

令和元年6月3日現在

機関番号：12601

研究種目：基盤研究(B) (一般)

研究期間：2016～2018

課題番号：16H03120

研究課題名(和文) 高次報酬懲罰系による社会的ジレンマの解決

研究課題名(英文) High-order reward punishment systems solves social dilemmas

研究代表者

鳥海 不二夫 (Toriumi, Fujio)

東京大学・大学院工学系研究科(工学部)・准教授

研究者番号：30377775

交付決定額(研究期間全体)：(直接経費) 13,000,000円

研究成果の概要(和文)：高次報酬懲罰系における社会的ジレンマの理論モデルの解析のための私的評価環境における解析手法を開発した。

高次報酬懲罰系社会構造の理論的フレームワークの構築で、一次報酬と二次報酬にリンケージがなくても、二次報酬行動が可視化されれば協調的社会が実現することを明らかにした。この結果は、被験者実験によって、他者のサンクション行動に対する信念が人々の協力行動に影響を与えることを示しことと整合性がある。また、二次の罰則は一次の罰則より高い協力率を維持することなどを明らかにした。

研究成果の学術的意義や社会的意義

高次報酬懲罰系における社会的ジレンマについて、理論、実験、シミュレーションの3点からアプローチし、社会的ジレンマの性質およびその解消法について明らかにできた。

本研究結果は社会的ジレンマが存在する社会システムの構築の際に、その設計に寄与すると考えられる。たとえば、公共財としての知識共有システムの利用促進などへの応用が期待される。

研究成果の概要(英文)：We developed a theoretical approach of private assessment for analyzing social dilemmas in higher-order reward and punishment systems. We also clarified that a cooperative society would emerge when players' actions of second-order rewards were visualized even if there is no linkage between first-order and second-order rewards. This result is consistent with the results of our experiments which show that beliefs on sanctioning behaviors by the others have a positive influence on players' cooperative behaviors. Moreover, our experiment shows that the second-order punishment maintains higher cooperation than the first-order punishment.

研究分野：計算社会科学

キーワード：社会的ジレンマ 公共財ゲーム マルチエージェントシミュレーション 被験者実験

1. 研究開始当初の背景

現実社会において、公共財に対するフリーライダーの存在に起因する問題は数多く存在し、その解決は重要な課題である。公共財ゲームの枠組みにおいてフリーライドを抑制する仕組みとして報酬・懲罰制度が存在する。被験者実験では Fehr and Gächter (2002) が利他的懲罰の存在を明らかにして以来、多くの実験研究において、協調を維持するために、人間は自発的な報酬や懲罰を行っていることを示してきた。一方、解析・シミュレーションの側からは、对人的懲罰と制度的懲罰の比較(Sigmund(2010)) や懲罰の報酬制度に対する優位性(Hilbe(2010)) など、主に公共財ゲームにおける協調問題の枠組みで分析するアプローチが主流である。しかし、これらの研究ではいわゆる協調問題という1次のフリーライダーの解決に焦点を当てており、2次以上(高次)のフリーライダー問題に言及したものはほとんどない。ここで、2次のフリーライダー問題とは、フリーライダーへの懲罰やフリーライドしなかった人への報酬といった制度の維持にかかるコストの拠出をさぼるフリーライダーが現れる問題である。この2次の問題は Axelrod(1986)の古典的研究において言及されている重要問題であるが、適切な解決策が発見されず現在にいたっている。その中で申請者は予備的研究において、高次の報酬・懲罰制度を組み込むことによる解決をシミュレーション研究で示し、ソーシャルメディアへの応用を試みた(研究業績[14][30])。なお、ここで、高次の報酬懲罰制度とは、2次以降のフリーライダーへの報酬懲罰が存在する制度を指す。予備的研究において高次の報酬・懲罰制度の有効性が明らかになっているが、このような制度が実際に存在し得るのか、また存在したとして人間は理論通り行動するのかといった現実への展開について課題が残されている。また、予備的研究において示した高次の報酬・懲罰制度を一般化して理論的な検討についても行う必要があると考え本計画を立案した。

2. 研究の目的

本申請では高次の報酬・懲罰系社会構造を提案し、社会的ジレンマ問題を解決し、持続可能な協調的社会を達成するための制度設計を行う。具体的には、研究期間内には以下の点を明らかにする。

高次報酬懲罰系社会構造に関して:

- (1) 理論的分析による新たなフレームワークの構築・理論解の解析的導出
 - (2) 被験者実験による理論解の実現可能性評価
 - (3) エージェントベースモデリングによる理論解との乖離の原因説明
 - (4) 実現可能な高次報酬・懲罰系社会構造の提案
- (1) では、従来個別に扱われていた報酬と懲罰の有効性比較と高次のフリーライド問題を統一的に扱うフレームワークを構築し、高次報酬懲罰系における協調を実現する条件を解析的に解くことで、設計すべき制度の指針が得る。
- (2) から、実際の人間行動と理論解にどのような乖離が存在するのかを明らかにし、現実社会に適応可能な制度を洗い出す。
- (3) によって、理論解と被験者実験との乖離原因を説明可能なモデルを構築し、解析的に解けない複雑な状況を考慮したシナリオをシミュレーションによって解く。
- (4) 実社会では実装されていない高次報酬懲罰制度がどのような効果を持ち、どのようなシナリオが生じうるのかを検証し、実現可能な高次報酬・懲罰系社会構造を明らかにする。

3. 研究の方法

以下の4点について分析研究を行った。

- (1) 高次報酬懲罰系社会構造の理論的フレームワークの構築
従来個別に扱われていた報酬と懲罰における高次のフリーライド問題を統一的に扱うことが可能となるような、高次報酬懲罰系社会構造の理論的フレームワークを構築する。フレームワークは、Axelrod によるメタ規範ゲームを基本とし、その拡張として実装する。
- (2) 解析解の導出
高次報酬懲罰系社会構造の理論解を解析的に求めることによって、持続的な協調的社会が実現可能な構造を明らかにする。解析解の導出には Evolutionarily Stable Strategy(ESS) を用い、すべての可能な構造に対して進化的に安定解を求める。また、単に実現可能性だけでなく、その安定性を数値解析し、限界理論値を導出する。まず、2次の協調報酬系における、各次元における協調と報酬の有無をまとめた 24 パターンについて解析を行い、その後高次の協調報酬系の議論を行う。ただし、高次の協調報酬系については解析解を求めることが困難である可能性が高いため、その場合は後述するエージェントベースモデルを用いた数値シミュレーションによる分析を行う。
- (3) 被験者実験の実施
被験者実験によって、協調的社会が実現するための構造を明らかにする。ここでは、考えられ得る全ての2次報酬懲罰系社会構造についてそれぞれ実験を行い、理論解との比較を行う。2次の協調報酬系におけるすべてのあり得る組み合わせ 24 パターンについて、協

調の実現率の分析と被験者の行動様式の分析を行う。

(4) 解析解と被験者実験の乖離を説明するモデル構築

解析解と被験者実験の結果の乖離を説明可能なモデルの構築を行う。特に、解析解における拘束条件及び、人が持つ限定合理性に着目したモデリングを行う。

4. 研究成果

まず、理論解析では、高次報酬懲罰系における社会的ジレンマの理論モデルの解析を行うため私的評価環境における解析手法を開発した。これにより公的評価環境との比較を理論的な側面から厳密に行えることが可能になった。その結果、私的評価環境下でも安定的に協力体制を維持できる Staying と呼ばれる規範を新たに発見することができた。この成果は[O4]として論文化され公表された。

被験者実験では、特定の規範が人間において観察されるかどうかを検証した。特に理論解析で予言された Staying 規範を人間が持ちうるということが可能なのかについて検証した。検証の結果、理論的予測通り、Staying 規範は人間に需要可能であることが確認できた。この成果は[O5]として論文化され公表された。

次に、解析解と現実の人間行動との乖離が生じるメカニズムを解明するために報酬・懲罰に関してサンクションに対する信念が行動に与える影響を検討する被験者実験をおこなった[Y1]。その結果、他者の公共財に対する行動の予測だけではなく他者のサンクション行動に対する信念が人々の協力行動に影響を与えていることが明らかになった。この成果は論文化され公表された。また認知の歪みが存在する事が報酬・懲罰の存在する公共財ゲームにどのような影響を与えうるかをさらに理論的に分析する解析をおこなった[Y2]。その結果、プロスペクト理論的な認知の歪みを持つ個人を想定すると、1次の懲罰しか存在しない環境でも2次ジレンマが発生せずに協力が安定することが解析的に明らかになった。認知に歪みがなければ常に非協力支配が唯一の安定にもかかわらず1次の懲罰で協力が安定することを示したことはこれまでの理論解析と現実の人間行動との乖離を埋めるだけではなく、協力の進化研究において認知の歪みを考慮する重要性を提示した重要な知見である。

高次報酬懲罰系社会構造の理論的フレームワークの構築では、二次の懲罰報酬系について網羅的な性質を明らかにした[T5]。また、ネットワークを導入した分析を行い[T1][T2][T3][T8]、ネットワーク構造がサンクション行動に影響を与えることを示した。また、報酬系における限界効用逓減を導入したモデルを構築し、その性質を分析した[T9][10][T12][T13]。その結果、効用逓減効果がある場合にはサンクション行動に対して多様性が生じることが明らかとなった。従来の理論的フレームワークの問題点であった一次報酬と二次報酬の実行には関連があるという仮定を崩しても、二次報酬行動が可視化されれば協調的社会が実現することを明らかにした[T7][T11]。

被験者実験の実施では、一次と二次の罰則あるいは報酬が、公共財ゲームにおける人間の行動様式に与える影響を明らかにするため、Sefton et al. (2007)を二次報酬系に拡張した被験者実験を実施した。検証の結果、二次の罰則は一次の罰則より高い協力率を維持することができるということ、二次の報酬では一次より高い協力率が困難である傾向があること、二次の罰則/報酬は一次の罰則/報酬ほどは頻繁に用いられないこと、一次の報酬を多く与える人ほど多くの二次の報酬を受ける傾向があり、一次の罰則を多く与える人ほど多くの二次の罰則を受ける傾向があることを示した。この結果は審査付き国際会議 ICSD 2017 及び HICSS 2018 にて発表した(Nishikawa, Akiyama, Okada, Toriumi, and Yamamoto)。また、囚人ジレンマにおける個人的利益と社会的利益の間のせめぎ合い、及び、自動車の手動/自動運転時の道徳的利益と功利論的利益の間のせめぎ合いが心的葛藤の増大、さらに、行動様式の変動に与える影響を分析した。この結果はHICSS2019 にて報告している。また、相互扶助ゲームにおける進化シミュレーションでの分析では、協力行動の進化を促す新たなメカニズムの提案を行った。この結果は、JAWS2018 にて優秀論文賞を受賞している[A5]。

5. 主な発表論文等

【雑誌論文】計 22 件

- [1] Uchida, S., Yamamoto, H., Okada, I., & Sasaki, T. Evolution of Cooperation with Peer Punishment under Prospect Theory. Games, 10(1), 11. (2019). <https://doi.org/10.3390/g10010011> [Y1]
- [2] Okada I., Sasaki T, Nakai Y. A solution of private assessment in indirect reciprocity using solitary observation. Journal of Theoretical Biology, 査読あり, 455, 7-15, 2018(10) [O4]
- [3] Okada I., Yamamoto H, Sato Y, Uchida S, Sasaki T. Experimental evidence of selective

- inattention of reputation-based cooperation. *Scientific Reports*, 査読あり, 8:14813, doi:s41598-018-33147-x, 2018(10) [O5]
- [4] Yamamoto, H., & Suzuki, T. Effects of beliefs about sanctions on promoting cooperation in a public goods game. *Palgrave Communications*, 査読あり, 4(1), 148. (2018). <https://doi.org/10.1057/s41599-018-0203-8> [Y2]
- [5] Okada, I., Yamamoto, H., Sato, Y., Uchida, S., & Sasaki, T. Experimental evidence of selective inattention in reputation-based cooperation. *Scientific Reports*, 査読あり, 8(1), 14813. (2018). <https://doi.org/10.1038/s41598-018-33147-x>
- [6] Kobayashi, T., Ogawa, Y., Suzuki, T., & Yamamoto, H. News audience fragmentation in the Japanese Twittersphere. *Asian Journal of Communication*, 査読あり, (2018). <https://doi.org/10.1080/01292986.2018.1458326>
- [7] Uchida, S., Yamamoto, H., Okada, I. & Sasaki, T. A Theoretical Approach to Norm Ecosystems : Two Adaptive Architectures of Indirect Reciprocity Show Different Paths to the Evolution of Cooperation. *Frontiers in Physics*, 査読あり, 6:14. (2018). <https://doi.org/10.3389/fphy.2018.00014>
- [8] Hanaki, N., Akiyama, E., and R. Ishikawa, "Effects of different ways of incentivizing price forecasts on market dynamics and individual decisions in asset market experiments," *Journal of Economic Dynamics & Control*, forthcoming, 査読あり, 88, 51-69, March (2018). <https://doi.org/10.1016/j.jedc.2018.01.018>
- [9] Hanaki, N., Akiyama, E., and R. Ishikawa, "Behavioral uncertainty and the dynamics of traders' confidence in their price forecasts," *Journal of Economic Dynamics & Control*, 査読あり, 88, 121-136, March (2018). <https://doi.org/10.1016/j.jedc.2018.01.020>
- [10] 大山知宏, 河又裕士, 秋山英三, エラーを起こすエージェントの割合が少数派ゲームにおける社会的効率に与える影響, *電子情報通信学会論文誌 D*, 査読あり, Vol.J101-D, No.9, 1354-1362, September (2018). <https://doi.org/10.14923/transinfj.2017SAP0011>
- [11] Yamamoto, H., Okada, I., Uchida, S., & Sasaki, T. A norm knockout method on indirect reciprocity to reveal indispensable norms. *Scientific Reports*, 査読あり, 7, 44146. (2017). <https://doi.org/10.1038/srep44146>
- [12] Sasaki, T., Yamamoto, H., Okada, I., & Uchida, S. The Evolution of Reputation-Based Cooperation in Regular Networks. *Games*, 査読あり, 8(1), 8. (2017). <http://doi.org/10.3390/g8010008>
- [13] Kengo Osaka, Fujio Toriumi, Toshihuru Sugawara Effect of direct reciprocity and network structure on continuing prosperity of social networking services *Computational Social Networks*, 査読あり, Vol.4 No.2 (2017)
- [14] Sasaki T, Okada I, Nakai Y. The evolution of conditional moral assessment in indirect reciprocity. *Scientific Reports*, 査読あり, 7, 41870, doi:10.1038/srep41870, 2017
- [15] Okada I, Sasaki T, Nakai Y. Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring. *Scientific Reports*, 査読あり, 7, 9737, doi: 10.1038/s41598-017-09935-2, 2017
- [16] 片平啓, 秋山英三, "参加周期の異なるプレイヤーが混在する少数派ゲーム," *情報処理学会論文誌*, 査読あり, Vol 58 No.1, 269-277, January (2017).
- [17] Akiyama, E., Hanaki, N., & R. Ishikawa, It is Not Just Confusion! Strategic Uncertainty in An Experimental Asset Market. *The Economic Journal*, 127(605) , 査読あり, F563-F580, October (2017). <http://doi.org/10.1111/eoj.12338>
- [18] Fujio Toriumi, Hitoshi Yamamoto and Isamu Okada Effect of Rewarding Systems in Social Media 2016 2ND ANNUAL INTERNATIONAL CONFERENCE ON COMPUTATIONAL SOCIAL SCIENCE(2016)
- [19] Yuki Hirahara, Fujio Toriumi, Toshiharu Sugawara Cooperation-dominant Situations in SNS-norms Game on Complex and Facebook Networks *New Generation Computing*, 査読あり, Volume 34, Issue 3, pp 273-290 (2016)[T2]
- [20] Fujio Toriumi, Hitoshi Yamamoto, and Isamu Okada: Exploring an effective incentive system in groupware *Journal of Artificial Societies and Social Simulation*, 査読あり, 19 (4) 6 (2016) [T5]
- [21] Sasaki T, Okada I, Nakai Y. Indirect reciprocity can overcome free-rider problems on costly moral assessment. *Biol. Lett.*, 査読あり, 20160341. 2016(7). doi: 10.1098/rsbl.2016.0341
- [22] Iwata, M., & E. Akiyama, Heterogeneity of link weight and the evolution of cooperation. *Physica A*, 査読あり, 448, 224-234, April (2016). <http://doi.org/10.1016/j.physa.2015.12.047>

【学会発表】計 19 件

- [1] 小澤 洋介, 島海 不二夫 ソーシャルメディアにおける金銭的報酬の効果 第 18 回社会システム部会研究会, 査読無し, (2019)[T13]
- [2] 岡田 勇, 特集「道徳判断の自動化をめぐる問題: 規範の選択と協力の進化」にあたって, *人工知能学会誌*, 査読無し, 34(2), 122-124, 2019

- [3] The Anh Han, Moniz Prereira Luis, 岡田勇, 進化的機械倫理のあらまし, 人工知能学会誌, 査読無し, 34(2), 152-159, 2019(3)
- [4] 山本仁志, 規範エコシステムアプローチによる規範と協力の共進化メカニズム, 人工知能学会誌, 査読無し, Vol.34 No.2, pp.160-167 (2019).
- [5] 鳥海 不二夫, 山本 仁志, 岡田 勇 ソーシャルメディアにおける二次報酬可視化の効果 第 15 回社会システム部会&第 59 回システム工学部会 合同研究会, 査読無し, (2018)[T11]
- [6] 三浦 雄太郎, 鳥海 不二夫, 菅原 俊治 ソーシャルメディアにおける限界効用逓減の効果 情報処理学会第 80 回全国大会講演論文集, 査読無し, (2018)[T12]
- [7] 三浦雄太郎, 鳥海不二夫, 菅原俊治 限界効用逓減の法則を考慮した SNS モデルによるレシピ共有 SNS の再現のためのパラメータ推定 第 15 回ネットワーク生態学シンポジウム, 査読無し, (2018)
- [8] 山本仁志, 協力社会を実現するための社会シミュレーション分析, 計測と制御, 査読無し, 57, pp.438-443 (2018).
- [9] 横手 美史暢, 秋山 英三. Axelrod の文化の伝播モデルにおけるエージェントの移動と全体情報の影響の分析, The Proceedings of the Joint Agent Workshop (JAWS) 2018, 査読あり, 8 pages, (2018).
- [10] 矢澤 直人, 秋山 英三. 繰り返し相互扶助ゲームにおける協力行動の進化を促すメカニズムの提案, The Proceedings of the Joint Agent Workshop (JAWS) 2018, 査読あり, 8 pages, (2018). 優秀論文賞 [A5]
- [11] Yutaro Miura, Fujio Toriumi, Toshiharu Sugawara Evolutionary Learning Model of Social Networking Services with Diminishing Marginal Utility 9th International Workshop on Modeling Social Media, WWW'18 Companion Proceedings of the The Web Conference 2018, 査読あり, pp.1323-1329 (2018)[T9]
- [12] Fujio Toriumi, Isamu Okada, Hitoshi Yamamoto Expectation for responses encourages social media participation: simulation study Social Simulation Conference 2017(2017)
- [13] Fujio Toriumi, Isamu Okada, Hitoshi Yamamoto Effect of Network Structure for Meta-Sanctions Game The 2017 International Symposium on Nonlinear Theory and its Applications (NOLTA2017), 査読あり, (2017)[T8]
- [14] 三浦雄太郎, 大阪健吾, 鳥海不二夫, 菅原俊治 ソーシャルメディアにおける限界効用逓減の効果 Workshop of Social System and Information Technology (WSSIT17), 査読無し, (2017)[T10]
- [15] 横手美史暢, 秋山英三, ``社会ネットワークのサイズと空間構造が文化圏形成に与える影響の分析,`` The Proceedings of the Joint Agent Workshop (JAWS) 2017, 査読あり, 47-52, (2017).
- [16] 小塚孝卓, 秋山英三, ``入札頼母子講における, 出資行動と入札行動の進化: エージェント・シミュレーションによるアプローチ,`` The Proceedings of the Joint Agent Workshop (JAWS) 2017, 査読あり, 122-129, (2017).
- [17] 秋山英三, 社会集団における協力形成問題へ向けて: 数理モデル, 計算機シミュレーションによるアプローチ, 組織科学, 査読無し, 51(2), 38-48, (2017).
- [18] Kengo Osaka, Fujio Toriumi and Toshiharu Sugawara Effect of Direct Reciprocity on Social Networking Services The 6th International Workshop on Emergent Intelligence on Networked Agents, 査読あり, (2016)[T1]
- [19] Kengo Osaka, Fujio Toriumi, and Toshiharu Sugawara Effect of Direct Reciprocity on Continuing Prosperity of Social Networking Services COMPLEX NETWORKS 2016, The 5th International Workshop on Complex Networks and their Applications, 査読あり, (2016)[T3]

〔図書〕(計 件)

〔産業財産権〕

出願状況(計 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年：
 国内外の別：

取得状況(計 件)

名称：
 発明者：
 権利者：

種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究分担者

研究分担者氏名：秋山英三

ローマ字氏名：**Eizo Akiyama**

所属研究機関名：筑波大学

部局名：システム情報系

職名：教授

研究者番号（8桁）：**40317300**

研究分担者氏名：岡田勇

ローマ字氏名：**Isamu Okada**

所属研究機関名：創価大学

部局名：経営学部

職名：准教授

研究者番号（8桁）：**60323888**

研究分担者氏名：山本仁志

ローマ字氏名：**Hitoshi Yamamoto**

所属研究機関名：立正大学

部局名：経営学部

職名：教授

研究者番号（8桁）：**70328574**

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。