめに

近年の情報化やグローバル化が進んだ社会において、情報の信頼性が著しく低下し問題視されている。例えば Web, 特に SNS における虚偽情報の発信と急速かつ大規模な拡散はフェイクニュースと呼ばれ、社会問題となっている [1]。このような問題を解決する糸口をつかむには原因を原理的にとらえる必要がある。つまり、そのような虚偽や欺瞞を含むコミュニケーションの意図の構造を明確化し、その社会的役割を探索することで問題解決につながると考えられる。

虚偽や欺瞞のコミュニケーションに注目した研究として Byrne らによって提唱された「マキャベリの知性仮説」がある。人間は社会生活において、協調のコミュニケーションの中に、虚偽や欺瞞のような相手を裏切るコミュニケーションを巧妙に交えることで、自己の利益を高めようと振る舞う。このようなコミュニケーションが求められる集団環境への適応とその相互作用によって、人間の知能が高度に進化したとする仮説がマキャベリの知性仮説である [2]。マキャベリの知性仮説に基づくと、問題視されている虚偽や欺瞞が入り混じるコミュニケーションは必ずしも悪ではなく、

むしろコミュニケーションシステムを複雑にすることで、人類の進化を促進している可能性もあると考えられる。

コミュニケーションの生起や進化に関する研究は、認知科学分野において盛んに行われている [3]。その中で Galantucci は実験記号論と呼ばれる手法を提唱した。実験記号論では、実験参加者に通常のコミュニケーションが制限された実験室実験環境下でビデオゲームを用いた協調的な課題を課すことで、二者間の相互作用により言語のような図形と意味の対応づけのシステム（以下、「コミュニケーションシステム」という）が新規に成立することが確認された [4, 5]。

Galantucci の研究を受け、金野らは Galantucci の実験環境をより制限することで分析を容易にする実験課題を設計した。実験では 2 × 2 マスのマップ上からペアがそれぞれ 1 回の移動によって、共通の場所で出会う協調ゲームを課題とした。初期状態において、ペアに参加するプレイヤーは互いの位置を知ることができず、移動に先立って、5 種類の図形を二つ組み合わせてメッセージとして交換することで、共通の移動先を決定することを求められた。メッセージに使用される図形は、実験開始の当初は意味が付与されておらず、課題を繰り返すことで徐々に意味づけが行われ、新規な言語的コミュニケーションシステムとして成立することが確認された [6]。

Galantucci や金野らが扱う課題では、実験参加者に対するゲーム内での目標は共通の課題が与えられていた。よって実験参加者の間には単純で協調的なコミュニケーションシステムしか生じなかった。このようなコミュニケーションシステムは昆虫などの社会における原始的なコミュニケーションのモデルとしては妥当性を持つだろう。しかしながら、現代の人間の社会では時として協調と裏切りの選択を迫られるような複雑なコミュニケーションも行われる。そのため、人間社

会のモデルという観点では十分とは言えない。

協調と裏切りの選択が求められる課題はゲーム理論分野においてジレンマゲームと呼ばれ、検討されてきた。その一種であるスタグハントゲームは、利得支配である協調行動とリスク支配である裏切り行動の二つの純粋戦略ナッシュ均衡が存在するジレンマである。Duguidらは、チンパンジーと人間を対象に通常の直接的なコミュニケーションが制限された環境下で、スタグハントゲーム課題を繰り返し課す実験を行った[7]。その結果、チンパンジーは相手の行動選択の不確実性を克服できず、裏切り行動に陥り小さな利益しか得られなかった。一方で、人間はアイコンタクトなどのコミュニケーションによって信頼関係を構築し、協力しつづけることで大きな利益を得ることが可能であったことが示された。これによりジレンマ環境において、コミュニケーションがプレイヤーの行動選択に作用する可能性が示された。

このような関連研究を踏まえ、本研究では裏切りのコミュニケーションが生起・進化する環境を再現する実験室実験を行うことで、裏切りのコミュニケーションが、コミュニケーションシステムの進化にどのように作用するのか検討する。実験では、通常の言語コミュニケーションが制限された環境にジレンマ状況を与えることで、参加者同士の相互作用の中にどのようなコミュニケーションが成立・変化するか探索する。具体的には意味が曖昧なコミュニケーションの生起やそれを利用した裏切りの意図の隠蔽、そして裏切りによるコミュニケーションの変化を検討する。

2. 実験課題

上記の検討を行うために課題とその実行環境を設計する。本実験では金野らの研究を踏襲し、実験課題としてゲームを用いた。ゲームは複数ラウンドから構成され、各ラウンドは「部屋割り当て」、「メッセージ交換」、「移動」の3フェーズから構成される(図1)。本研究では金野らの研究を拡張し、マップを3×3に拡大し、メッセージとして使用できる図形の数を10種類に増やした。さらにジレンマ状況を構築するため、2名のプレイヤーに加え、「報酬」を新たにマップに加えた。ゲーム中、各プレイヤーは隣の部屋に存在するプレイヤー、あるいは報酬を視認することが可能であった。両プレイヤーのメッセージ交換・部屋移動が終わると次のラウンドがはじまり、2名のプレイヤーのいずれかが報酬の配置された部屋に移動すると各プレイヤーおよび報酬はマップ上に再割り当てされ、次のラウンドが始まる。このような状況の中で、報酬の

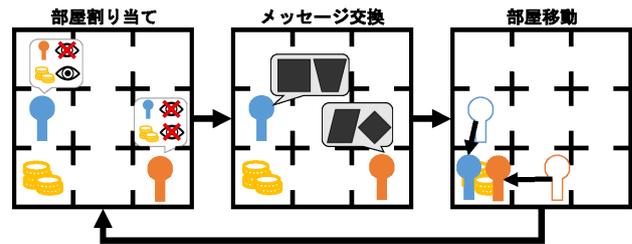


図1: 実験課題

独占（報酬を一人で取得する）と共有（報酬を二人が同時に取得する）のジレンマが生じる設計を目指す。

また、コミュニケーションシステムを複雑にするには、その基礎となるコミュニケーションシステムを構築させる必要がある。そのため実験では、実験参加者にジレンマ環境を与える前に、共有でしか得点が得られないゲームを設けたあと、ジレンマ要素を含む環境に移行させた。

3. 実験1: 事例研究

3.1 方法

本研究では、まず事例研究として少人数を対象とした実験を行った[8]。実験では、独占によって得られる得点 (m :monopoly) と共有によって得られる得点 (s :sharing) の比率が実験参加者のパフォーマンスに大きな影響を与えると考えられた。よって実験1では実験参加者に m が s の2倍であると教示した $m > s$ 条件と m と s が等しいと教示した $m = s$ 条件、 m が s の $1/2$ であると教示した $m < s$ 条件を設定した。なお、いずれの条件においても報酬を取らなかったラウンドは失敗 (f :failure) とし、 f は0ポイントとした。

上記の利得構造はゲーム理論におけるジレンマゲームを参考に配分した。しかしながら、本実験課題とゲーム理論におけるジレンマゲームでは行動を選択し結果を得る前にメッセージ交換フェーズが設けられているか否かという点で大きく異なる。そのような過程が存在することで実験参加者には、自分のメッセージに意味づけし、相手のメッセージの意味を解釈する認知的コストが課される。さらに本研究では制限時間を課してゲームを行うため、メッセージ交換や移動にかかる時間もコストとして考えられる。つまり、本研究の実験課題では、実験参加者の思考や残り時間などの要素によって利得構造が動的に変化する。よって、本実験課題における利得構造を利得表に一意に落とし込むことは困難である。一方で、メッセージ交換フェーズが存在することによってコミュニケーションシステムの形成と進化の観察、分析を実現することが可能に

なっている。

3.2 結果と考察

3.2.1 行動分析

裏切りのコミュニケーションの検討を行うには、まずそのようなコミュニケーションがどのような条件下で生起するのか検討する必要がある。

図2は、各条件においてペアを組んだ二人のプレイヤー (p1,p2) が協調フェーズとジレンマフェーズで報酬を取得した累積回数の推移を示している。図2aから、 $m > s$ 条件のジレンマフェーズにおいて他条件に比べプレイヤー間の報酬を取得した回数に大きな差が生じていることが確認される。これはこのペアにおけるプレイヤー間の関係性が、搾取をする側（報酬を独占する側）とされる側に分かれていたと解釈できる。この結果を支持する回答は、実験後のアンケートにおいても得られている。

これにより、ゲーム内における実験参加者のパフォーマンスが、実験で設定されるパラメータによって変化する可能性が示唆された。

3.2.2 メッセージ分析

また、裏切りのコミュニケーションは、複数の意味で解釈できるような曖昧な言語表現を利用して行われると考えられる。協調フェーズで成立したコミュニケーションシステムがそのようなコミュニケーションシステムに変化したか検討を行うために各プレイヤーのメッセージログデータについても分析を行った。

本研究では生起しうるコミュニケーションシステムの組み合わせを10種類の図形を9部屋+1未使用から選ぶ順列(362880通り)と仮定して、実際の行動データ(実際の行き先)と各条件・各プレイヤーの各コミュニケーションシステムを用いるモデル(あるコミュニケーションシステムを用いたときメッセージから想定される行き先)との一致率を求めた。この一致率をウィンドウ幅を10ラウンド、ストライド1ラウンドで分析した。その可視化の結果が、図3のヒートマップである。x軸は前述したウィンドウごとにまとめて分割された各ラウンドでとったパフォーマンスである。y軸は本研究において仮定したコミュニケーションシステムのすべての組み合わせのゲーム全体での平均一致率を上位から順に並べたものである。この図では一致率が1に近づくほど白く明るく、0に近

づくほど黒く暗く着色されている。つまり、各ヒートマップ内で明るい行があればその行が一致率の高いコミュニケーションシステムであり、明るい列があれば、多くのコミュニケーションシステムに当てはまるようなパフォーマンスであったという意味となる。

以上の前提において図3bを見ると、実験中盤から終盤にかけて一致率が一定程度収束している一方で、最終盤においても一致率が高い意味付けは複数確認でき、意味が一意に定まっていなかったことがわかる。これにより、本実験課題においてコミュニケーションシステムの意味付けの曖昧性が生じていた可能性が考えられる[9]。裏切りのコミュニケーションの成立には、虚偽や欺瞞を含む情報を発話した後に実際に行動する必要がある。つまり、実験における実験参加者のパフォーマンスとコミュニケーションシステムの関係性を確認することで実験参加者がどのようなコミュニケーションを行っていたか明らかにできると考えられる。

図4は行動とコミュニケーションシステムがそれぞれどのように関連しているか可視化したグラフである。x軸はラウンド数、y軸はジレンマフェーズにおける各条件におけるプレイヤー間の報酬取得数の差と図3で検討された一致率のスライド毎の標準偏差の推移を示している。本分析で用いた標準偏差は下式から求めることができる。

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

本分析において n はヒートマップの行数、 i は一致率の順位、 x_i はある推定されるコミュニケーションシステムにおける一致率、 \bar{x} はヒートマップ一列の平均である。

図において、報酬取得数の差が大きいとそのペアは裏切り行動を主に選択し、差が小さいと協調行動を主に選択していたと解釈できる。また仮に、本実験内でコミュニケーションシステムの意味づけが明確に行われているとすれば、実際に使われている意味付けルールの一一致率は非常に高く、使われていないものの一一致率は低くなるため標準偏差は高くなると予想される。逆に、コミュニケーションシステムの意味づけが曖昧であれば、一致率は一律の値になるため標準偏差は低くなると考えられる。

以上を踏まえて、図4aを見ると、序盤から中盤にかけて報酬取得数の差が開き、中盤から終盤にかけて一致率の標準偏差が下降していることが確認できる。よって $m > s$ 条件ではジレンマフェーズにおいて、意

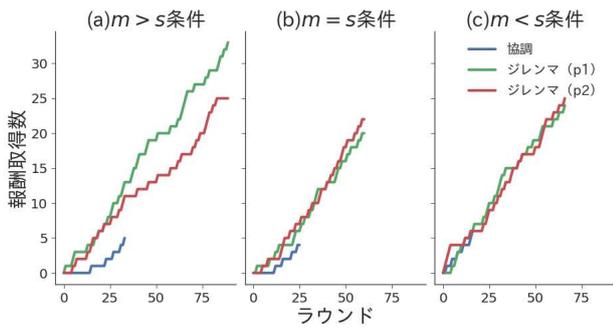


図 2: 報酬取得数の推移

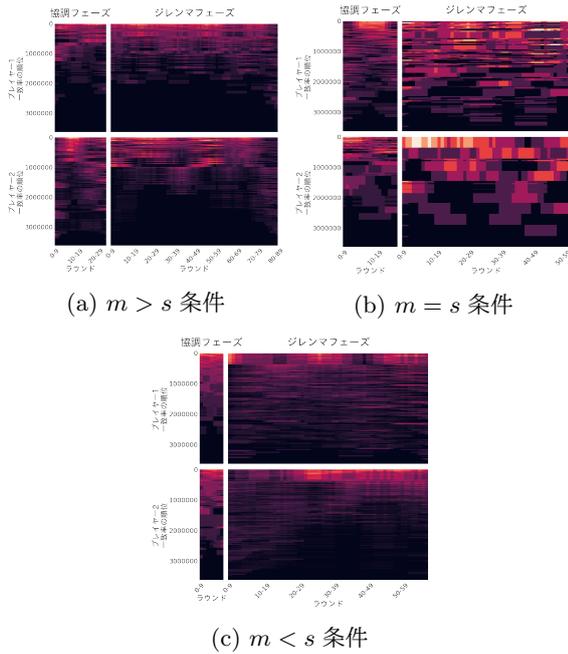


図 3: 各プレイヤーにおける各コミュニケーションシステムと実際の行動の一致率

味付けが曖昧な、虚偽や欺瞞の多いコミュニケーションが行われていたと考えられる。一方で図 4c では、報酬取得数の差は終始 0 付近で横ばい一方で、序盤に標準偏差が下降した後に中盤にかけて上昇し、終盤まで横ばいに推移している。このことからジレンマフェーズの $m > s$ 条件では、最初は意図せず曖昧な意味づけを行っていたが、徐々に意味付けが明確で安定したコミュニケーションに落ち着いたと考えられる。

以上のように、実験 1 ではペアごとに m, s, f を設定し、個別のデータに注目して検討を行った。しかしながら本実験は実験対象が 3 ペア 6 人と少ないため、実験結果が実験課題や条件ではなく、実験参加者の特性に依存していた可能性がある。そのため、パフォーマンスやメッセージについての分析手法の信頼性には疑問が残る。

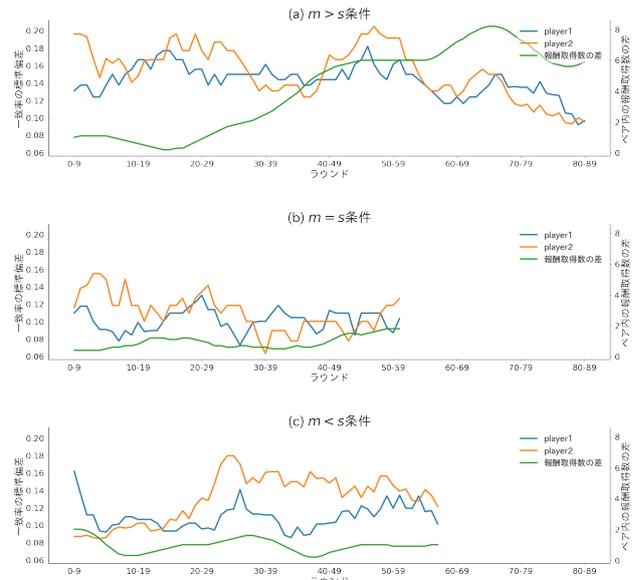


図 4: 報酬取得数の差と一致率の標準偏差の推移

4. 実験 2: 集団実験

実験 1 では、プレイヤーごとのデータに注目したミクロな視点での検討が行われた。その結果を踏まえた上で、より多くのペアを対象とした実験を行い、マクロな視点で大局的に分析することで、より一般的な知見が得られ、検討できると考えられる。また、実験 1 であげられた分析手法の信頼性の課題に対し、集団実験で得られる結果から、どのような分析手法が妥当であるか検討することを目指す。本実験に類似する集団を対象にした研究として、協調ゲームに関する小嶋らの研究がある [10]。小嶋らの研究では既に実験が実施されているため、本研究ではその実験手順を踏襲する。

4.1 方法

実験は 17 ペア、34 人を対象に実施した。実験参加者は静岡大学の学部 3 年生であった。実験課題は概ね実験 1 と同一の環境で実施した。システムを集団実験へ対応させる際、協調ゲームとジレンマゲームを共通のペアで行うが、パートナーは互いに誰かわからない匿名の状態でのマッチングする設定とした。また実験 1 ではペアごとに m, s, f を設定したが、実験参加者の特性への依存の可能性が指摘された。そのため本実験では結果の差異が個人に依存するものなのか明らかにするために実験 1 における中間にあたる $m = s$ 条件に統一して実施した。そのような状況下で、協調ゲームとジレンマゲームを合わせて 1 時間半かけて実験を実施した (表 1)。

表 1: 実験の流れ

時間 (m)	フェーズ
15	協調ゲームの説明
30	協調ゲームの実施
5	ジレンマゲームの説明
30	ジレンマゲームの実施
10	アンケート
90	所要時間合計

4.2 結果と考察

各プレイヤーの累計報酬取得数の推移を図5に示す¹。図中で同じ色で示された線は同じペアのプレイヤーを示している。ペア内においてより報酬を取得したプレイヤーは実線、より報酬を取得しなかったプレイヤーは点線で示している。グラフ中で実線のみが示されているペアは、プレイヤー間の報酬取得がラウンドを通して完全に一致したことを示している。ペア3とペア6のように、プレイヤー間で報酬取得が完全に一致したペアが存在する一方、ペア2のように、プレイヤー間で報酬取得に大きな差が生じたペアも存在した。実験2における報酬の条件は $m = s$ に統一されている。そのため、報酬取得数に差が生じた原因は、ゲームにおける報酬設定ではなく、ペア内でのメッセージの性質によると考えられる。今後、図3や図4で示した分析を実験2においても行うことで、メッセージの曖昧さと、コミュニケーションにおける欺瞞や騙しの発生がどのように相互に関係しながら変化するかを検討できると考えている。

5. まとめ

本研究で実験記号論に基づき、メッセージ付きジレンマゲームを用いることで、コミュニケーションシステムが成立し、変化する過程を観察した。実験1で得られた行動ログデータおよびメッセージログデータの分析やアンケートの回答により、実験内で新規に成立した言語的コミュニケーションシステムに曖昧な意味づけが生じていた可能性が示唆された。さらに、このような結果を踏まえ、集団を対象とした実験を実施した。集団実験の結果の分析によって、メッセージ付きジレンマゲームにおけるコミュニケーションシステムの分析として、よりマクロな視点で分析が行われることが期待される。

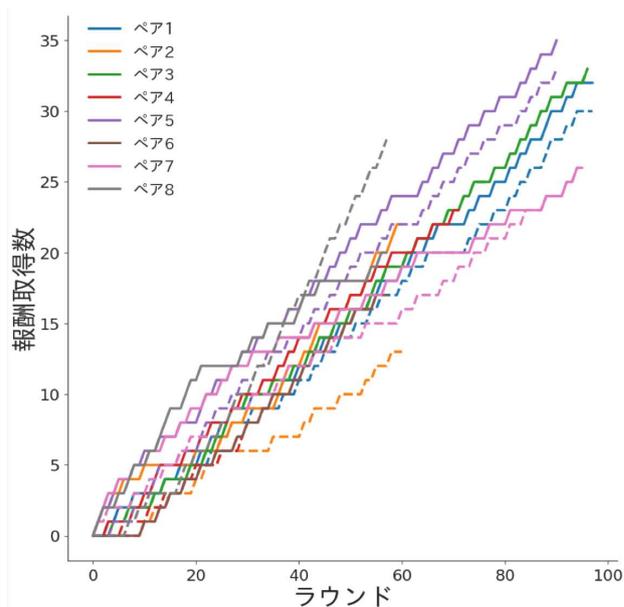


図 5: 各プレイヤーの報酬取得数の推移

文献

- [1] 笹原和俊. フェイクニュースを科学する: 拡散するデマ、陰謀論、プロパガンダのしくみ. DOJIN 選書. 化学同人, 2018.
- [2] Richard W Byrne. *The thinking ape: Evolutionary origins of intelligence*. Oxford University Press on Demand, 1995.
- [3] Thomas C Scott-Phillips and Simon Kirby. Language evolution in the laboratory. *Trends in cognitive sciences*, Vol. 14, No. 9, pp. 411–417, 2010.
- [4] Bruno Galantucci. An experimental study of the emergence of human communication systems. *Cognitive science*, Vol. 29, No. 5, pp. 737–767, 2005.
- [5] Bruno Galantucci. Experimental semiotics: A new approach for studying communication as a form of joint action. *Topics in Cognitive Science*, Vol. 1, No. 2, pp. 393–410, 2009.
- [6] 金野武司, 森田純哉, 橋本敬. 言語的コミュニケーションシステムの創発に関する実験的アプローチ. 計測と制御, Vol. 53, No. 9, pp. 801–807, 2016.
- [7] Shona Duguid, Emily Wyman, Anke F Bullinger, Katharina Herfurth-Majstorovic, and Michael Tomasello. Coordination strategies of chimpanzees and human children in a Stag Hunt game. In *Biological sciences*, Vol. 7, p. 1796, 2014.
- [8] 井上直紀, 森田純哉. 騙しの生起条件を探るメッセージ付きジレンマゲームの設計. HCG シンポジウム, 2018.
- [9] 井上直紀, 森田純哉. メッセージ付きジレンマゲームにおけるコミュニケーションシステムの成立と変化. 人工知能学会, 2019.
- [10] 小嶋暁, 紅林優友, 森田純哉. 新規なコミュニケーションシステムの形成に及ぼす自閉傾向の影響. 日本認知科学会, pp. 174–183, 2018.

¹システムのトラブルにより 8 ペアのための結果である。