

# 聖人の調和と囚人のジレンマ

—利他性を考慮したゲーム理論分析—

河合 伸<sup>a</sup>

## 要約

本論文は、ゲーム理論の代表的ゲームである「囚人のジレンマ」について、従来の研究ではほとんど考慮されてこなかった人間の「利他性」を導入することにより、「非協力」かつ「有限回」のゲームであっても囚人のジレンマ状態が解消されるケースがあることを示すことを主眼としている。囚人のジレンマ・ゲームの前提条件は、(i)非協力ゲームであること、(ii)自分の利得のみに関心があること、(iii)有限回のゲームであることである。このなかで、(i)および(ii)の条件が成立していても、(iii)の条件が成立していない場合、すなわち、無限回繰り返しゲームによって、囚人のジレンマが解消されることが知られているが、ここでは、(i)と(iii)の条件が成立していても、(ii)の条件を緩和し、自身の行動が自分の利得のみならず、他人の利得に及ぼす影響にも関心がある場合、その関心がある程度高ければ、囚人のジレンマは解消され、「聖人の調和」が達成されることを示す。また、孔子の教えに基づいて囚人のジレンマ・ゲームを解いた場合にも「聖人の調和」が達成されることを示す。

JEL 分類番号 : D91, D11

キーワード : 囚人のジレンマ、利他性、贈り物ゲーム

---

<sup>a</sup> 河合伸 東日本国際大学経済経営学部准教授 skawai@tonichi-kokusai-u.ac.jp

## 1. イントロダクション

### 1.1 目的

本論文は、ゲーム理論の代表的ゲームである「囚人のジレンマ」について、従来の研究ではほとんど考慮されてこなかった人間の「利他性」を導入することにより、「非協力」かつ「有限回」のゲームであっても囚人のジレンマ状態が解消されるケースがあることを示すことを主眼としている。

本論文では、この囚人のジレンマを解決する鍵として人間の利他性を考慮する。現代の主流派経済学では、人間の利他性に信頼をおかず、利己心のみで行動する主体を前提として議論を進めることで、利他心に基づく行動までは考慮しないという慣習がある。しかし、現在行き詰まりつつあるグローバル資本主義経済において、新たな処方箋を見出すには人間の利他性に焦点を当て、それを信頼し、そこに訴えかける政策・方向性しかないと考えられるものである。

### 1.2 方法

囚人のジレンマ・ゲームの前提条件は3つある。1つ目はプレイヤー同士が「非協力」であること、2つ目は個々のプレイヤーが自分の利得しか考えないこと、3つ目は一回限り、あるいはたかだか有限回繰り返しゲームであることである。ここで「非協力」という意味は、事前に話し合っ協力することができない、そして得られる利得の配分について拘束力のある取り決めができない、ということである。したがって、通常、非協力であるから他人のことは一切考えないで行動するという具合に、非協力であることと個人の利得のみを考えて行動することは同じことであるかのように扱われているが、「非協力」であっても、他人の利得の変化を考慮して、それも含めて行動する他者への「思いやり」を行動仮説として議論することは可能である。

そこでまず、議論の土台となる囚人のジレンマについて標準型ゲームにおける囚人のジレンマを示した後、無限回繰り返しゲームによって利己的動機に基づく利他的行動の説明となる「贈り物ゲーム」を分析し、その問題点を洗い出す。次に、利他性を考慮した効用関数を用いた場合の囚人のジレンマを解き、利他性を示す「徳」の度合いに応じて、「協調の失敗」や「聖人の調和」など様々なナッシュ均衡が現出することを示し、それを「贈り物ゲーム」に応用する。

## 2. モデル

### 2.1 徳因子の導入

ここでは、一回限りの囚人のジレンマ・ゲームにおいて、利他性を考慮することで、囚人のジレンマが解決することを示す。ここでは非協力ゲームであることは維持しつつ、相手の戦略に対して、自身の戦略を選択する際に、自分の利得のみを比較するのではなく、相手の利得がどうなるのかについてある程度思いを馳せる個人を想定する。つまり相手の立場に立って考える「思いやり」を表す指標を考える。その思い

やりの指標を「徳因子」と名付け $\alpha$ とおく。 $\alpha$ は $0 < \alpha < 1$ とする。やはり人間である以上、自分のことに一番関心があることを前提とし、 $\alpha$ は限りなく1に近くはなるものの1とはならないものとする。すると表4におけるプレイヤー1の行動を考えると、プレイヤー2が協調を選択している場合に、プレイヤー1が協調を選択するときを考慮する利得は次のように表される。

$$H + \alpha H \quad (7)$$

このように他者の利得（効用）を「徳因子」でウェイト付けして考える類似の手法は、アロー（Arrow, 1981）およびバットら（Bhatt, Ogaki, & Yaguchi, 2015）にあり、その簡略化版は利他主義の包摂の例として紹介している岡部 [岡部光明, 2017]にある。バットらは徳因子を自分がその対象となる相手に費やした時間の関数として内生的に捉えている点で興味深いだが、ここでは徳因子の値はパラメータとして外生的に与えることにする。また、(7)式の第1項を直接的利得、第2項を間接的利得と呼ぶことにする。このゲームにおいて注意しておきたいのは、相手の利得を考慮するにあたって、相手も自分のことを考慮してくれている場合、その相手の間接的利得も包含すべきかどうかの問題が生じるが、ここでは相手の直接的利得のみを考慮するものとする。

同様にして各利得のマスにおけるプレイヤー1の利得のみを表したのが次の表6である。ここで、

$$G > H > L > M, 2H > (G + M) > 2L \quad (8)$$

である。(8)式は先の一般的な囚人のジレンマの条件に $(G + M) > 2L$ を加えた形となっている。ここで改めて(8)式の意味について考えてみよう。(8)式の最初の不等式は各プレイヤーが直接的利得のみを比較したときに相手の出方に関係なく常に裏切ることが最適となっているための条件である。2番目の不等式は各戦略の組み合わせにおける両者の利得の合計についての条件となっている。お互いに協調しているときがその値が最も高く、そして互いに裏切っているときが最も低く、どちらか一方が裏切っているときの利得の合計がその中間に位置することを示している。

表6 徳因子を考慮した利得表

1 \ 2	協調 (C)	裏切り (D)
協調 (C)	$H + \alpha H, H + \alpha H$	$M + \alpha G, G + \alpha M$
裏切り (D)	$G + \alpha M, M + \alpha G$	$L + \alpha L, L + \alpha L$

出典：筆者作成

各プレイヤーは同質的としているため、この利得表は、プレイヤー2にとっても同様である。ここで次の定理が成立する。

## 2.1 利他性を考慮したナッシュ均衡

**定理3（聖人の調和）** 利他性を考慮した囚人のジレンマ・ゲームにおいて、各プレイヤーの徳因子  $\alpha$  が、

$$\alpha \geq \max\left\{\frac{G-H}{H-M}, \frac{L-M}{G-L}\right\} \quad (9)$$

ならば、互いに協調することがナッシュ均衡点となる。

徳因子  $\alpha$  が(9)式の条件を満たしているプレイヤーを「聖人」と呼び、そのナッシュ均衡を「聖人の調和」と呼ぶことにする。ここで、(5)式と(9)式を比較すると非常に似通っていることがわかる。実際、 $(G-H)/(H-M)$ の値は同じである。これは(6)式と(10)式の導出の仕方を考えるとわかるが、しっぺ返し戦略の(6)式は2期間の比較において今期も来期も互いに協調することによって得られる自分の利得の割引現在価値と今期裏切るが来期には裏切られる場合の自分の利得の割引現在価値を比較しているのに対して、思いやりを考慮した場合は、それを今期において相手の立場を自分のことのように置き換えて比較しているという違いはあるものの、比較する値は同じとなっているためである。

定理3の結果の重要な含意は一回限り、あるいは有限回の囚人のジレンマ・ゲームであっても、相手への思いやりがある程度あれば、非協力の環境において、お互いに協調を選択することが合理的に説明できるということである。

ここで次の補題が導かれる。

**補題1** 囚人のジレンマの利得が条件式(8)に加え、 $L+H > G+M$ ならば、このゲームのナッシュ均衡は、徳因子  $\alpha$  の大きさに応じて次のように特徴付けられる。

表7 徳因子の範囲と解の特徴1

徳因子の範囲	$\alpha < \frac{G-H}{H-M}$	$\frac{G-H}{H-M} \leq \alpha < \frac{L-M}{G-L}$	$\alpha \geq \frac{L-M}{G-L}$
解の特徴	囚人のジレンマ	協調の失敗	聖人の調和

出典：筆者作成

補題1における重要な含意は、徳因子  $\alpha$  の値が協調の失敗の範囲にある場合においても、お互いに協調することはナッシュ均衡の一つとなっていることから、たとえ有限回であっても繰り返しゲームにおけるしっぺ返し戦略やトリガー戦略のナッシュ均衡点として協調解が現れることを説明することができる。そして、協調の失敗であるがゆえに、囚人のジレンマもナッシュ均衡として起こりうることもあわせて指摘しておきたい。これは囚人のジレンマの実験において有限回の試行においてすら、お互いに協調を選択するケースが有意に観察されることと整合的である。このような場合、これまでは個人はそこまで将来を見通すことはできないとする限定合理性や、無限回

ではなく無期限でありさえすればよいという形で説明がなされてきたが、新たな根拠を提供するものである。

## 2.2. 贈り物ゲーム再考

ここでは、利他性を考慮した場合に、II-3の贈り物ゲームがどのように修正されるのかを考察する。2人の個人1と2がお互いに財の贈り物をする状況を考えよう。財は2種類とし、財1と財2の消費量 $x_1$ と $x_2$ に対して、個人1の効用関数は徳因子 $\alpha$ の下、

$$U_1 = u_1(x_1, x_2) + \alpha u_2(x_1, x_2) = x_1^1 x_2^1 + \alpha x_1^2 x_2^2$$

とする。個人2も同様に対称的な効用関数を持つものとする。

個人1と2の財の初期保有ベクトルを、それぞれ(5,1)と(1,5)とする。個人の戦略は次の2通りとする。

C：自分の多く持っている方の財を1単位相手に贈る

D：贈らない

このゲームの利得表は次の表9で表される。

表9 利他性を考慮した贈り物ゲーム

1 \ 2	C	D
C	8+8 $\alpha$ , 8+8 $\alpha$	4+10 $\alpha$ , 10+4 $\alpha$
D	10+4 $\alpha$ , 4+10 $\alpha$	5+5 $\alpha$ , 5+5 $\alpha$

出典：筆者作成

この状況から、修正された利得構造は補題2の条件を満たしているので、ナッシュ均衡は次のようになる。

表10 利他性を考慮した贈り物ゲームの解の特徴

徳因子の範囲	$\alpha < \frac{1}{5}$	$\frac{1}{5} \leq \alpha < \frac{1}{2}$	$\alpha \geq \frac{1}{2}$
解の特徴	囚人のジレンマ	チキンゲーム	聖人の調和

出典：筆者作成

つまり、贈り物をする機会がたとえ1度限り（もしくは有限回）であっても徳因子の値が1/2以上、言い換えると、自分の直接利得に対する思いの半分くらいの思いやりで相手の気持ちを考えてあげられるのであれば、互いに贈り物をするのがナッシュ均衡となる。

## 3. 結論

本論文では、従来の「囚人のジレンマ」に利他性を導入した場合に、1回限りのゲームにおいても囚人のジレンマが解消され、「聖人の調和」が達成されることを示すことができた。さらにそれを贈り物ゲームに応用して、無限回繰り返しゲームでなくても利他的動機さえあれば利他的行動が生まれうるとの結果を得た。

今後の研究課題として、本論文で考察した同質的な2人同時プレイのゲームをn人同時プレイへ一般化することや、異質的な個人が参加するゲームへの拡張、不完全情報の導入などにより、定理の普遍性を確認することが挙げられる。

## 引用文献

- Arrow, K. (1981). "Optimal and Voluntary Income Distribution,". In Rosefielde(ed.), *Economic Welfare and the Economics of Soviet Socialism: Essays in Honor of Abram Bergson* (pp. 267-288). Cambridge University Press.
- Barrow, R. (1974). "Are Government Bonds Net Wealth?". *Journal of Political Economy*, 82, pp.1095-1117.
- Becker, G. (1974). "A Theory of Social Interactions,". *Journal of Political Economy*, 82, pp.1063-1093.
- Bhatt, V., Ogaki, M., & Yaguchi, Y. (2015). "Normative Behavioral Economics Based on Unconditional Love and Moral Virtue". *Japanese Economic Review*, 66(2), pp.226-246.
- Nash, J. (1951). "Non-Cooperative Games,". *Annals of Mathematics, Second Series*, 54(2), pp.286-295.
- Phelps, E. (1975). *Altruism, Morality, and Economic Theory*. Russell Sage Foundation.
- アダム・スミス 村井章子＋北川知子 [訳]. (2014). 『道徳感情論』. 日経 BP 社.
- アダム・スミス 大河内一男監訳. (1976). 『国富論 I』. 中央公論社.
- 宇沢弘文・渡辺格. (2017). 『生命・人間・経済学－科学者の疑義』. 日本経済新聞社.
- 岡田章. (1996). 『ゲーム理論』. 有斐閣.
- 岡部光明. (2017). 『人間性と経済学』. 日本評論社.
- 後藤隆一. (2007). 『人間主義経済学序説』. 循環社会研究所.
- 松井彰彦・清水武治. (2006). 『ゲーム理論』. 三笠書房.
- 泉田成美・柳川隆. (2008). 『プラクティカル 産業組織論』. 有斐閣アルマ.
- 大西昭. (1979). 『人間主義経済学－世界経済の未来への指針』. 第三文明社.
- 土居丈朗. (2002). 『入門 公共経済学』. 日本評論社.
- 堀元. (2001). 「利他性と効用相互依存」. 著: 井堀利宏・岡田章・伴金美・福田慎一 [編], 『現代経済学の潮流 2001』 (ページ: 51-69). 東洋経済新報社.