

令和 4 年 6 月 2 日現在

機関番号：32643

研究種目：若手研究

研究期間：2018～2021

課題番号：18K13276

研究課題名（和文）協力の文脈における人間の行動規則の推定：計算論モデルからのアプローチ

研究課題名（英文）Estimating human behavioral rules in cooperation: a computational modeling approach

研究代表者

堀田 結孝 (Horita, Yutaka)

帝京大学・文学部・准教授

研究者番号：90725160

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：本研究の目的は、経済ゲームを用いた行動実験、実験データへの計算論モデルの当てはめ及び情報量基準に基づくモデル比較のアプローチを組み合わせ、人々の協力行動を説明する上で高い説明力を有するモデルを探ることであった。一連の実験研究の結果、同じ相手と繰り返し交換を行う状況と毎回相互作用相手が変わる状況いずれにおいても、相手の行動のみに反応する単純な互惠モデルよりも、自分と相手の前回の行動両者を考慮するモデルが多様な参加者の個人差を含む実験データにおいて高い予測力を有することが示唆された。

研究成果の学術的意義や社会的意義

協力行動の実験研究において、人々の行動規則を理解する上で互惠性の重要性が認識されている。一方、理論的には、協力の進化において単純な互惠戦略以外の有効性も指摘されてきた。本研究結果は、協力の文脈における人々の行動を理解する上で単純な互惠性のみでは不十分であることを示しているとともに、単一のモデルで協力の文脈における行動のパリエーションを効率的に予測できる可能性を示唆している。本研究結果は現実の人間の行動規則の素朴な理解に留まらず、協力の進化における理論モデルに対してもフィードバックをもたらすことが期待される。

研究成果の概要（英文）：The purpose of this study was to explore models with high explanatory power in explaining people's cooperative behavior by combining behavioral experiments using economic games, fitting computational models to experimental data, and model comparison approaches based on information criteria. The results of a series of experimental studies suggested that, in both situations of repeated interactions with the same person and situations in which the interactive partners change each time, a model that considers both the previous behavior of the participant and that of the other participant has higher predictive power in explaining experimental data containing individual differences among various participants than a simple reciprocity model that considers only previous behavior of the other participant.

研究分野：社会心理学

キーワード：互惠性 利他性 協力 罰 信頼 計算論モデル

### 1. 研究開始当初の背景

人は自己利益を犠牲にしても見知らぬ他者に協力することが、囚人のジレンマゲームをはじめとした経済ゲーム実験を通して頑健に示されてきた。なぜ人は協力するのかという問いに対して、心理学では公正感や愛他的動機など、向社会行動(pro-social behavior)の背後にある様々な心理メカニズムが記述されてきた。しかし、心理学で記述されてきた人の向社会性に関するモデルは抽象的な概念による表現が多く、行動の進化などの数理的な検証へ応用させることが困難であったといえる。それに対し、行動経済学や進化ゲーム理論においては、効用関数ないしは確率論モデルのかたちで利他行動を数理モデル化し、利他行動の進化の過程を説明する試みがなされている。

しかしその一方、理論モデルが現実の人間行動を反映していなければ意味がない。協力の進化を説明する理論モデルが現実の人間行動を反映しているかを実証研究で検討し、その結果を元に理論を再構築するプロセスが利他行動の進化を理解する上で重要な営みであると言える。

### 2. 研究の目的

本研究の目的は、経済ゲームを用いた行動実験、実験データへの計算論モデルの当てはめ及び情報量基準に基づくモデル比較のアプローチを統合し、協力の文脈における人々の行動を説明する上で高い説明力を有するモデルを探ることであった。

### 3. 研究の方法

囚人のジレンマゲームのような金銭的インセンティブを導入した経済ゲーム実験を行い、参加者の行動データを取得した。そして、実験データに協力行動を予測する様々なモデルを当てはめ、実験での参加者の行動パターンを予測する上で最適なモデルを情報量基準に基づくモデル選択を通して検討した。

### 4. 研究成果

#### (1)

これまでの多くの実証研究において、相手の協力に対して協力を返報する互惠性が、人々の利他行動を表す一般的な特徴として素朴に理解されている。しかし、理論的には応報戦略 (Tit-for-Tat: TFT) に代表される互惠的戦略は必ずしも常に進化的に安定な戦略となるわけではなく、行動や他者の行動に対する認識にエラーが生じる状況等においては強化学習の類型 (Win-stay lose shift: WSLS) が互惠的戦略よりも有利になり得ることも指摘されている。近年、公共財ゲーム (public goods game) のような複数の他者同士で資源の交換が行われる状況においては、強化学習のモデルが実証データをうまく説明できる可能性も指摘されている (Burton-Chellew, Nax, & West, 2015; Horita, Takezawa, Inukai, Kita, & Masuda, 2017)。

これらの背景を踏まえ、本研究では社会的相互作用の基礎となる二者間の交換に焦点を当て、実験における人々の行動パターンのバリエーションを探ることを試みた。参加者たちの間で金銭の分配を行う相互作用実験を実施した。相互作用のタイプとして、相手からの直接の返報がなされる直接互惠の状況 (direct reciprocity: 図 1A) と、相手からの直接の返報が見込めない一般交換の状況 (generalized reciprocity: 図 1B) の2つを設定した。更に、理論モデルで強化学習が有利に働く状況として前提とされているエラーが生じる状況 (相手が前回協力していても、“非協力した”と逆の情報がフィードバックされるなど)のもとで、参加者には他の参加者たちと相互作用を行ってもらった。これらの実験ゲームに参加してもらい、参加者の行動データを取得した。取得したデータに、互惠性や強化学習を反映させた計算論モデルを当てはめ、実験データに対して当てはまりの良いモデルを探った。

研究の結果、いずれの相互作用のタイプにおいても、相手の前回の行動のみに反応する単純な互惠モデルよりも、自身と相手の前回の行動両者を考慮するモデルが多様な参加者の個人差を含む実験データにおいて高い予測力を有することが示された。それぞれの状況での行動パターンを細かく見ると、直接互惠の状況では相手が前回単に協力ないしは非協力したかよりも、自分も相手も共に協力した相互協力及び共に非協力をした相互非協力の結果が次回の協力の確率の上昇及び下降に大きく影響しており、単純な WSLS のような行動パターンとも異なることが示唆された (図 2)。この傾向はむしろ、多くの人々の協力の動機が相互協力の達成の目標に基づ

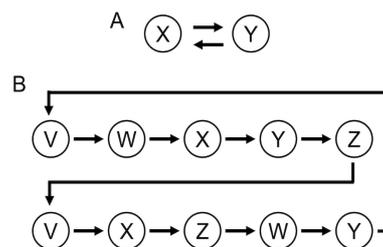


図 1: Horita (2020)の Figure 1 より

いているという社会心理学における議論 (Pruitt & Kimmel, 1977) とも共通している。ただし、一般交換の状況では相互協力及び相互非協力が次回の協力の確率に及ぼす効果が小さいことも示唆された。本研究では TFT や WLSL に代表されるモデルを対象としていたが、他のモデルも比較対象として、直接互恵及び一般交換における行動規則が単一のモデルのパラメータの違いとして説明できるものか、それぞれの状況での行動規則が質的に異なる行動モデルに基づくかをより深く探ることが今後の課題である。研究成果は PLoS One に掲載された (Horita, 2020)。

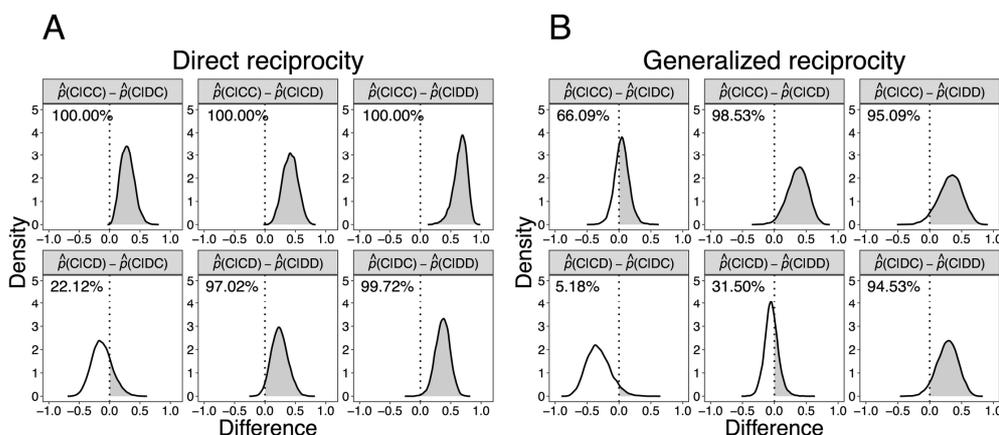


図 2: Horita (2020)の Figure 4 より

#### (引用文献)

- Burton-Chellew, M. N., Nax, H. H., & West, S. A. (2015). Payoff-based learning explains the decline in cooperation in public goods games. *Proceedings of Royal Society: B*, 282(1801):20142678. <https://doi.org/10.1098/rspb.2014.2678> PMID: 25589609
- Horita, Y., Takezawa, M., Inukai, K., Kita, T., & Masuda, N. (2017). Reinforcement learning accounts for moody conditional cooperation behavior: Experimental results. *Scientific Reports*, 7:39275. <https://doi.org/10.1038/srep39275> PMID: 28071646
- Horita, Y. (2020) Greater effects of mutual cooperation and defection on subsequent cooperation in direct reciprocity games than generalized reciprocity games: Behavioral experiments and analysis using multilevel models. *PLoS ONE*, 15(11): e0242607. <https://doi.org/10.1371/journal.pone.0242607>
- Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, 28, 363–392. <https://doi.org/10.1146/annurev.ps.28.020177.002051>

#### (2)

近年、自集団には協力する一方、他集団に対して非協力的に振る舞う“偏狭な利他性(parochial altruism)”がヒトの協力の進化と大規模な協力社会の形成において重要な役割を果たしている可能性について議論されており、偏狭な利他性から協力の進化を説明する理論モデルも提案されている。他集団に対する非協力傾向の背後には、他集団成員からの非協力的行動の確率を過度に見積もる傾向が関わっていることが予想される。本研究では近年注目されている偏狭な利他性に関する背景も踏まえ、他者に関する情報が不確実な状況における、他者の非協力傾向の推測及びそれによって引き起こされる行動を予測するモデルを検討した。

先制攻撃ゲーム (Simunovic et al., 2013) や社会的価値志向性 (van Lange et al., 1997) を測定する課題などを用いた実験を通して、上述の目的を検討した。相手の行動が明らかになる前に、相手に対して先に非協力的に振る舞う者あるいは他者との相互作用を避ける者がどの程度見られるか、また、それらの行動が生起する確率を予測する上で予測力の高いモデルを探った。経済ゲームでの非協力的行動と相関することが示されているパラノイア傾向 (他者が危害を加えようとしていると推測する傾向) (Raihani & Bell, 2018) にも注目し、他者の非協力的行動の見積もりと関連が予想される変数として検討の対象とした。

研究の結果、まず、非協力的な行動が非協力者本人の自己利益の損失につながる行動 (スパイト行動) であったとしても、“相手は自身に対して非協力的に振る舞ってくるだろう”と予測する者が多く見られ、他者の非協力的行動に対する過度な見積もりは一般的に見られる傾向であることが示唆された。また、相手の攻撃性や競争性の推測において、パラノイア傾向が強い予測力を有する変数である可能性が示唆された。更に、他者から非協力を受ける前に先に攻撃の手を打つ“先制攻撃”や他者との相互作用を避ける行動の生起には、相手の競争性を過度に推測する傾向が関わっていることがモデル選択を通じた解析の結果から示唆された。いくつか行った研究成果のうち、一つは *Frontiers in Psychology* に掲載されている (Horita, 2021)。

#### (引用文献)

- Horita, Y. (2021). Conjecturing harmful intent and preemptive strike in paranoia. *Frontier in Psychology*,

12:726081. doi: 10.3389/fpsyg.2021.726081

Simunovic, D., Mifune, N., & Yamagishi, T. (2013). Preemptive strike: an experimental study of fear-based aggression. *Journal of Experimental Social Psychology, 49*, 1120–1123. doi: 10.1016/j.jesp.2013.08.003

Raihani, N. J., & Bell, V. (2018). Conflict and cooperation in paranoia: a large-scale behavioural experiment. *Psychological Medicine, 48*, 1523–1531. doi: 10.1017/S0033291717003075

Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*, 733–746. doi: 10.1037/0022-3514.73.4.733

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 3件／うち国際共著 0件／うちオープンアクセス 3件）

1. 著者名 Horita, Y.	4. 巻 15
2. 論文標題 Greater effects of mutual cooperation and defection on subsequent cooperation in direct reciprocity games than generalized reciprocity games: Behavioral experiments and analysis using multilevel models.	5. 発行年 2020年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0242607
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0242607	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 堀田結孝	4. 巻 8
2. 論文標題 教育方法の評価へのベイズ統計モデリングの応用	5. 発行年 2020年
3. 雑誌名 帝京大学学生カウンセリング研究	6. 最初と最後の頁 7-17
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 堀田結孝	4. 巻 53
2. 論文標題 人間の協力行動に関する実験ゲーム研究と組織管理への応用可能性	5. 発行年 2019年
3. 雑誌名 組織科学	6. 最初と最後の頁 33-42
掲載論文のDOI（デジタルオブジェクト識別子） 10.11207/soshikikagaku.53.2_33	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Horita, Y., & Takezawa, M.	4. 巻 9:1921
2. 論文標題 Cultural differences in strength of conformity explained through pathogen stress: a statistical test using hierarchical Bayesian estimation.	5. 発行年 2018年
3. 雑誌名 Frontiers in Psychology	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1098/rspb.2011.0060	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Horita, Y.	4. 巻 12:726081
2. 論文標題 Conjecturing harmful intent and preemptive strike in paranoia.	5. 発行年 2021年
3. 雑誌名 Frontiers in Psychology	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/fpsyg.2021.726081	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

[学会発表] 計5件(うち招待講演 0件/うち国際学会 0件)

1. 発表者名 土田修平・中島彩花・堀田結孝・竹澤正哲
2. 発表標題 強化学習モデルを用いた協力行動の個人差の検討
3. 学会等名 日本社会心理学会第59回大会
4. 発表年 2018年

1. 発表者名 堀田結孝
2. 発表標題 公募シンポジウム「文化差の起源とダイナミクス：文化進化論からの視座」(話題提供)
3. 学会等名 日本心理学会第82回大会
4. 発表年 2018年

1. 発表者名 堀田結孝・竹澤正哲
2. 発表標題 伝染病の蔓延と集団主義傾向の関連の再検討：階層ベイズモデリングによる検証
3. 学会等名 日本人間行動進化学会第11回大会
4. 発表年 2018年

1. 発表者名 堀田結孝・竹澤正哲・犬飼佳吾・喜多敏正・増田直紀
2. 発表標題 強化学習による（気まぐれな）条件付き協力行動の説明
3. 学会等名 第22回実験社会科学カンファレンス
4. 発表年 2018年

1. 発表者名 堀田結孝
2. 発表標題 パルノイアと危害的意図の推測:先制攻撃ゲームを用いた検討
3. 学会等名 日本社会心理学会第62回大会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関