

# 人間社会の秩序を支える罰

帝京大学文学部心理学科 講師

堀田結孝 (ほりた ゆたか)

## Profile—堀田結孝

2011年、北海道大学大学院文学研究科博士課程修了。博士（文学）。日本学術振興会特別研究員PD、国立情報学研究所ビッグデータ数理国際研究センター特任助教を経て、2016年より現職。専門は社会心理学、進化心理学。論文は「Punishers may be chosen as providers but not as recipients」「Observation enhances third-party punishment only among people who were not hot-tempered」（いずれも *Letters on Evolutionary Behavioral Science* に掲載）など。



本稿では、社会心理学や行動経済学などの知見を交えながら、罰と秩序維持との関係についての研究を紹介する。罰と聞くと穏やかではないイメージを抱かれるかもしれない。しかし、犯罪者や脱税者を取り締まる警察署や税務署など、規範逸脱者への罰が社会を支える上で必要不可欠であることは疑いようのない事実である。そして、規範逸脱者への罰を促す心理メカニズムが私たちに強く備わっていることを、これまでのさまざまな研究が示している。

## 社会的ジレンマと罰

まず罰の話に入る前に、罰が必要となる状況である「社会的ジレンマ」について説明する。社会的ジレンマとは、個人の利益追求が集団利益の損失を招く状況のことを指す。ハーディンによる「共有地の悲劇」を挙げながら、どのような状況であるかを説明する (Hardin, 1968)。農民たちが自身の所有する羊を自由に放牧できる共有地がある。羊をたくさん放牧すれば、羊をたくさん育てることができ、農民は多くの利益を享受できる。しかし、共有地周辺の農民全てが自分の利益だけを考えて羊を過放牧すれば共有地の草は枯れ果て、誰も放牧できなくなってしまふ。すなわち、社会的ジレンマとは個人が利己的に振る舞うことで、結局全員の損失につながってしまう状況である。社会的ジレンマは、学校の掃除や地球温暖化など身近な共同体から国家レベルまでさまざまな現実社会の協力の問題に当てはめることができ、人間が集団を

築いて生活する以上、避けることのできない問題であるといえる。

ここで、社会的ジレンマの解決策として容易に思いつくのは、利己主義者を罰する制度の導入であろう。罰を受けて損をすることになるならば、利己主義者も罰を避けて集団のために協力するようになるだろう。この解決策は一見良いアイデアのように思えるが、理論的にはそう単純にうまくいかない。なぜならば、利己主義者を見張ったり罰を与えに行ったりなど、罰の行為自体にもコストが掛かるからである。自分は罰のコストを負担せず誰かに罰を任せるほうが得ならば、誰も罰のためのコストを負担しないだろう。罰の導入は結局、「罰のコストを支払うことに協力するか」という新たな社会的ジレンマを生み出すことになり、この問題は「二次のジレンマ」として知られている (Yamagishi, 1986)。

## 利己主義者への罰に関する実験

このように、理論的には罰による社会的ジレンマの解決は困難であるが、現実の人間は自己利益を顧みず積極的に利己主義者を罰することが行動経済学などの実験研究で示されている。以下に、代表的な例をいくつか紹介する。

まずは、「公共財ゲーム」と呼ばれる社会的ジレンマ状況を反映した実験ゲームである。公共財ゲームでは、実験参加者同士でグループを作ってお金を使ったゲームを行う。仮に、4人のグループがあるとする。4人にはそれぞれ、

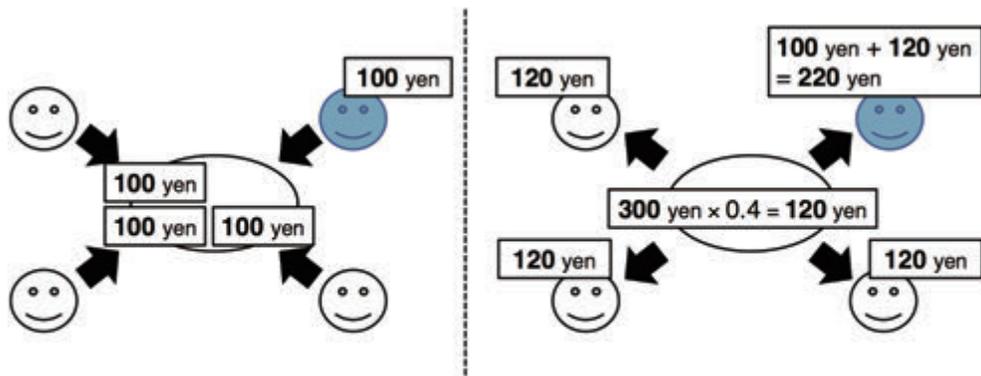


図1 公共財ゲームの例

元手として100円が実験者から与えられる。4人はそれぞれ、元手の100円の中からいくらを「公共財」に提供するかを決める。提供しなかった分は、自分の手元に残る。公共財に提供された金額の合計を0.4倍した金額が、グループの各メンバーに分配される。例えば、3人が100円を公共財へ提供し、1人が全く提供しなかった場合、公共財に提供された金額の合計は300円で、その0.4倍の120円が4人に分配される。全く提供しなかった人は、自分の手元に残した100円を合わせて、合計220円を受け取る(図1)。このように、公共財ゲームでは個人にとってはお金を提供しない(集団のために協力しない)ほうが得である。しかし、誰も提供しなければ獲得金額は全員元手の100円のまま、これは全員が100円を提供したときの獲得金額である160円(=400円×0.4)よりも少ない。この公共財ゲームを繰り返し行くと、回数を経るごとに公共財への提供額が低下していく傾向にある。

では、お金を提供しない人を罰する仕組みを加えたらどうなるだろうか。フェアとゲヒター(Fehr & Gächter, 2002)は、公共財への提供額の決定後に他のメンバーを罰するオプションを加えて実験を行った。他のメンバーを罰する場合には、自分の手元の金額から罰のための費用を支払う必要があり、支払った費用の3倍の金額が罰のターゲットから差し引かれた。フェアとゲヒターの実験では、参加者のうち約84パーセントが少なくとも1回は罰のための費用を払い、罰のある条件は罰のない条件よりも公

共財への提供額が大きく上昇する傾向にあった。その後も罰付きの公共財ゲーム実験は数多く行われ、多くの人々がコストを掛けて利己主義者を罰する傾向にあることが示されている。

次に、「最後通牒ゲーム」を紹介する。最後通牒ゲームでは、分配者と受け手の間でお金の分配を行う。分配者は受け手に分け方を提案し、受け手はその提案を受け入れるか拒否するかを決める。受け入れた場合は、お金は分配者の決めたとおりに分けられる。拒否した場合は、分配者と受け手両者とも受け取れるお金がゼロになる。例えば分配者に800円、受け手に200円という分け方を受け手が提案された場合、ゼロ円よりかは200円をもらったほうがましなので、受け手が自己利益を追求する人間ならば、どんなに不平等でも常に受け入れるはずである。しかし、この予測とは反し、これまでの数多くの最後通牒ゲーム実験は、受け手が不平等分配を非常に高い確率で拒否することを示している(Camerer & Thaler, 1995)。このシンプルな実験は、自己利益を犠牲にしても利己主義者を罰したいという欲求が人々に強く備わっている証拠として挙げられてきた。

公共財ゲームと最後通牒ゲームでは、利己主義者を罰する者は利己的行動の直接の被害者であった。これに対し、「第三者罰ゲーム」を用いた実験研究は、人々は直接の被害者の立場でなくても、利己主義者を罰する傾向にあることを示している。第三者罰ゲームでは、まず分配者と受け手の間でお金の分配が行われる。仮に、分配者が1,000円を自分に800円、受け手

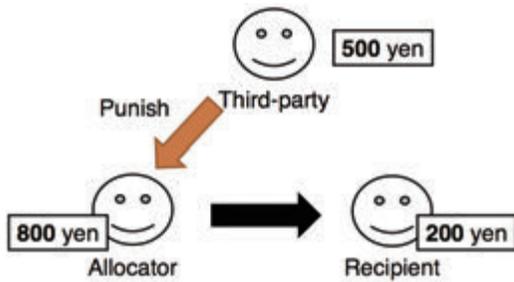


図2 第三者罰ゲームの例

に200円で分けたとしよう。次に、別の第三者にその分配結果が知らされ、第三者が分配者を罰するかを決める。第三者には500円が渡され、そのお金のいくらかを使って分配者のお金を減らすかを決める。第三者が支払ったお金の3倍の金額が、分配者の獲得金額から差し引かれる(図2)。最後通牒ゲームでは受け手が分配者を罰する(分配を拒否する)かを決めたが、第三者罰ゲームでは受け手とは異なる別の第三者が決める。初めて第三者罰ゲーム実験を行ったフェアとフィッシュバツハーの研究では(Fehr & Fischbacher, 2004)、第三者の役割となった参加者のうち約60パーセントが不平等分配をした分配者に罰を与えた。その後の研究でも「第三者罰」を行う人が少なからず存在することが確認されている。

### 罰の背後にある心理メカニズム

これらの例からもわかるように、現実の人間は自己利益を犠牲にしてでも利己主義者に対して罰を行う。罰が規範を支えるために行われるものならば、罰は協力を追求する動機と強く結びついていることが予想される。行動経済学でも、罰行動は公正さに対する選好の現れである規範維持行動として捉えられてきた。

しかし、これには懐疑的な意見もある。例えば最後通牒ゲームでの拒否のような仕返しとしての罰には、社会的公正を侵害されたことへの義憤だけでなく、自分が軽く扱われたことへの私憤を晴らす側面もありえる。山岸ら(Yamagishi et al., 2012)は最後通牒ゲームで不平等分配を拒否した者が他の実験ゲームで必ずしも協力的に振る舞うわけではないことを示

し、最後通牒ゲームでの拒否は必ずしも公正や協力をめざした行動ではなく、軽視に対する反発としての行動の現れである可能性を主張している。仕返しとしての罰は相互協力を目標とした動機とは無関係であるが、結果として利己主義者を抑制し、規範維持としての機能を果たすのかもしれない。その一方で、第三者罰は協力傾向と関連する可能性を示す知見もある(李・山岸, 2014)。

### 罰と評判との関係

現実の人間には利己主義者を積極的に罰する傾向が備わっている一方、依然として先述の二次のジレンマの問題は残されている。二次のジレンマの解消の鍵の一つとして近年、罰の背後にある評判の利益が注目されている。利己主義者を罰することで周囲から好意的に評価され、罰を行使しなかった者よりも有利に扱われるならば、一時的なコストを払ったとしても長期的な視点で見れば罰は報われることになる。

では、罰を行使することで良い評判を獲得できるのだろうか。これに関しては近年検討されているが、罰のメリットを示す研究もあれば、デメリットを示す研究もある(詳細なレビューは、Raihani & Bshary, 2015を参照)。例えば筆者の過去の研究(Horita, 2010)では、公共財ゲームや第三者罰ゲームでの罰行使者は、他の人から信頼に足る人物として評価される一方、親しみにくい印象を抱かれやすいことが示唆されている。これらのことから、非協力者を罰した人は単純に「良い人」と評価されるわけではなく、罰がもたらす評判には多様な側面があると解釈できる。例えば一つの側面として、「公正な意図を持った人間」という評判が考えられる。罰が復讐などの私怨ではなく協力的な意図による正当な行動と認識されれば、公正な振る舞いが重視される場面では罰行使者は信頼に足る相互作用の相手として選ばれる可能性にある(Barclay, 2006)。他にも罰の評判として、強さを示す側面も考えられる(Hilbe & Traulsen, 2012)。利己主義者を罰することで「不公正な扱いをされたら必ず復讐する」こと

を周囲に信じ込ませれば、将来に渡って不公正な扱いをされ続けなくて済む。逆に罰のコストを惜しむ人間は、周囲から不公正な扱いをされ続ける結果に甘んじることになってしまう。このように、他者に強さを示すことが必要な文脈では、利己主義者を許さない人間であることの評判は有利に働くと予想される。最後通牒ゲームでの拒否のような軽視に対する反発としての罰には、このような利益が期待できるかもしれない (Yamagishi et al., 2012)。

その一方、強さの誇示は親しみにくい人という評判をもたらすことで、相互作用の相手として避けられるデメリットを伴うことも予想される。つまり、罰が行使者に有利に働くかは、罰行使者の置かれた状況や行動がどう帰属されるかにも依存する可能性にある。罰と評判の関係については引き続き検討が必要な課題であるが、罰行動の背後にある動機や有利になる環境にも注目するなど、多様な観点からのアプローチが必要であろう。

### おわりに

以上、罰と秩序維持との関係についての研究例を紹介してきた。ここで紹介してきた研究では、警察や裁判所といった中央集権によるフォーマルな罰制度ではなく、共同体成員が自らの手で利己主義者を罰するインフォーマルなかたちの罰を扱っている。中央集権制度は近代において作られたものであり、私たちヒトの進化の歴史はそれらの制度のない世界で満ちていたはずである。現代においてさえも、学校や職場での揉め事をいちいち裁判所に訴えたりなどしないように、公的な制度に頼る場面のほうが少ないといえるかもしれない。2009年にノーベル経済学賞を受賞した政治学者オストロムは、世界各地の共有資源の多くが共同体の構成員たちによる自発的な制裁によって維持されてきた点を指摘している (Ostrom, 1990)。インフォーマルな状況においても人々はいかにして秩序問題を解決してきたのか、解決に導く心理メカニズムとは何かについての問いは、人間社会の大規模な協力の成り立ちを理解する上で

重要な問題であるといえよう。

### 文 献

- Barclay, P. (2006) Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 27, 325-344.
- Camerer, C. & Thaler, R. (1995) Anomalies: ultimatums, dictators and manners. *The Journal of Economic Perspectives*, 9, 209-219.
- Fehr, E. & Gächter, S. (2002) Altruistic punishment in humans. *Nature*, 415, 137-140.
- Fehr, E. & Fischbacher, U. (2004) Third-party punishment and social norms. *Evolution and Human Behavior*, 25, 63-87.
- Hardin, G. (1968) The tragedy of the commons. *Science*, 162, 1243-1248.
- Hilbe, C. & Traulsen, A. (2012) Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific Reports*, 2, 458.
- Horita, Y. (2010) Punishers may be chosen as providers but not as recipients. *Letters on Evolutionary Behavioral Science*, 1, 6-9.
- 李楊・山岸俊男 (2014) 強い互酬性仮説の検証：協力行動と罰行動の関係。『心理学研究』85, 100-105.
- Ostrom, E. (1990) *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Raihani, N. J. & Bshary, R. (2015) The reputation of punishers. *Trends in Ecology & Evolution*, 30, 98-103.
- Yamagishi, T. (1986) The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology*, 51, 110-116.
- Yamagishi, T., Horita, Y., Mifune, N., et al. (2012) Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences USA*, 109, 20364-20368.