



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〔論文内容及び審査の要旨〕
Author(s)	土田, 修平
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	https://hdl.handle.net/2115/94963
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	Shuhei_Tsuchida_review.pdf, 審査の要旨



学位論文審査の要旨

博士の専攻分野の名称：博士（文学）

氏名： 土 田 修 平

	主査	教 授	竹 澤 正 哲
審査委員	副査	教 授	高 橋 伸 幸
	副査	准教授	小 川 健 二

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

・当該研究領域における本論文の研究成果

本論文は、人間の協力行動の研究において、その重要性が長く論じられてきた罰の進化合理的基盤を検証したものである。他者の利益を増加させるために自らコストを負担する協力が、なぜ人間社会において存在しているのかという問いを巡り、多様な領域で研究が展開される中で、協力の維持において、非協力者に対する罰が重要な役割を担っていることが見出されてきた。だが罰にはコストが伴うため、自らは罰を与えないことが合理的となる。一方、たとえコストが伴うとしても、現実世界では非協力者に罰を与える人々が存在し、集団内の協力が罰によって維持されている。この謎に対して、「罰行使者は、他者から良い評判が得られる、あるいは公正な人間として取引相手に選ばれやすいことで利益を得ている」という仮説の下、多くの実験研究が行われてきた。だが一貫した証拠は得られておらず、「なぜ罰を与えるのか」という謎は、未解決の状態が長く続いている。

申請者は、この問題に斬新な視点から取り組んだ。罰を受ける非協力者にとっての合理性という視点である。罰を受けることで損失を被るならば、罰を回避するために協力することで自己利益を最大化できるだろう。もちろん、罰から生じる損失と非協力から得られる利益のバランスによって、罰に対して反応すべきか決まる。だが、協力の進化ゲーム理論モデルのほとんどは、非協力者をどれだけ罰を受けても非協力し続ける存在として定義してきた。罰に対する感受性、すなわち非協力者にとっての合理性視点は、協力の進化の文脈では、ほとんど考慮されてこなかったのである。

研究1（第2章）では、罰と協力の共進化を説明するためには、協力者が連携して罰のコストを下げる必要があると主張する著名な理論モデルの再分析が行われた。そして、このモデルでは一回でも罰を受けると協力し続ける非協力者（日和見主義者）が想定されており、その前提を取り除くと、たとえ罰のコストが低くても、罰と協力の共進化は生じないことが見出された。研究2（第3章）では、独自に進化ゲーム理論モデルを構築し、罰の感受性を連続的なパラメータで表現した。もし罰に敏感に反応し、わずかな罰を受けただけでも協力を転ずる傾向、すなわち罰に対する高い感受性が進化するならば、罰行使者が負うコストが減少する。さらに相互作用する期間が長くなるほど、非協力者が協力者へ転ずることで、罰行使者はより大きな利益を得られる。申請者は、進化ゲーム理論モデルの解析とエージェントベースモデル・シミュレーションを通して、罰に対する高度な感受性と罰が共進化し、その結果として協力が高いレベルで維持されることを見出した。なお研究1と研究2の成果は、査読付き国際学術誌への投稿に向けて、それぞれ別の英語論文として執筆が進められている。

研究3（第4章）では、罰に対する高い感受性に類似した現象として、象徴罰に着目して実験研究が行われた。金銭的な損失が発生せず、ただ単に不満を伝えられるだけでも、非協力行動が抑制されるという現象である。申請者は複数の条件を比較した結果、金銭罰と象徴罰の双方を自由に選択して行使できる条件では、非協力者に対してはまず象徴罰が行使されることを見出した。その結果、金銭罰のみが行使可能な通常の条件と比較し、金銭罰の行使量は減少しているにも関

わらず、高いレベルの協力が維持されることを見出した。これは、象徴罰が脅しとして機能し、非協力者は象徴罰に対して敏感に反応することで、金銭罰を回避していることを示唆する結果である。研究4（第5章）では、公共財問題ゲームをプレイした参加者の行動データを強化学習の計算論モデルを用いて分析した。その結果、高い学習率、すなわち直近の経験に反応する程度の高さと協力的な傾向の間に関連が見出された。いずれも、研究1・2の理論分析の結果と整合性が高い結果である。

・学位授与に関する委員会の所見

（罰を受ける側である）非協力者の合理性という斬新な視点を導入したことで、罰の進化を巡る謎は、ほぼ解消されたといっても過言ではなく、審査委員会では一連の研究の成果が高く評価された。口頭試問では、罰に対する感受性というメカニズムの領域一般性についてどう考えるかという質問、論文内で金銭罰と象徴罰が共通の神経科学的な基盤を持つ可能性が論じられているが、その論拠に対する疑問、本研究の枠組みによって、見知らぬ他者に対する一回限りの協力まで説明可能であると考えるかといった質問がなされた。申請者はいずれの質問に対しても過不足なく回答し、自身の研究が適用可能な範囲と限界を理解していることが示された。

以上から、審査委員会は全員一致で、学位申請者に博士（文学）の学位を授与することが妥当であるとの結論に達した。