

# 繰り返し囚人のジレンマゲームにおける 表情センシングを活用した人の感情と行動の分析

小菅 雷太郎<sup>1,a)</sup> 若林 直希<sup>2</sup> 山本 仁志<sup>3</sup> 秋山 英三<sup>4</sup> 栗原 聡<sup>1</sup>

**概要:** 社会シミュレーションにおけるエージェントは人間の行動を反映した振舞いが求められる。人間のノンバーバルな要素を行動モデルに組み込むために、我々は社会的ジレンマにおける人間の表情に着目した。本研究では、1対1繰り返し囚人のジレンマをプレイする人間の感情を、表情センシングにより推定し選択行動に対する感情および結果の観測に対する感情を調査した。また、表情センシングによる感情推定の結果と社会的心理指標とを比較した。この実験により、ジレンマ状態における表情に現れる感情が、協力行動時と非協力行動時で異なることがわかった。

**キーワード:** 実験社会心理学, 行動経済学, 表情分析

## 1. はじめに

社会シミュレーションでは社会現象をモデル化することで実際の社会現象を再現し、これから実施する施策の効果を推定することができる。社会シミュレーションの精度向上のためには、各エージェントの動きが実際の人の行動に近づく必要がある。人の行動は、ルールベースで決まるものではなく、様々な要素が反映されている。したがって、エージェントのモデル構築のために人の行動を規定する要因を抽出することは重要な課題である。

これまでの社会シミュレーションのエージェントの行動はゲーム理論をベースに人の合理性を仮定していた。一方で、人間は必ずしも自己最適を求めず、相手との関係を気にした行動選択をすることがある。このような人間の合理的でない部分をエージェントの行動モデルに組み込むことが提案されている。

Fuら [1]の研究では、空間型囚人のジレンマを用いた裏切りの誘惑と受容という2つのパラメータを元にしたシミュレーションで、感情伝達が個人の戦略に影響を与えることを示した。Phelpら [2]の研究では、繰り返し囚人の

ジレンマを用いて、社会的ジレンマの状況におけるプロンプトにより性格付けされたLLMエージェントの協力行動を評価した。

このように言語ベースの情報を用いた研究が多い一方で、非言語情報を用いた研究は少ない。本研究では、人間の限定合理性を非言語情報から抽出する。繰り返し囚人のジレンマの枠組みを用いて、社会的ジレンマ状態を仮想的に作り出し、行動選択中と結果確認中の表情センシングによる感情推定値を出力する。これにより、ジレンマ状態における行動選択中や結果確認中の表情に現れる感情の傾向をみる。

## 2. 背景

ここでは、本研究に用いた繰り返し囚人のジレンマと表情分析、質問紙調査で用いた一般的信頼について説明した後、関連研究について述べる。

### 2.1 繰り返し囚人のジレンマ

囚人のジレンマは、ゲーム理論における非協力ゲームの一つである。このゲームでは、相手の行動を知らずにCかDの2つの選択肢の中から一つを選択し、相手の選択との組み合わせにより得られる利得が変化する。この関係を表1に示す。表1の数字はそれぞれの行動の組に対する利得(P1の利得, P2の利得)である。

繰り返し囚人のジレンマとは、囚人のジレンマが繰り返し行われるゲームである。繰り返し囚人のジレンマゲームにおいては、直前までの結果を考慮しながら戦略を決定す

<sup>1</sup> 慶應義塾大学理工学部  
Faculty of Science and Technology, Keio University  
<sup>2</sup> 慶應義塾大学院理工学研究科  
Faculty of Science and Technology, Keio University  
<sup>3</sup> 立正大学経営学部  
Faculty of Business Administration, Risho University  
<sup>4</sup> 筑波大学システム情報系  
Faculty of Engineering, Information and Systems, University of Tsukuba  
a) rai7835gm@keio.jp

る。一般に、協力をとり続けることが最適戦略であるが、相手の行動に合わせた多様な戦略が提唱されている。

表 1 囚人のジレンマの利得表

P1 \ P2	C	D
C	(10, 10)	(-20, 30)
D	(30, -20)	(-10, -10)

## 2.2 表情分析

ここでは、本研究で用いた顔画像から表情を抽出し感情を推定する方法について説明する。本研究では、Py-Feat[3]を用いて、表情画像から感情を推定する。Py-FeatではAction Unit (AU) と呼ばれる「表情における筋肉の一部分の動き」を分析し、AUと感情カテゴリとの対応付けにより感情の推定が行われている。感情推定値は怒り、嫌悪、恐れ、喜び、悲しみ、驚き、無表情の7種類の感情である確率である。

## 2.3 一般的信頼

一般的信頼感とは、「相手についての情報がない場合の相手の信頼性に対する”デフォルト値”」として定義される[4]。一般的信頼では、他者一般の信頼性に対する信念としての信頼を測定することができる。垣内、山岸[5]の研究では、囚人のジレンマを用いた実験で一般的信頼の高い人間が一般的信頼の低い人間に比べ相手の行動により敏感に反応していることを示している。一般的信頼は一般的信頼下位尺度に対し、5件法で回答を求めることにより測定する。本実験の事後質問紙調査では、一般的信頼下位尺度のうち「ほとんどの人は信用できる」、「たいていの人は人から信頼された場合、同じように相手を信頼する」、「ほとんどの人は他人を信頼している」、「ほとんどの人は基本的に正直である」を用いた。

## 3. 関連研究

囚人のジレンマとセンシング技術を用いた研究は多く存在する。中でも、視線、脳波などのセンシング技術を用いた研究が多い。前田ら[6]は、1回限りの囚人のジレンマゲームにおいて視線計測を実施し、相互協力による結果と相互非協力による結果を並べて表示させた場合、協力者は、非協力者に比べて、相互協力による結果に注意を払うことを示したJahngら[7]は、EEGハイパースキャンを用いて、2人の繰り返し囚人のジレンマゲームを対面で実施する場合と非対面で実施する場合の脳波を測定した。この実験で対面時のほうが非対面時に比べて皮質全体の活動が活発になっていることが示された。

表情分析と囚人のジレンマの関連研究を以下に示す。Reedら[8]は1対1の囚人のジレンマについて、事前に協

力の約束をしたときの表情をセンシングし、その結果を後に実施する囚人のジレンマの協力・非協力の予測因子になるかということ調査した。この結果、約束時にデュシェンヌスマイルをする参加者が協力する可能性が高いことが示唆された。杉浦ら[9]の研究では、繰り返し囚人のジレンマゲームをプレイする被験者の表情計測を実施した。この研究では、表情解析ツールPy-Featを用いて表情シグナルを提示したときの被験者の表情を計測した。実験の結果、非協力的な人は、協力的な人に比べて、裏切られたときに怒り、嫌悪、悲しみ、驚きの表情が出やすいことがわかった。

本研究は、繰り返し囚人のジレンマゲームにおける行動選択および結果確認における人の表情センシングによる感情推定を実施し、その結果を社会心理指標との関係と比較する点で、以上の研究と異なるものである。社会的ジレンマ状態における人の表情の動きを抽出することで、合理的でない人の行動の要因の分析に貢献することができるという。

## 4. 実験

本研究の目的である、繰り返し囚人のジレンマをプレイするときの人の表情をセンシングするための実験を設計し、実施した。今回は、行動選択中の表情と結果確認中の表情に着目し、感情推定値（以下感情値）の平均を比較することで、繰り返し囚人のジレンマにおける人の感情を表情解析により調査する。本章では、実験の概要と使用したソフトウェア、事後質問紙調査について説明する。

### 4.1 実験の概要

本実験は、倫理審査委員会の承認を得て実施した（立正大学研究倫理委員会倫 05-1号）。実験は2023年11月30日、12月7日の2日間都内私立大学において実施され7名が参加した。実験参加にあたっては、実験の目的、実施方法、実験結果の利用方法について説明し、同意を得た。また、通常の謝金に加えゲームの最終得点に応じた追加報酬があることを伝達した。

実験室は、被験者が対戦相手や周囲を気にせず実験ができるように個室を用意した。これにより、被験者の表情が外環境に影響を受けることを最小限にした。被験者は、1対1の繰り返し囚人のジレンマゲームをプレイする。すべての被験者について繰り返しは20回であったが、被験者には繰り返し回数は決まっていないと教示した。対戦相手は人である。被験者同士のペアが決まらない場合にはダミーの対戦相手を用意した。

繰り返し囚人のジレンマの行動選択データの収集はノートPCを用いて行った。また、同時に表情センシングを行うために、ノートPCのカメラで被験者の顔を撮影した。実験終了後には、被験者に質問紙調査を実施した。

#### 4.2 オンライン繰り返し囚人のジレンマゲーム

被験者がPCを用いて繰り返し囚人のジレンマゲームを実施できるよう、oTree[10]を用いてゲームのソフトウェアを実装した。マウス操作でOption I (C) か Option J (D) を選択するように指示した。ゲーム画面は図1、図2の通りである。また、実験実施時にはホストPCにLinuxサーバーを立て、実験参加者が使用するPCと接続した。実験では以下の利得表に基づいて利得を計算した。実験に用いた利得表は表1を採用した。

#### 4.3 表情記録ソフトウェア

行動選択中と結果確認中の被験者の顔の動画をPC内蔵カメラを用いて30FPSで撮影した。撮影した画像をもとに、表情解析ツールPy-Featを用いて怒り、嫌悪、恐れ、喜び、悲しみ、驚きの6種類の感情である確率(感情値)を出力した。

#### 4.4 事後質問紙調査

事後質問では、被験者に対し一般的信頼を含む質問に対し、そう思う・ややそう思う・どちらともいえない・あまりそう思わない・そう思わないの5段階で回答を求めた。また、実験時に考えていたことや感じたことについて自由記述で回答を求めた。

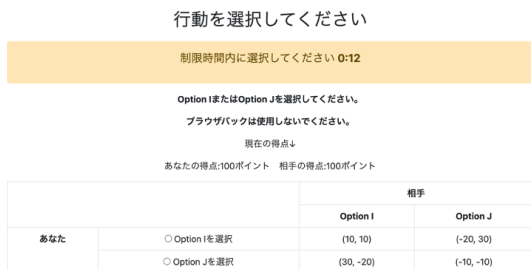


図1 ゲームの行動選択画面

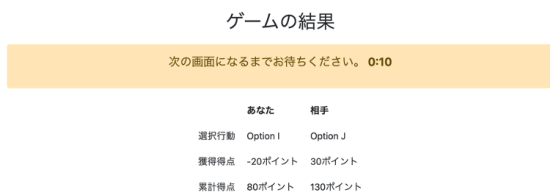


図2 ゲームの結果確認画面

### 5. 結果・考察

#### 5.1 行動選択中の感情値

まず、行動選択中における感情値の平均を協力(C)と非協力(D)で比較を図3に示す。6種類の感情の中で悲しみが最も感情値が高かった。また悲しみの感情値の平均は、CとDで差はないことがわかる。これは、実験参加者が行動を選択しているときに一貫して悲しみの感情が検出されていることを示している。悲しみ以外の感情値ではCを選ぶときにはDを選ぶときに比べ、怒り、驚きが高いことがわかる。

以上から、行動選択中には選択に関わらず一貫して悲しみの表情が現れており、Cを選択するときにはDを選択するときと比べ怒り、驚きの感情が高いことがわかった。

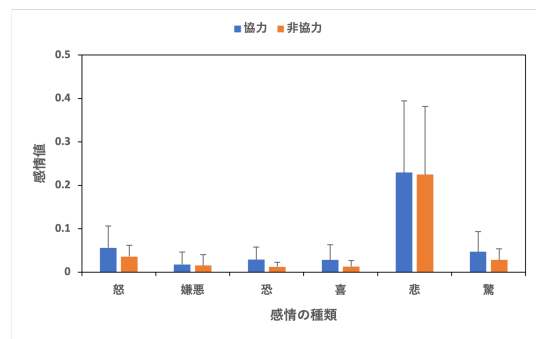


図3 行動選択中の感情値の平均

#### 5.2 結果確認中の感情値

次に、結果確認中における感情値の平均をゲームの結果によって比較した。ゲームの結果はCC, CD, DC, DDの4種類である。CCでは自分相手ともに点数が加算され、CDでは自分は減点され相手は加点される。DCでは自分は加点され相手は減点される。DDでは自分相手ともに減点される。

図4は結果確認中の感情値の平均を結果の種類別に示している。6感情のうち行動選択中と同様に悲しみが最も検出されていた。行動選択中と異なり、結果確認中においては悲しみの感情値の平均は結果の種類によって異なっていた。悲しみの感情値の平均をみると、 $DC > CC > DD > CD$ の順に大きい。この違いは、悲しみの感情値以外の感情値の影響を受けたものと考えられる。怒りと驚きの感情値についてみていく。怒りの感情値は $CD > CC > DD > DC$ の順に大きく、驚きの感情値は $CC > CD > DD > DC$ の順に大きい。このことから、怒りと驚きの感情値が結果の種類によって異なった分、結果の種類によって悲しみの感情値が異なることがいえる。

以上から、結果確認中には結果に関わらず悲しみの感情値が最も検出されていることがわかった。また、4つの結

果の中で、怒りの感情値は  $CD$ 、驚きの感情値は  $CC$  のときに最も高いことがわかった。

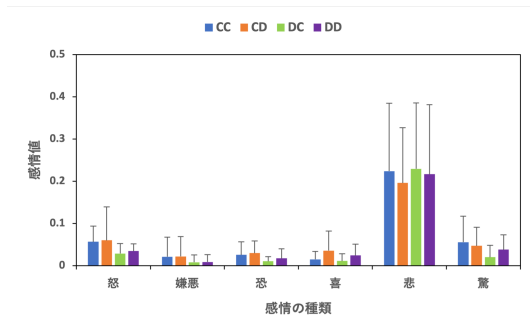


図 4 結果確認中の感情値の平均

### 5.3 一般的信頼による結果確認中の感情値の比較

実施した質問紙調査をもとに、一般的信頼による結果確認中の怒りの感情値の平均を比較する。一般的信頼は、一般的信頼下位尺度に対し、5 件法で回答を求めた。その結果をそう思う・ややそう思う・どちらともいえない・あまりそう思わない・そう思わないを 5・4・3・2・1 として数値化した。被験者 7 名を数値化した得点の中央値をもとに一般的信頼の高い群と低い群に分け、それぞれの結果確認中の怒りの感情値の平均を比較した。

図 5 は一般的信頼の高さによる結果を確認したときの怒りの感情値の違いを示している。一般的信頼の高い人の方が結果確認中において、全ての結果に対して怒りの感情値が高いことがわかった。また、一般的信頼の高い人の怒りの感情値は、 $CD > CC > DC > DD$  の順に大きいのに対し、一般的信頼の低い人の怒りの感情値は、 $CC > DD > CD > DC$  の順に大きいことがわかった。

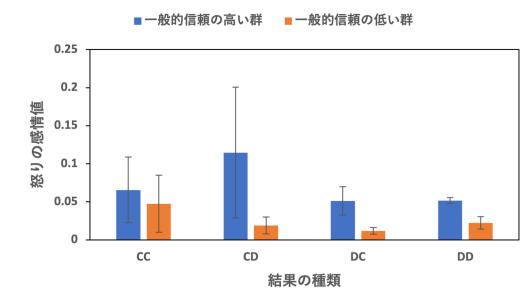


図 5 一般的信頼による結果観測時の感情値の比較

## 6. 結論・展望

本研究では、繰り返し囚人のジレンマゲームにおける人間の感情に着目し、表情センシングにより感情を推定することで、行動選択中や結果確認による感情の変化を調査した。そして、その結果の個人差を質問紙による社会的心

理指標を用いて検討した。実験の結果、協力を選択するときの表情が非協力を選択するときと比べて、怒り、驚きの表情が出やすいことがわかった。結果確認中においては、 $CD$  のときには怒りが、 $CC$  のときに驚きの感情が表情に現れている傾向が見られた。また、結果確認中において一般的信頼の高さによって怒りの感情値の違いがあることが示唆された。今後追加の実験を実施し、統計的な議論をしていきたい。

今後の展望としては、複数の表情解析モデルを用いて、表情からの感情推定の精度を高めていきたい。さらに、バイタルデータや視線データなどを組み合わせたマルチモーダルな実験の実施も検討する。以上の点を改善した上で、より多くの実験参加者を対象に実験を実施し、ジレンマ状態における感情や意思決定の過程を明らかにしたい。また、より実社会に近い状況として繰り返し囚人のジレンマの形態を対面に変更した実験を実施する。そして、得られた分析結果やモデルを用いてジレンマ状態にある人間の行動の要因を抽出し、エージェントのモデルに反映することで社会シミュレーションの精度向上に寄与したい。

**謝辞** 本研究は、国立研究開発法人情報通信研究機構の助成を受けたものである。

### 参考文献

- [1] Xiao Fu, Xuesong Liu, Yuhan Hu, Shaoyu Huang. *Emotion-based renewal strategies to promote cooperation in spatial prisoner's dilemma games*. Applied Mathematics and Computation, Volume 455, 2023.
- [2] Steve Phelps and Yvan I Russell. *Investigating emergent goal-like behaviour in large language models using experimental economics*, arXiv preprint arXiv:2305.07970, 2023.
- [3] Jin Hyun Cheong, Eshin Jolly, Tiankang Xie, Sophie Byrne, Matthew Kenney, Luke J. Chang. *Py-Feat: Python Facial Expression Analysis Toolbox*. arXiv:2104.03509, 2021.
- [4] 山岸俊男 小杉素子. 一般的信頼感と信頼性判断. 心理学研究, 69:349–357, 1998.
- [5] 垣内理希 山岸俊男. 一般的信頼と依存度選択型囚人のジレンマ. 社会心理学研究, 12(3):212–221, 1997.
- [6] Keisuke Maeda, Hirofumi Hashimoto, and Seiichiro Tanida. *Cooperators pay more attention to the outcome of mutual cooperation in the one-shot prisoner's dilemma game: Empirical evidence from an eye-tracking study*. Letters on Evolutionary Behavioral Science, 14:8–12, 2023.
- [7] Jaehwan Jahng, Jerald D. Kralik, Dong-Uk Hwang, and Jaeseung Jeong. *Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the prisoner's dilemma game*. NeuroImage, 157:263–274, 2017.
- [8] Lawrence Ian Reed, Katharine N. Zeglen, and Karen L. Schmidt. *Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma game*. Evolution and Human Behavior, 33(3):200–209, 2012.

- [9] 杉浦巧, 岸本大輝, 山田悠司, 秋山英三, 山本仁志, 栗原聡. マルチモーダルセンシングを用いたジレンマゲームにおける人の協調行動の分析. In 第 23 回データ指向構成マイニングとシミュレーション研究会 (SIG-DOCMAS)(WSSIT2023), 2023.
- [10] Daniel L. Chen, Martin Schonger, and Chris Wickens. oTree — an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.