

非協力者への罰行動についての心理学的・適応論的検討

学位論文内容の要旨

本論文は、非協力者への罰行動の心理メカニズムと適応基盤を検討した一連の研究成果をまとめたものである。近年、行動経済学・人類学・生物学などの分野で、人間の協力行動の進化に関する研究が活発に行われている。特にこれらの研究の文脈で、非協力者への罰行動（非協力者の利益を減らす行動）が、人間に見られる広い協力的社会を支える重要な鍵であるとして注目されている。特に近年の一部の行動経済学者や人類学者らは、非協力者への罰行動を、公正さを追求する選好に基づいて生じる“集団全体の秩序や規範の維持につながる利他行動”として理解している。しかし、“非協力者の利益を減らす行動”の中には、公正さの追求とは無関係な目的に基づく行動が含まれる可能性も考えられる。本研究では一連の実験及びシミュレーション研究を通して、近年の行動経済学で検討対象とされている非協力者への罰行動の全てが、必ずしも公正さの追求に基づいて生じる利他行動ではない可能性を指摘することを試みた。研究の背景及び目的については、第 1 章及び第 2 章で述べられている。

第 3 章では、非協力者への罰行動の全てが、必ずしも公正さを追求する選好に基づくわけではないことを指摘するために行われた実験研究を紹介している。具体的には、非協力者への罰行動の例として捉えられている“最後通告ゲームでの拒否”に注目し、その背後にある動機を探る目的の実験が行われた。最後通告ゲームでは、“提案者”と“受け手”のプレイヤーの間で金銭の分配を行う。提案者が金銭の分配方法を受け手に提案し、受け手がその提案を受け入れるか拒否するかを決める。受け手が受け入れた場合は提案者が決めたとおりに金銭が分配されるが、受け手が拒否した場合は両者とも獲得金額がゼロになる。これまでの実験研究で多くの受け手が不公正分配の提案を拒否する傾向にあることが確認されており、最後通告ゲームでの拒否は公正さを追求する選好に基づく罰行動——不公正な結果を是正する行動、不公正な相手に対する懲罰行動——として解釈されている。これに対し、第 3 章の研究では、拒否によって公正結果へ回復あるいは相手へ懲罰ができない最後通告ゲームである“一方的最後通告ゲーム”を実施し、このゲームでも不公正分配の拒否が見られるかを検討した。一方的最後通告ゲームでは、受け手が分配を拒否した場合、受け手の獲得金額のみがゼロになり、提案者は自分に分配した金額をそのまま受け取る。つまり、一方的最後通告ゲームでは、拒否によって提案者の利益には何ら損害を及ぼすことができないばかりか、提案者と受け手の間の利益格差を余計に広げることになる。この一方的最後通告ゲームに加え、受け手が拒否した事実が提案者に知らされない（それどこ

るか、提案者は受け手が受け入れもしくは拒否の選択をすることすらも知らない）“プライベート一方的最後通告ゲーム”を実施した。プライベート一方的最後通告ゲームでは、提案者に心理的な罰（不公正な扱いに対する不満を伝えること）すらできない。すなわち、これら二種類の一方的最後通告ゲームでも拒否が確認された場合、最後通告ゲームでの拒否の背後に公正さに対する選好（不公正の是正、不公正他者への懲罰）以外の動機が含まれる可能性が示唆されることになる。第1実験の結果、二種類の一方的最後通告ゲームでもかなりの割合（通常の後通告ゲームでの拒否率の約半分）で拒否が確認された。この傾向は、実験方法を変化させた第2実験と第3実験でも再現された。これらの実験結果は、罰行動（拒否）の背後に公正さに対する選好とは直接無関係な動機——不公正な扱いを受け入れることで低下する同一性を保護する動機——が相対的に重要な役割を果たしている可能性を示している。第3章の最後では、同一性保護に基づく罰の適応的意味として、感情的なコミットメント戦略としての側面、すなわち常に不公正な扱いを拒絶する態度を取り続けることで、不公正を甘受しない人間であるという評判を維持する側面として解釈できる余地について考察した。

第4章では、非協力者への罰行動の進化可能性——非協力者を罰することによって利益の獲得が見込めるか——を探る目的で行ったシナリオ実験の内容を紹介している。第4章の研究の目的は、罰行動は相互作用の中で直接個人的に利益の獲得が見込める行動であるか、それとも個人的利益の獲得とは無関係な集団利益を維持するための利他行動として解釈せざるを得ないかを検討することである。第4章の研究では、様々な実験ゲーム場面对象としたシナリオ実験を通して、非協力者を罰することで相互作用の中で個人的利益の獲得が見込めるかどうかを網羅的に検討した。評定者には、ある実験ゲームでコストを払って非協力者の利益を減らした者が登場するシナリオが呈示された。その後、評定者には、罰行使者を好意的に評価するか、また、いくつかの実験ゲームをプレイするとしたら罰行使者を相手として選びたいと思うか、更に相手にしたとしたら自分及び相手はどう行動すると思うかなどについて尋ねられた。これらの質問への回答を分析した結果、非協力者への罰行使には、1）報酬の被分配者としては選ばれにくく、親しみが持てない人物という印象を他者に与えやすいといった不利益がある一方で、2）報酬の分配者として選ばれやすく、信頼に足る相互作用の相手として認識される傾向にあり、3）罰行使者は他者から不公正な扱いをされにくいという利益がある可能性が示された。これらの結果は、罰行使者は報酬決定者（リーダー）として選ばれやすい可能性にあること、また他者からの搾取を抑制できる傾向にあることを示しており、非協力者への罰の行使にも個人的利益が伴う可能性があることを示唆している。

第5章の研究では、第3章及び第4章で示唆された罰行動の個人的利益のうち、他者からの搾取を抑制する利益に焦点を当て、搾取の回避といった個人的利益の確保に基づく罰行動が進化可能かどうか、また、そのような個人的な罰によっても相互協力関係の形成が導かれるかどうかを、コンピュータ・シミュレーションを通して検討した。最後通告ゲームと一方的最後通告ゲームがプレイされる状況で、最後通告ゲームと一方的最後通告ゲームのいずれのゲームでも一貫して不公正な扱いを拒否する戦略と、公正な分配をする戦略

が進化し得る理論的条件を探った。その結果、不公正分配の拒否が感情（不公正に対する直接の反応）を介して生じるという前提のもとならば、一貫して不公正な扱いを拒絶する罰戦略が進化すると共に、他者に罰を避ける誘因を与えることで公正行動の進化を促し、相互協力関係の形成を導き出す可能性があることが示唆された。

第 6 章では、研究成果のまとめと、研究成果が近年の協力行動及び罰行動に関する研究にもたらすインプリケーションについて述べている。本論文で示した一連の研究成果は総合して、非協力者への罰行動の全てが公正さに対する選好に基づく行動ではないと同時に、必ずしも集団利益に貢献するためだけの利他行動（自己犠牲行動）ではない可能性を示唆している。本研究成果は、近年の罰行動に関する研究に対して、罰行動の進化可能性について代替解釈の余地——個人的利益の獲得を通じた進化——を提供するものである。

学位論文審査の要旨

主 査 教 授 山 岸 俊 男

副 査 准教授 高 橋 伸 幸

副 査 教 授 瀧 川 哲 夫

学 位 論 文 題 名

非協力者への罰行動についての心理学的・適応論的検討

本論文は、集団内での非協力者に対する罰行動の進化可能性が、集団淘汰を前提とすることでのみ理解可能なのか、それともより直接的な個人的利益から理解可能なのかを、経済ゲーム実験、シナリオ実験、そしてコンピュータ・シミュレーションを通して検討することを目的としている。また、行動経済学者が罰行動を研究する際に一般的に用いられる最後通告ゲームをとりあげ、最後通告ゲームにおける受け手による不公正提案への拒否行動を、不公正な提案を行った提案者に対する罰行動として理解すべきかどうかという問題を検討することを、副次的な目的としている。

第1実験の結果は、これまで不公正回避への選好および不公正な意図を持つ相手への罰として理解されてきた最後通告ゲームにおける不公平提案の拒否行動が、実はそうした理由だけでは説明できず、別の説明を必要としていることを明らかにしている。続くシナリオ実験の結果は、罰行使者は集団状況での報酬決定者（つまりリーダー）として選ばれやすい可能性があること、また他者からの搾取を抑制する傾向があることを示しており、非協力者に対する罰の行使に直接の個人的利益が伴う可能性があることを示唆している。第5章で紹介されている3つのコンピュータ・シミュレーションの結果は、一方的最後通告ゲームでの拒否行動と公正分配の進化には、二つのゲーム間での行動の一貫性が何らかの方法で導入される必要があることを明らかにしている。学位申請者はこの一貫性が怒りの感情により担保されると考え、怒りの感情に基づく行動を導入したシミュレーションを実施することで、この解釈の妥当性を検証している。

本論文の成果は、非協力者に対する罰行動の適応基盤が、報酬決定権を持つプレイヤーとして他者から選ばれやすくなるという点と、他者からの搾取を抑制するというコミットメント問題解決にある点を実験とコンピュータ・シミュレーションにより示した点にある。これまでの研究では、社会的ジレンマ状況における協力促進のために非協力者に対する罰が有効であることが繰り返し指摘されているが、罰行動そのものの個体レベルでの適応基盤については明らかにされてこなかった。そのため非協力者に対する罰の進化を説明するためには、何らかの集団淘汰の原理を採用しなければならないという議論が経済学・人類学の一部で受け入れられてきた。集団淘汰による罰行動の説明は、そのインプリケーショ

ンとして、集団内での協力は集団間での攻撃性と共進化するというものであり、この原理が成立するかどうかは社会科学全体にとって極めて大きな意味を持っている。従って集団淘汰による罰および協力の説明を採用するかどうかは慎重に検討する必要がある。申請者の研究は、罰行動に関しては、より直接的な個人的利益による説明が可能であることを指摘するものであり、社会科学全体に対して持つ意義は大きい。また、最後通告ゲームが人々の公正追求行動の測定に広く用いられている現状を考えると、そこでの不公平提案に対する拒否が必ずしも公正さを求める行動でもなく、また不公正な対戦相手を罰するための行動でもないという知見は、今後の最後通告ゲーム研究において重要な意味を持ってくるであろうと考えられる。この点も、本論文の成果として強調されるべき点である。

上にまとめられた本研究の成果は、人間社会における協力行動を維持する一つのメカニズムとしての非協力者に対する罰行動が、必ずしも集団淘汰を前提としなければ説明できないわけではないことを示した点で、社会科学全体にとって大きな意義を有している。審査の過程では、実験研究とコンピュータ・シミュレーション研究がそれぞれ想定している状況が正確に対応していない点、コンピュータ・シミュレーション研究がいわゆる「思考実験」の枠を越えないのではないかという点などについての指摘がなされたが、そうした研究の限界は本研究全体が持つ意義を大きく損なうものではないという点で、審査委員の意見は一致している。

本審査委員会は、本論文に示された申請者の研究の成果を高く評価し、全員一致で、本論文を博士（文学）の学位を授与されるにふさわしいものであるとの結論に達した。