

サンクションの誤推定がもたらす協調行動：ワンショット公共財ゲームによる検討

○山本仁志 遠藤はるか（立正大学）

The effect of misestimation of sanction in public good game

*H. Yamamoto and H. Endo (Rissho University)

概要— 社会的ジレンマに関する理論研究やシミュレーション研究によって、2次のジレンマを解消するためには懲罰よりも報酬のほうが協力の進化に有効であることが示されてきた。一方、プロスペクト理論における損失回避性から、人間は報酬による利得より懲罰による損失を大きく評価することが予測される。本研究では、ワンショット公共財ゲームによる被験者実験によって懲罰と報酬の効果を比較する。実験の結果、懲罰のほうが報酬よりも協力率の増加に効果的であること、懲罰に対する推測が報酬に対する推測よりも大きいことがわかった。更に、人の持つ互惠性は報酬ではなく懲罰に発揮されることがわかった。

キーワード: 公共財ゲーム, サンクションシステム, プロスペクト理論

1 はじめに

人間をはじめ社会構造を有する複数の生物種において、競争環境にありながらも他者を助け役立とうとするような利他行動、向社会的行動が観察されることはよく知られている。これらの行動を促すメカニズムの解明は、流動的な社会において安定的なソーシャル・キャピタルを提案することにつながり社会的意義を有する。

社会的ジレンマ状況における協力を促進するためには多くの可能性が検討されており、互惠性による協力の進化^{1, 2)}を代表的なアプローチとして、他にも血縁やネットワーク構造による協力の進化が提案されている³⁾。一方で、現実社会においては制度やシステムによって協力を促す仕組みが導入されている。例えば、協力をしなければ罰を与え、協力をすれば報酬を与える「サンクション」^{4, 5, 6, 7)}や、オークションサイトに見られるよう評価制度のようにこれまでの協力行動について示す「評判システム」^{8, 9)}、行動を第三者が確認しているというシグナルによる「監視シグナル」¹⁰⁾などがある。

公共財ゲームにおけるサンクションの効果については、理論研究¹¹⁾やシミュレーション研究¹²⁾によって、高次のジレンマを解消するためには懲罰のようなマイナスのサンクションよりも報酬のほうが協力の達成に有効であることが示されてきた。一方で、被験者実験による実証研究ではこれらの効果について、報酬のほうが効果的であるという報告¹³⁾や懲罰のほうが効果的であるという報告⁶⁾といった報告があり一貫した結果は得られていない。また、そもそも2次のジレンマが生じないという報告¹⁴⁾や、報酬が高次のジレンマを防ぐ効果を持つ¹⁵⁾といった結果も報告されている。Ballietら¹⁶⁾によるメタ分析では、全体的にはわずかに懲罰のほうが効果が大きい、ワンショット公共財ゲームでは懲罰の優位性があり、繰り返しゲームにおいては報酬が有効であると報告されている。

実社会にサンクションシステムを実装することを想定すると、報酬より懲罰のほうが容易であり一般的にも実装例が多い。協力が達成された社会において、懲罰システムは少数の裏切りに対して実働するため懲罰

コストは低く維持できるが、報酬は常に多数の協力者に払われ続ける必要がある。Hilbe and Sigmundの理論研究¹⁷⁾においても懲罰のほうが報酬よりも安定的だと示されている。

プロスペクト理論^{18, 19)}で説明される損失回避性によって、人間は同じ額の利得と損失について、損失を大きく評価する傾向があることが知られている。そのため非協力を選んだときに懲罰によって被る損失の心理的インパクトのほうが協力に対する報酬のそれよりも大きくなると考えられる。また、確率加重関数と価値観数の性質から、発生確率が低いマイナス利得については過大評価し、逆に発生確率が高いプラス利得については過小評価することがわかっている。サンクションがあることで全体の協力率が高くなると見積もった状態では、低い確率の負の事象である「自分だけが裏切られて懲罰される時の損失」は大きく評価され、高い確率の正の事象である「周囲と同様に協力した上で報酬を得られる利得」は小さく評価されると考えられる。こうしたメカニズムによって懲罰による損失回避が働くため、懲罰のほうが報酬よりも協力率を引き上げる効果が大きいことが予測される。

理論上、サンクションシステムのあるジレンマ状況では二次、三次と高次のジレンマが発生し、一次のサンクションだけでは協力率の増加は達成されないとされていた²⁰⁾。しかしながら、多くの研究においてサンクションシステムが存在していることで、実際に罰や報酬の行使がなくとも協力率が増加することが明らかになっている^{21, 22, 23)}。小野田ら¹⁴⁾は、こうした理論上と現実との差の間には、サンクションシステムによって導かれる心理過程、つまり規範の過大視が起こっている、と指摘した。小野田らの研究によって、サンクションシステムがあることで協力を促す“状況拘束的協力者”は、一次的ジレンマの協力者が行う罰行使の程度を実際よりも高く見積もることがわかった。この懲罰の過大視が、1次のサンクションシステムが存在しているだけでジレンマを解消する原因であると考えられる。この主張は、本研究が指摘する人々が損失回避性を持つことで懲罰を過大評価することと整合的である。一方で、報酬の利得に対しては懲罰の損失と比較して過小に評価されると考えられる。そこで本研究

では、懲罰・報酬のそれぞれのサンクションシステムがある公共財ゲームにおいて、協力率の差異だけでなく、サンクションに対する推定も測定することで懲罰と報酬の効果の違いとそのメカニズムの解明を目指す。

2 実験

実験は、2015年10月23日(金)に、東京都内私立大学2年の大学生2クラス、計110名を対象に行われた。回答を行わない、途中で回答を止める等の受講者を除いた100名(うち男66、女34名)を有効回答として分析対象とする¹。罰条件に割り当てられたクラスの有効回答数は43、報酬条件の有効回答数が57である。講義の最後に実験に協力してもらうことを講義開始時にあらかじめ伝えておき、講義の最後20分間で実験を行った。今回は、実際に金銭が発生しない場面想定法で行われたため、参加者に対する説明の中で、ゲームで得た金額が実際に支払われるという想定で実験に協力するよう強調した。

質問紙配布後、紙面とパワーポイントを使いゲームのルールが説明された。ゲームは一回戦と二回戦に分かれており、それぞれサンクションなし条件とサンクションあり条件であった。また、2クラスは、それぞれ罰条件と報酬条件の質問紙が配られた。実験は、一回きりの公共財ゲームが用いられており、ランダムに組まれた3人グループで行った。

公共財ゲームの基本的なルールは、最初に元手となる800円がグループの全員に配られ、それを主催者に預けると倍の1600円が他の2人に均等に配分される、というものであった。参加者はこのゲームにおいて元手を預けるか・預けないかを選択した。サンクションあり条件においては、罰制度あるいは報酬制度のために元手とは別に400円が渡されて、この中から支払った金額の倍の額が非協力者から引かれる、あるいは協力者に渡されるというルールがプラスされた。

サンクションあり条件において、“サンクションあり条件の囚人のジレンマゲームに協力するか”、という質問だけでなく、“サンクションにコストを支払うか”、“支払う場合にはいくら支払うか”、また、“この実験に参加している人はサンクションにいくら支払うと思うか”、を回答してもらった。

また、実験の最後に、一般的信頼・互恵性と罰あるいは報酬に対する恐れや期待を測定するための7項目の設問にも5件法で回答してもらった。項目はそれぞれ、“ゲームにおいて、「自分以外のグループのメンバーが、800円を預けないのではないか？」と恐れを抱いた”、“二回目のゲームにおいて、「自分以外のグループのメンバーが、仕返し制度を使うのではないか？」と恐れを抱いた(二回目のゲームにおいて、「自分以外のグループのメンバーは、ありがとう制度を使うだろう」と期待を抱いた)”、“ほとんどの人は信頼できる”、“大抵の人は人から信頼された場合、同じように相手を信頼する”、“自分は信頼できる人と信頼できない人を見分ける自信がある”、“誰かに助けってもらったら、自分もまた他の誰かを助ける”、“人に親切にすると、結局はめぐりめぐって自分にいいことがあると考えている”の7つである。アンケートまで記入が終わった人から

¹サンクションなしで協力し、サンクションありで非協力を選択した計3名も実験状況を理解していないと判断し無効回答とした。

順番に質問紙を回収し、解散となった。

3 結果

3.1 罰と報酬の効果の分析

罰条件、報酬条件のそれぞれの群のもともとの協力率に差異がないことを確認するため、サンクションなし条件の協力率に違いがあるか分析した(Fig. 1上パネル)。分析の結果有意差は認められなかった($F(1, 98) = 1.199, p = .276$)。続いて、罰と報酬のどちらがジレンマ状況における協力の維持に効果があるか検証するため、サンクションあり条件における協力率の差を分析した(Fig. 1下パネル)。分析の結果、罰条件の協力率が高いことが確認された($F(1, 98) = 4.391, p = .039$)。

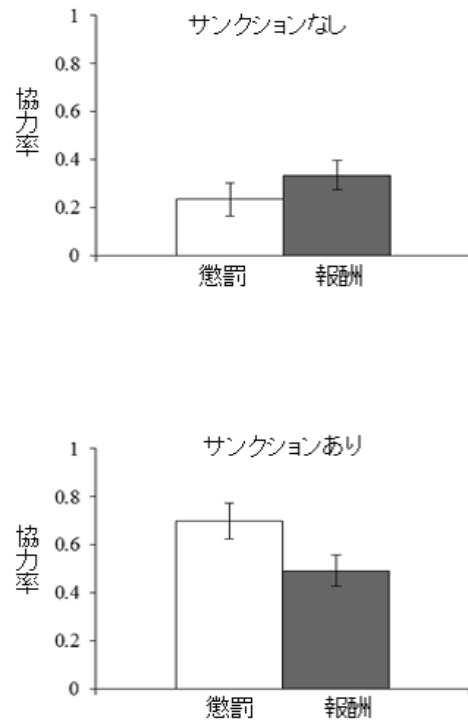


Fig. 1: サンクションの有無による協力率の差異

罰条件において協力率が高いことが確認されたが、その違いが生み出された要因を探るため、実際にサンクションに使われた「サンクション行使額」(Fig. 2上パネル)と、周囲の参加者がサンクションに使う額を推定した「サンクション推定額」(Fig. 2下パネル)を比較する。実際にサンクションに使用された額については差がなかった($F(1, 98) = .459, p = .499$)が、懲罰のほうがサンクション推定額が高い傾向が見られた($F(1, 98) = 3.703, p = .057$)。

3.2 誰がサンクションを過剰推定するのか

続いて、サンクションを過剰に推定するのは誰なのかを詳細に検討する。以降の分析ではサンクションの行使・推定額が行動戦略によって異なるのかどうかを探索的に分析する。

最初に公共財への貢献の有無とサンクションの行使・推定について分析する。サンクションなし条件の行動とサンクションあり条件の行動の組み合わせと、サンクション行使額とサンクション推定額の間を分析した。組み合わせは、常に裏切りのDD、サンクションあ

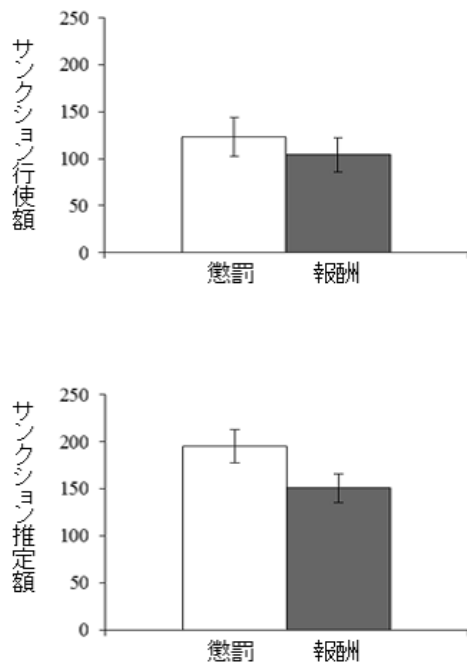


Fig. 2: サンクションの行使額と推定額

り条件のみで協力する DC、常に協力の CC である²。先行研究¹⁴⁾では状況拘束的協力者 (DC に該当) の推定値とそれぞれのタイプの行使額を比較しているが、より一般的な検討を行うため、すべての群の行使額と推定額を対象として分析を行う。罰条件・報酬条件においてそれぞれ DD/DC/CC の 3 タイプとサクシヨンの行使額・推定額の 2 要因の分散分析を行った。罰条件の分析の結果、行使/推定の主効果のみ有意であった ($F(1, 80) = 4.493, p = .029$) (Fig. 3 左パネル, 表 1)。行動タイプの効果と交互作用については有意な効果は得られなかった。つまりサンクションに実際に行使された額よりも推定額のほうを高く見積もっているが、行動タイプによる違いは観察されない。同様に報酬条件においても 2 要因の分散分析を行った。分析の結果、行使/推定の主効果 ($F(1, 108) = 2.766, p = .099$) が有意傾向であり、行動タイプの主効果が有意であった ($F(2, 108) = 9.845, p = .000$) (Fig. 3 右パネル, 表 2)。交互作用については有意な効果は得られなかった。報酬条件においても行使額より推定額のほうが高くなっている。Holm 法を用いた多重比較によって、行使額においてタイプ DD が他の 2 タイプより低く、また推定額において DD タイプが DC タイプより低くなっている。

3.3 互惠性による協力・サンクションの影響

ここまでの分析では、ゲーム中の行動戦略と協力行動とサンクションの関係を分析してきた。本節では心理的態度が協力行動・サンクションにもたらす影響を分析する。

本研究では互惠性を Putnam²⁴⁾ が “generalized reciprocity” として述べた「いずれ誰かが私を助けてくれるだろうから、私はあなたに何か見返りを期待せずこれをやる」と同様の概念として扱う。これは「情けは人の

²CD の組み合わせは前述のとおり分析から外している。

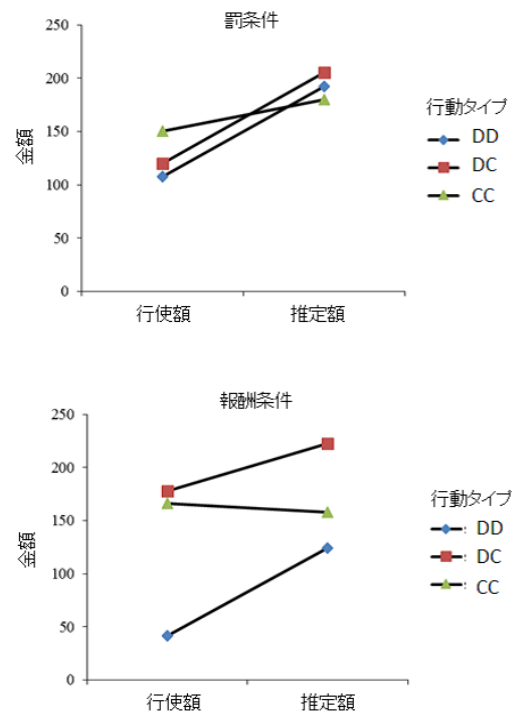


Fig. 3: 罰・報酬における行動タイプと推定額・行使額

為ならず」と解釈が可能である。実験終了後におこなった質問紙調査の中から上記の互惠性に対応する 2 項目を採用しこれらの回答の単純加算によって被験者の互惠性得点とする。互惠性得点を算出する項目は「誰かに助けてもらったら、自分もまた他の誰かを助ける」「人に親切にすると、結局はめぐりめぐって自分にいいことがあると考えている」の 2 項目である ($\alpha = .526$, 平均 = 7.290, $SD = 2.114$)。また全参加者の互惠性得点の平均値より高い群を「互惠性高群」平均値より低い群を「互惠性低群」として 2 群に分割した。

3.3.1 互惠性と協力行動

互惠性 (低群・高群) とサンクションの種類 (罰・報酬) による協力行動の違いを分析するために 2 要因の分散分析を行った。互惠性の主効果 ($F(1, 96) = 5.016, p = .027$)、サンクションの種類的主効果 ($F(1, 96) = 4.277, p = .041$) が有意であった (Fig. 4)。互惠性の高い人ほど協力率が高いという直観的な予想に整合的な結果となっている。

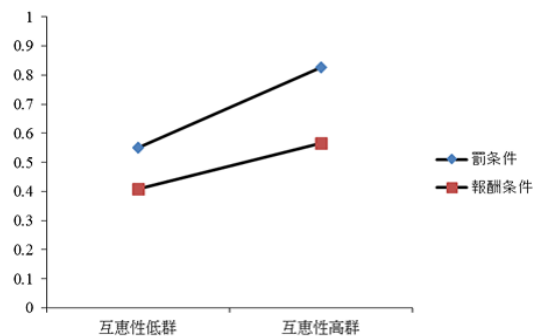


Fig. 4: 罰・報酬における互惠性と協力率

Table 1: 行動タイプと行使額・推定額による分散分析（罰条件）

条件	タイプ	DD		DC		CC		主効果		交互作用
罰	行使/推定	行使	推定	行使	推定	行使	推定	タイプ	行使/推定	0.325
	金額	107.69 155.25	192.31 132.05	120.00 147.26	205.00 94.45	150.00 158.11	180.00 113.53	0.093	4.943*	

上段:平均値、下段:標準偏差 ** $p < .01$, * $p < .05$, + $p < .10$

Table 2: 行動タイプと行使額・推定額による分散分析（報酬条件）

条件	タイプ	DD		DC		CC		主効果		交互作用
報酬	行使/推定	行使	推定	行使	推定	行使	推定	タイプ	行使/推定	1.817
	金額	41.41 94.54	124.14 115.43	177.78 120.19	222.22 120.19	165.79 133.39	157.89 112.13	9.845**	2.766+	

上段:平均値、下段:標準偏差 ** $p < .01$, * $p < .05$, + $p < .10$

3.3.2 互恵性とサンクション行使・推定

サンクションの行使額ならびに推定額について、互恵性（低群・高群）とサンクションの種類（罰・報酬）の2要因の分散分析を行った。行使額において互恵性の主効果 ($F(1, 96) = 6.431, p = .013$)、交互作用 ($F(1, 96) = 4.957, p = .028$) の効果が有意であった。一方推定額においては、サンクションの種類の主効果 ($F(1, 96) = 3.605, p = .061$) のみが有意傾向であった。

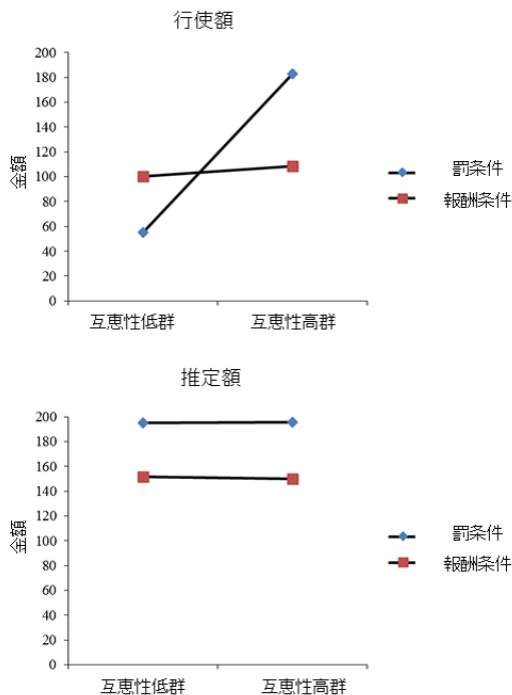


Fig. 5: 罰・報酬における互恵性と推定額・行使額

3.4 一般的信頼による協力・サンクションの影響

一般的信頼を測定する項目としては「ほとんどの人は信頼できる」「大抵の人は人から信頼された場合、同じように相手を信頼する」の2項目を採用し、単純加算によって被験者の一般的信頼得点とした ($\alpha = .571$, 平均 = 5.330, $SD = 1.897$)。また全参加者の一般的信頼得点の平均値より高い群を「信頼高群」平均値より低い群を「信頼低群」として2群に分割した。

3.4.1 一般的信頼と協力行動

一般的信頼（低群・高群）とサンクションの種類（罰・報酬）による協力行動の違いを分析するために2要因の分散分析を行った。サンクションの種類の主効果 ($F(1, 96) = 4.277, p = .041$) と交互作用 ($F(1, 96) = 7.263, p = .008$) が有意であった (Fig. 7)。これまでの分析で明らかのように罰条件において協力率が高くなっている。罰条件では一般的信頼による協力率の違いはないが、報酬条件において一般的信頼が低い群の協力率が低くなっている。

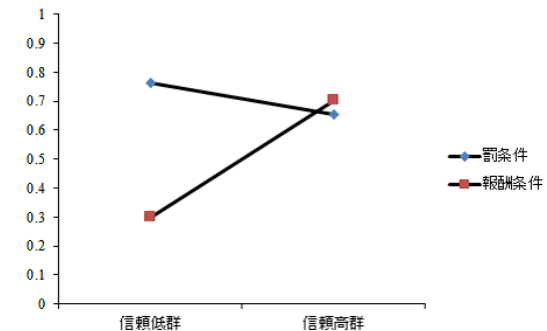


Fig. 6: 罰・報酬における一般的信頼と協力率

3.4.2 一般的信頼とサンクション行使・推定

サンクションの行使額ならびに推定額について、一般的信頼（低群・高群）とサンクションの種類（罰・報酬）の2要因の分散分析を行った。行使額において交互作用 ($F(1, 96) = 7.616, p = .007$) の効果が有意であった。一方推定額においては、サンクションの種類的主効果 ($F(1, 96) = 4.852, p = .048$)、交互作用 ($F(1, 96) = 10.536, p = .002$) が有意であった。

行使額について、互恵性とは逆の傾向がみられている。互恵性に関しては罰条件において高群が高い行使額を示しているのに対し、一般的信頼では逆に高群のサンクション行使が低くなっている。一方、報酬条件においては、互恵性については差が見られないが、一般的信頼については高群で高い行使額を示している。

推定額について、これまでの分析同様罰条件で推定額が高くなっている。しかし、低信頼群において罰の推定が高くなっており、逆に報酬の推定は低くなっている。

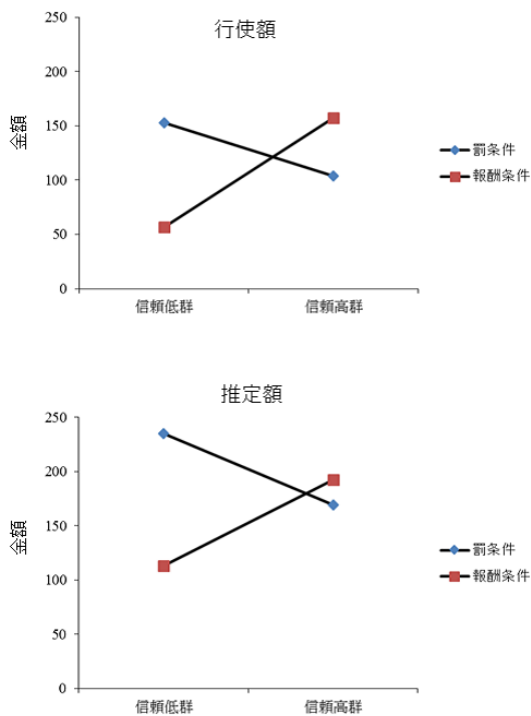


Fig. 7: 罰・報酬における一般的信頼と推定額・行使額

4 考察

ワンショット公共財ゲームにおいて、報酬よりも罰の方が協力行動を導く効果が高いことが明らかになった。また、サンクション推定額においても、報酬条件よりも罰条件の方が高かった。これらの結果から、人々が報酬への期待より罰への恐れを大きく抱いているために、罰の方が効果的に働いたと考えられる。

そして、このサンクション推定額の差を生んでいるのは、報酬条件における非協力者であることが明らかになった。サンクションの有無による行動の組み合わせによるサンクションの行使額と推定額の関係进行分析したところ、報酬条件において有意差がみられた。常に非協力をとる人 (DD) は、サンクションがあることで協力に転じる人 (DC) と比較して行使額も推定額も低く見積もっている。

この結果は、サンクションに貢献するかどうかと公共財に貢献するかどうかの分析からも支持される。罰条件において有意差は見られなかった一方で、報酬条件において、「裏切りかつサンクションを行わない」戦略タイプと、「協力かつサンクションを実行する」戦略タイプの推定額に差があった。つまり、報酬への期待値が低い人が裏切りへとシフトするために、ジレンマの構造を変革するというサンクション本来の効果が発揮されていないといえる。

互恵性とサンクションの行使・推定額から興味深い示唆が得られた。行使額について互恵性高群が高くなることについては妥当な結果と言えるが、驚くべきことに罰条件において互恵性の高低で行使額に大きな開きがあった。また多重比較の結果、互恵性高群において報酬よりも罰に大きな金額が行使されている。この結果は、互恵性の高い人達は非協力者を強く罰することを示している。互恵性は「自分の協力はいつか返報

される」という信念を指す故に報酬において強く発揮されると直観的には考えられる。しかし実際には罰において強く発揮されている。つまり互恵性は「協力に対して報酬する」という形ではなく「非協力に対して罰する」というマイナスの返報として表出していると考えられる。Pedersen ら²⁵⁾ は、実験で観察される第三者罰は嫉妬の感情が元になっているものがあると指摘しており、人間の互恵性は利己的に利益を得ようとするものをスパイト的に排除することで協力関係を維持することに有効に機能していると考えられる。

5 まとめ

社会的ジレンマ状況において、人々の協力行動を導くために、罰や報酬といったサンクションシステムが有効とされてきた。しかし、多くの研究が行われる中で、罰と報酬のどちらがより効果的かについては様々な条件によって異なる結果が報告されている。本研究では、人の持つ損失回避性に着目し、ワンショット公共財ゲームにおいてはサンクションの大きさについての推定が、罰と報酬で逆の効果を持つのではないかと考え、これを測定する実験をおこなった。

実験の結果、サンクションシステムの導入によって協力率は増加し、更に懲罰のほうが報酬よりも協力率を増加させることが分かった。また、協力率の差が生じた原因を探るため、他者がサンクションを行使する量を推定させたところ、懲罰において他者のサンクション行使量を多く見積もっていることがわかった。これは人間の損失回避的な性向が懲罰による損失を大きく見積もるため生じていると考えられる。

謝辞

小野田竜一氏 (北海道大学), 小林哲郎氏 (City University of Hong Kong), 鈴木貴久氏 (NII) からは貴重な意見とディスカッションの機会を頂戴した。ここに謝意を表す。また本研究は JSPS 科研費 26330387, 15KT0133 の助成を受けて実施された。

参考文献

- 1) Karl Sigmund. Moral assessment in indirect reciprocity. *Journal of theoretical biology*, 299(2) 25/30 (2011)
- 2) Martin a Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437(7063) 1291/1298 (2005)
- 3) Martin a Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805) 1560/1563 (2006)
- 4) Ernst Fehr and Simon Gächter. Altruistic punishment in humans. *Nature*, 415(6868) 137/140 (2002)
- 5) Ernst Fehr, Urs Fischbacher, and Simon Gächter. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1) 1/25 (2002)
- 6) Matthias Sutter, Stefan Haigner, and Martin Kocher. Choosing the carrot or the stick/ Endogenous institutional choice in social dilemma situations. *The Review of Economic Studies*, 77(4) 1540/1566 (2010)
- 7) Manfred Milinski and Bettina Rockenbach. On the interaction of the stick and the carrot in social dilemmas. *Journal of theoretical biology*, 299 139/143 (2011)
- 8) Chrysanthos Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10) 1407/1424 (2003)

- 9) Yannis Bakos and Chrysanthos Dellarocas. Cooperation without enforcement/ a comparative analysis of litigation and online reputation as quality assurance mechanisms. *Management Science*, 57(11) 1944/1962 (2011)
- 10) Melissa Bateson, Daniel Nettle, and Gilbert Roberts. Cues of being watched enhance cooperation in a real-world setting. *Biology letters*, 2(3) 412/4 (2006)
- 11) Isamu Okada, Hitoshi Yamamoto, Fujio Toriumi, and Tatsuya Sasaki. The Effect of Incentives and Meta-incentives on the Evolution of Cooperation. *PLOS Computational Biology*, 11(5) e1004232 (2015)
- 12) Fujio Toriumi, Hitoshi Yamamoto, and Isamu Okada. Influence of Payoff in Meta-Rewards Game. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 18(4) 616/ 623 (2014)
- 13) Aljaz Ule, Arthur Schram, Arno Riedl, and Timothy N Cason. Indirect punishment and generosity toward strangers. *Science*, 326(5960) 1701/4 (2009)
- 14) 小野田 竜一, 松本 良恵, and 神 信人. 社会的ジレンマにおける 協力促進要因としての規範の過大視. Center for Experimental Research in Social Science Working Paper Series (92) (2009)
- 15) Toko Kiyonari and Pat Barclay. Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment. *Journal of personality and social psychology*, 95(4) 826/42 (2008)
- 16) Daniel Balliet, Laetitia B. Mulder, and Paul A. M. Van Lange. Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4) 594/615 (2011)
- 17) Christian Hilbe and Karl Sigmund. Incentives and opportunism: from the carrot to the stick. *Proceedings of the Royal Society B: Biological Sciences*, 277(1693) 2427/2433 (2010)
- 18) Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica: Journal of the Econometric Society*, 47(2) 263/291 (1979)
- 19) Amos Tversky and Daniel Kahneman. Advances in Prospect Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5(4) 297/323 (1992)
- 20) R.M Axelrod. An evolutionary approach to norms. *American Political Science Review*, 80(4) 1095/1111 (1986)
- 21) M.D. Caldwell. Communication and sex effects in a five-person Prisoner's Dilemma Game. *Journal of Personality and Social Psychology*, 33(3) 273/280 (1976)
- 22) Kaori Sato. Distribution of the cost of maintaining common resources. *Journal of Experimental Social Psychology*, 23(1) 19/31 (1987)
- 23) Toshio Yamagishi. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51(1) 110/116 (1986)
- 24) R. Putnam. *Bowling Alone*. Simon & Schuster (2000)
- 25) Eric J Pedersen, Robert Kurzban, and Michael E McCullough. Do humans really punish altruistically? A closer look. *Proc. R. Soc. B*, 280(1758) (2013)