



# HOKKAIDO UNIVERSITY

|                     |                                                                                                                                                                       |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title               | 協力と罰の共進化 : 罰に対する感受性を中心とした理論的および実証的研究 [全文の要約]                                                                                                                          |
| Author(s)           | 土田, 修平                                                                                                                                                                |
| Description         | この博士論文全文の閲覧方法については、以下のサイトをご参照ください。<br><a href="https://www.lib.hokudai.ac.jp/dissertations/copy-guides/">https://www.lib.hokudai.ac.jp/dissertations/copy-guides/</a> |
| Degree Grantor      | 北海道大学                                                                                                                                                                 |
| Degree Name         | 博士(文学)                                                                                                                                                                |
| Dissertation Number | 甲第16499号                                                                                                                                                              |
| Issue Date          | 2025-03-25                                                                                                                                                            |
| Doc URL             | <a href="https://hdl.handle.net/2115/95195">https://hdl.handle.net/2115/95195</a>                                                                                     |
| Type                | doctoral thesis                                                                                                                                                       |
| File Information    | Shuhei_Tsuchida_summary.pdf                                                                                                                                           |



# 学位論文内容の要約

博士の専攻分野の名称：博士（文学）

氏名： 土 田 修 平

## 学位論文題名

協力と罰の共進化：罰に対する感受性を中心とした理論的および実証的研究

私たち人間 (*Homo sapiens*) は、社会の中で互いに助け合いながら生きている。例えば、クラスメイトと一緒に授業の課題に取り組んだり、会社の同僚とプロジェクトに貢献したり、地域の住民と清掃ボランティアに従事することなどが挙げられる。協力は、社会的秩序の基盤であり人間を理解する上で重要なテーマである。特に、協力は進化の観点から大きな謎である (Axelrod, 1986)。この謎は、社会的ジレンマ・パラダイムを用いることで鮮明になる。社会的ジレンマとは、個人の利益と集団の利益が対立する状況である (Dawes, 1980; Van Lange et al., 2013)。全員が協力することが集団全体にとっては望ましいが、個々人の観点からはコストを負担せずに利益だけを享受しようとするただ乗り (フリーライド) の誘惑がある。個人にとって自己利益を最大化する選択肢は、他者の行動に関わらず非協力を行うことである。ダーウィンの進化論の観点から考えると、自らコストを負担して集団のために協力する協力者は淘汰されるはずだと予測される。しかし、現代においても協力は淘汰されずに頻繁に観察されている。なぜ、人間社会には、大規模な協力が広範に存在しているのだろうか。

この謎は、実験と理論の両方から探究されてきたが、一貫して罰の重要性が見出されてきた。非協力をすることで罰を受けるのならば、協力の誘因が生まれる (Olson, 1965; Ostrom, 1990)。実際、罰の機会が存在すれば、高い協力水準が維持されることが、実験と理論の双方で頑健に観察されてきた。しかし、非協力者を罰するためには、集団メンバーを監視し、違反者を特定し、報復されるリスクを負うためコストが伴う。個々人にとっては、コストを負担せず罰を行わないことが自己利益を最大化する一方で、誰も罰を行わなければ協力者が搾取されてしまう。つまり、罰自体が社会的ジレンマの構造を持っているのである (二次のフリーライダー問題)。そのため、罰によって協力が維持されることは事実だとしても、罰自体がいかんして進化し得たのかを説明しなければならない。

本論文では、この謎を解明するために、非協力者の罰に対する感受性が、協力と罰の進化において重要な役割を果たすこと主張する。すなわち、もし非協力者が、小さな罰を受けただけで改心し、協力に転向する日和見主義者であるならば、協力と罰の進化を上手く説明できることを論じる。具体的には、進化ゲーム理論モデルの解析、エージェントベースモデル・シミュレーション、集団実験、計算論モデルの解析を通して検討を行った。

本論文は6章からなる。まず第1章では、協力の進化に関する背景知識を提供した。最初に、研究史を概観し、研究パラダイムとして社会的ジレンマを導入した。社会的ジレンマの中でも、本論文では集団における協力を検討するため公共財ゲームを用いた。公共財ゲームは、経済ゲーム実験の一つとして広く実験室実験が行われており、豊富な蓄積のうち代表的な知見を紹介した。次に、本論文の研究方略として適応論的アプローチを説明した。適応論的アプローチは、現在において観察される行動や心理を、進化のアナロジーを通して探求を進める方略である。進化と自然淘汰について説明し、協力の進化を説明する代表的な理論として、血縁淘汰と互惠性を概説した。次に、多くの研究を通して罰が協力を維持するために重要な役割を果たしているにも関わらず、罰には二次のフリーライダー問題が伴うため、なぜ罰が進化するのか十分に理解されていないことを論じた。最後に、協力と罰の共進化の理論的説明として近年注目されている文化的集団淘汰を簡単に紹介した。本論文では、文化的集団淘汰の直接的な検証を行っていないが、文化的集団淘汰を用いずとも伝統的な進化ゲーム理論の枠組み内で協力と罰の共進化を説明できることを主張する。

第2章では、罰に対する感受性の重要性を明らかにした。出発点として、協力と罰の進化の分野

で大きな影響力を持つ連携罰モデル(Boyd, Gintis & Bowles, 2010, *Science*) の再解析を行った。これは、集団メンバーが連携して罰のコストを分担し、非協力者の利得を効率的に引き下げること、協力と罰の進化が説明できると主張したモデルである。連携罰モデルは、人類学的知見を反映した複数の要因が同時に投入された複雑な数理モデルであった。そこで、実験計画法のロジックに基づき、連携罰モデルから要因の効果を丹念に一つずつ外しながら検証した。その結果、原著者らの主張に反して、協力と罰の進化は、連携罰によってではなく、一回でも罰を受けると即座に協力へと転ずる非協力者がモデルの中で想定されていたからであることを見出した。さらに、罰の進化を説明したと主張する複数の理論モデルをレビューし、いずれの研究においても、罰を受けたら協力者へ転ずる日和見主義者が想定されていることを論じた。

第3章では、罰に対して鈍感な非協力者の中から、罰に対して敏感な非協力者が進化しうるかを検討した。より一般的なモデルを独自に構築して、罰に対する感受性と罰が共進化しうることを立証した。進化ゲーム・シミュレーションの結果、罰に対する感受性の低い鈍感な非協力者が支配する集団の中に、罰に対して敏感な感受性の高い非協力者が侵入できることが見出された。そして、罰に対する感受性の上昇に伴い、非協力者に罰を与える戦略と協力戦略が共進化することが明らかとなった。さらに厳密な数理解析を行った結果、罰に対する感受性は一定程度までは遺伝的浮動によって増加するが、閾値を超えると淘汰が生じ、高い感受性が適応的になることが示された。

第4章は、罰に対する感受性の一般性を実証的に検討した。第3章の理論的研究から、人々は小さな罰に対しても敏感に反応して、協力を転向することが予想される。これは、象徴罰と呼ばれる現象に酷似しており、人々は金銭的な罰を受けずとも単に不満を伝えられるだけで、協力が維持されることが知られている(Masclet, et al., 2003)。しかし、先行研究では公共財ゲームの繰り返し回数が短く、対立仮説として参加者が象徴罰とコストのかかる金銭罰を混同している可能性が拭えない。そこで、通常よりも長い50ラウンドの公共財ゲーム実験を行い、金銭罰と象徴罰の効果を比較検討した。その結果、象徴罰によって協力は長期的にも維持されることが明らかとなった。また、人々は金銭罰と同様に象徴罰を受けることで、協力的に振る舞うようになることが確認された。特に、金銭罰と象徴罰の双方を行使可能な条件では、象徴罰が金銭罰の予期的なシグナルとして頻繁に利用されることで、金銭罰のコストを抑えながら高いレベルの協力維持に貢献していることが示された。

第5章は、罰に対する感受性を強化学習という、より一般的な計算論モデルによって再解釈することを試みた。罰を受けると行動を変えろという行動原理は、強化学習の枠組みを用いて理路整然と理解することができる。また、近年、他者の利益をも考慮する社会的選好を、強化学習によって理解できるという主張も注目を集めている。そこで、公共財ゲーム実験を実施し、強化学習モデルのフィッティングし、協力性の個人差を強化学習モデルにおけるパラメータの違いとして記述することを試みた。その結果、向社会的な人ほど、学習率が高いという傾向が見られた。これは従来の社会的選好による説明とも整合することが論じられた。

第6章では、各研究で得られた知見に基づいて、本論文の主張の妥当性を検討した。本論文の各研究で明らかとなったことをまとめ、本論文の主張の妥当性を検討した。最後に、今後の展望として、協力と罰の共進化を解明するためには、象徴罰の理解が重要であること、より積極的に強化学習の枠組みを、協力行動と罰行動の研究に拡大する必要があることが論じられた。