

社会規範論におけるルールの外部性と観察者視点について (野方宏教授退任記念号)

メタデータ	言語: ja 出版者: 静岡大学人文社会科学部 公開日: 2013-03-12 キーワード (Ja): キーワード (En): 作成者: 田島, 慶吾 メールアドレス: 所属:
URL	<a href="https://doi.org/10.14945/00007078">https://doi.org/10.14945/00007078</a>

## 論 説

# 社会規範論におけるルールの外部性と 観察者視点について

田 島 慶 吾

### 問題の所在—現在の社会規範論の問題性—

現在のゲーム理論中心の社会規範の研究<sup>(1)</sup>は多くの成果を生み出してきたが、にもかかわらず、なぜ、人が利己的顧慮を離れて、規範的な行為を採るのか、という社会規範論の最も基本的な問題を解決しているようには思えない。本稿はゲーム理論的社会規範論の批判的評価に基づき、行動規則（ルール）への顧慮とルールの外部性（観察者視点）が従来の社会規範論の難点を克服し、今後の社会規範論の展開において重要な把握であることを明らかにするものである。

### I. 社会規範（Social Norm）を定義する。

本節と次節ではゲーム理論的社会規範論を概観し、更に問題点を明らかにしよう。

#### 1. 社会規範の日常的意味

社会規範という言葉で我々が普通思い浮かべる意味は何であろうか<sup>(2)</sup>。社会規範は、第一に、「ある社会的状況における行為選択の指示」、つまり、「状況Wでは、行為Xは可能である、行為Yをするべきである、行為Zをしてはならない」という行為選択の基準・指示である。つまりこれは

<sup>(1)</sup> これまで経済学において、社会規範は第一に、集合行為問題（共通資源問題）として、第二に、効率性の問題（不完全情報や取引費用の存在による不完全な契約に対する代替）として、最後に、Anomalities問題、つまり、伝統的な利己的モデルでは説明できない一見非利己的、反利己的動機に基づく行動が存在するのはなぜか、という問題として提起されてきた。経済学の伝統的想定である「合理的利己主義者」モデルによる予測の失敗（predictive failures）または調整の失敗（coordination failures）は、行為の規則性（習慣、慣行、法の制約を受けた行動、規範的行動）により説明でき、また、失敗を回避できるという立場に基づく。この解決は、社会規範の「調整」機能に求められよう。つまり、特定の社会的状況において、行為者の行為が規範により一定の方向に導かれることにより、行為の規則性が生まれ、この規則性により、予測の失敗が説明され、調整の失敗が解決されるとする立場である。共通資源問題については、Hardin (1993 [1982]), Axelrod (1997), Taylor (1987), Ostrom (2005), 山岸 (1991) を、新制度派経済学と規範との関係はKoford and Miller (1991) を、行動経済学についてはFehr and Fischbacher (2004) をそれぞれ参照のこと。法と経済学の立場については飯田 (2004) を参考。

<sup>(2)</sup> 社会規範の最も一般的な定義は、「社会の成員によって一般的に遵守される振る舞い（conduct）または行為（action）に対する指示された手引き（prescribed guide）」（Ullmann-Margalit, 1977, p.13）である。

行為の「望ましさ、正しさ」(両者がどのように定義されるにせよ)を基準とした行為選択、行為そのものの選択<sup>(3)</sup>に関わる。第二に、社会規範は利己的利害を離れた行為の選択という側面を持つ。「状況Wでは行為Yをするべきである、行為Zをしてはならない」という場合、状況Wでは行為Zが最も利己心に叶うかもしれないが、その行為をあえて行わない、という利己心の抑制の側面<sup>(4)</sup>である。この利己心問題は、社会規範は「破る」ことができる、ということを示している。社会規範は個人に選択の機会、つまり、規範に従った行為を選択するか、または規範を破るか、の2つの行為の選択の機会を与えるものと見なすべきであろう。これは社会規範の維持にはサンクションが必須であることを示している。第三に、社会規範は状況Wでは、多くの人が行為選択の基準に従い行為するという「行為の規則性」を表すであろう。「多くの人が……行為選択の基準に従い行為する」ことにより、「多くの人が」が「状況Wでは、行為Yをし、Zを行わない」<sup>(5)</sup>という行為の規則性が生まれよう。従って、社会規範は個人的な性向、資質、価値・道徳観とは区別される。つまり、社会規範は個人の性向には還元できない、行為選択に関わる「客観的な」、つまり、全ての個人においてではなくとも、多数の個人に共有される基準であろう。以上の要素をまとめれば、社会規範の「日常的な」定義が与えられよう。

「社会規範とは、自己利益への顧慮を離れ(つまり、利己的な行為を抑制し)、社会の全ての成員においてではないが、多くの成員において共有される行為そのものの「良し悪し」を判断する基準である。」

本稿はこの義務論的な行為選択の基準、利己心の抑制、行為の規則性を社会規範の必須要件と考えるが、これらの要件は後述のゲーム理論的社会規範論の要件とは必ずしも一致しない。なぜ一致しないのか、その理由を与えることは本稿の課題の一つであるが、その前に、ゲーム理論における社会規範論とはどのようなものであるかを理解する必要がある。

## 2. ゲーム理論は社会規範をいかに理解するか。

最初に述べたように、現在社会規範論の考察は主としてゲーム理論によって行われている。こ

<sup>(3)</sup> 「社会規範はどんな行為が人々の集合にとって適切で正しい(あるいは不適切で正しくない)と見なされるかを特定する」(Coleman, 1990, p.274). Bicchieri (2006) もまた、社会規範と慣習とを前者における「義務と規範的期待」とによって区別した (Bicchieri, 2006, p.42).

<sup>(4)</sup> 「規範が問題となるのは、行為者の非規範的選好に基づいた場合に比較して、規範的選好に基づく選好が、厚生を最大化しない場合である。」(Griffith and Goldfarb, 1991, p.92)

<sup>(5)</sup> 「ある行為を行わない」という行為の規則性は後述するゲーム理論では排除される。だが、ある規範が、この状況で行為Zを行うべきではない、と指示する場合、「ある行為を行わない」ということもまた規範的行為であろう。ある状況で「ある行為を行わない」という選択肢はゲームの「外」にであることを意味している。注②を参照。

のゲーム理論的社会規範論が定義する社会規範は前項の「日常的」な社会規範とは異なっているが、なぜ、異なるのかを考察する前に、規範はどのような状況で必要とされるか、行為者にとって最善な行為とは何か、行為の規則性とは何か、という観点からゲーム理論的社会規範論を概説する必要がある。

### ① 問題的社会状況—社会規範の必要性—

ゲーム理論的社会規範論では、全ての社会的状況で、社会規範が必要とされるわけではなく、特定の社会的状況が「問題的」である場合に、規範の必要性が生まれるとする<sup>6)</sup>。この状況は問題的社会状況 (problematic social situation) と呼ばれ、パレート改善的な状態を実現できる行為があるのに、戦略的に相互に独立した利己的合理的な行為者はこれを実現できない状況 (Cf. Raub and Voss, 1990, p.81; Voss, 2001, p.110) と理解される (行為の調整による相互利益)。ゲーム理論はこの社会的状況を純粋調整ゲーム、囚人のジレンマゲーム、チキンゲーム、両性の戦いゲーム等<sup>7)</sup>として定式化し、協力、慣習、公平性、互酬性、信頼等を取り扱ってきた。

### ② 「最善な行為」と「ナッシュ均衡」

ゲーム理論的社会規範論の多くは利己的選好を想定している。これは自己の物的・金銭的利得 (のみ) への関心により行為を選択することを示す。最近のゲーム理論では、この利己的選好の他に、「社会的選好 (social preference)」を想定する場合がある (Ben-Ner and Putterman, 1998, pp.17-26; Bowles, 2004, pp.109-19)。社会的選好とは自己の物的、金銭的利得に加えて、他者の物的、金銭的利得への関心を示す選好である。利他主義、公平な分配への関心、互酬性などの他、妬み、悪意等も含まれる (Fehr and Fischbacher, 2004, pp.2-4)<sup>8)</sup>。利己的選好の持ち主であれば、最も多くの自己利得 (利益) をもたらす行為を選択しよう。社会的選好の持ち主であれば、他者の物的、金銭的利得を勘案し、自己の利得が減ることがあっても、自分の他者配慮的な選好順序に応じた行為選択を行う。いずれにしても、行為者は選好順序に応じた行為選択を行う。ある社会的状況において、行為者のそれぞれが、相手の行為選択を所与として、「自分にとって最善な行為」を選択する時、この選択は戦略的に合理的であり、この結果生まれる状況 (「ナッシュ均衡」) は

<sup>6)</sup> 社会規範が必要となる社会的状況とは何であるかについては議論がある。Colemanは規範の需要を「行為の外部性」に求めた。Cf. Coleman (1990, Ch.10)

<sup>7)</sup> ゲーム理論的社会規範論で、本稿で取り扱わない問題はバーゲニング問題である。これは調整ゲームの一種である「両性の戦い」ゲームにより表現される。

<sup>8)</sup> 社会的選好は、「他者顧慮的選好 (Other-regarding preferences)」と「過程顧慮的選好 (Process-regarding preference)」に細分される。「他者顧慮的選好」とは、自己利益及び他者利益への関心を示し、「過程顧慮的選好」とは、自己の利益または他者の利益という結果がいかに生じたかについての関心を意味し、「価値、倫理、行動の基準」への選好とほぼ同じ意味とされる (Ben-Ner and Putterman, 1998, p.20)。

「自己拘束的」である。つまり、自分にとって最善と思われる行為を選択しているので、その行為から逸脱する誘因がない、という意味で「自己拘束的」である。ナッシュ均衡が社会規範論に関わるのは、この自己拘束性のためである。規範とは行為者自らが自らに課す行為のルールであると理解される。

### ③ 行為の規則性

ナッシュ均衡は必ずしも行為の規則性を意味しない。行為の規則性とは、特定の社会的状況が繰り返される場合に、複数の行為者に共通して見られる行為の斉一性を意味する。他方、行為に規則性が見られる状態がそのまま規範が成立している状態ではない。従って、ナッシュ均衡に行為の規則性が見られるためには、(私にとっての最善の行為、あなたにとっての最善の行為)が「同じ行為」<sup>9)</sup>でなくてはならない。この場合に、行為の規則性が見られ、かつこれはナッシュ均衡である。

### ④ ゲーム理論による社会的規範の定義

Voss (2002) による定義が選好、行為の規則性、ナッシュ均衡の諸概念を用いたゲーム理論的社会的規範の最も一般的な定義であろう。Vossによれば、「社会規範とは、以下の条件における行為者の集団P内での行為の規則性Rである。1. Rは集団Pにおける行為者間で繰り返される相互行為において生じる。2. 集団Pのほとんど全ての行為者が、集団Pの他のほとんど全ての行為者がRに従うことを選好するという条件下で、Rに従うことを選好する。3. 集団Pのほとんど全ての行為者が、他のほとんど全ての行為者がRに従うことを選好することを信じている。4. Rは、繰り返される相互行為における一つのナッシュ均衡である。」(Voss, 2004, p.108) 更に、Vossは社会規範と慣習とを区別している。「5. 社会規範そのものは、集団のメンバーが、Rに従わない場合は、(負の) サンクションによって罰せられることが(正の蓋然性をもって) 期待されるようなRである。他方、慣習とは、反復的な調整状況における調整均衡であるようなRである。」(Voss, 2004, p.109)

上記の定義は、社会規範を反復的社会状況における「行為の規則性のあるナッシュ均衡」と見なすというゲーム理論的社会的規範論の基本的な考えを反映している。

<sup>9)</sup> 行為の規則性に関しては二種類が考えられている。第一は、全ての行為者が同じ行為を採るという場合であり、第二は、行為者には「タイプ」があり(第一の場合には行為者は同じタイプあるいはタイプによる区別はない)、タイプAの行為者は行為xを、タイプBの行為者は行為yを選択するというものである。この場合分けを行った時でも、タイプAまたはタイプBの行為者は全て、同じxまたはyを選択すると考えることができ、行為の規則性が見られる。

### 3. ゲーム理論的社会規範論内における相違点

社会規範は複雑な社会的事象であるために、ゲーム理論家の間で意見が完全に一致しているわけではない。この意見の不一致な点こそ、更なる社会規範論の展開に必須な点であろう。その相違点とは以下のものである。

#### ① 社会規範は「社会的な望ましさ」を実現するべきか。

「行為の規則性があり、かつナッシュ均衡」であるという状態は、必ずしも「社会的に望ましい状態」ではないかもしれない。「社会的に望ましい状態」をパレート効率性で定義すれば<sup>00</sup>、「行為の規則性があり、かつナッシュ均衡」はパレート劣位であるかもしれない。また、逆に、パレート効率的な状態はナッシュ均衡でないかもしれない。

#### ② 社会規範の形成にサンクションは必要か。

社会的な望ましさが必要とした場合、全ての行為者がこれを実現するよう行為を選択すれば、問題ないが、もし一部の行為者がこの行為ではなく、別の行為を選択（「逸脱」）した場合、この違反者に負のサンクションが与えられることが予期されよう。また、行為者もこのサンクションを予期して、逸脱行為は行わないであろう。これにより社会規範は維持されるが、サンクションは社会規範において必須の条件であるのであろうか？

#### ③ 社会規範に従った行為の性格

最後に行為の規則性の性格についてである。社会規範と行為の規則性との間には微妙な問題がある。社会規範とは行為の規則性を生み出す「何か」であろうか？あるいは行為の規則性そのものに現れる「何か」であろうか？「行為の規則性そのものに現れる何か」とは行為選択の結果である「秩序または均衡」と理解される。行為の規則性あるナッシュ均衡とは、この意味で「秩序」である。この時、社会規範とは行為の規則性そのものに現れている。つまり、行為選択の結果が行為の規則性であり、この規則性そのものが行為選択の指示である。そして、規範的行為とは、この状態を実現する行為、ナッシュ均衡戦略である。他方、行為の規則性を生み出す「何か」とは、行為する以前に、行為選択がそれに基づくもの、つまり、個人の性向や価値観が想定されよう。あるいは、義務感に基づき行為を選択するという場合もこれに含まれよう。行為の規則性を生み出す「何か」を仮に規範意識とすれば、行為者は行為の選択以前に前もって規範性を持つ

<sup>00</sup> 社会的合理性はふつう、パレート効率性により定義されるが、ハルサニ基準（平均効用最大化）、ヒックス・カレドア基準の場合もある。法と経済学の立場に立つEllickson（2001）は、「ヒックス・カレドア基準」（規範の存在により得られる個々人の便益が、規範の存在により失われる損失よりも大きい）に依っている。

ており、行為の規則性とは、この行為選択に先立つ規範意識により生み出される。この最後の点は重要であるので第Ⅲ節「『行為そのものの選択』とは何か。」で詳論し、次節では残りの2点の相違点および選好論からゲーム理論的社会規範論の類型的把握を試みよう。

## Ⅱ. 社会規範論の類型

### 1. 慣習と社会規範

第一の、そして主流的な立場は、社会規範を（後述するように慣習という言葉がふさわしいが）、行為の規則性をもつ「複数の安定したナッシュ均衡の一つ」（Sugden, 2004 [1986], p.33）とするものである。この時、社会規範とは行為の規則性そのものにおいて現れる状態＝「均衡」である。また、この立場では、複数の安定したナッシュ均衡がある場合、そのどれが選択されるかに関してパレート効率性または改善性で定義される社会的望ましさは必要とされない。非効率的なナッシュ均衡であっても慣習と見なす。「慣習」論<sup>11)</sup>を展開したのは、ヒュームの「慣習」論に進化ゲームを適用したSugden（2004 [1986]）である。Sugdenは、進化ゲームによる「慣習」論の嚆矢をなし、「2つ以上の安定均衡を持つゲームにおけるひとつの安定均衡」を慣習とした。複数のナッシュ均衡（ナッシュ均衡が一つしかない場合、そのナッシュ均衡は慣習ではない）の中で、進化的に安定な戦略を慣習とする、というその後の進化ゲームを用いた社会規範論の基本的着想を示した<sup>12)</sup>。Sugdenによる慣習の定義は以下のものである。

「集団Pの成員の行為の規則性Rは、成員が反復的状况における行為者である時、(1) 誰もがRに従う、(2) 誰もが他の行為者がRに従うことを期待する、(3) 誰もが、Rに従うならば、Rに従うことを選好する、以上の条件が、全ての成員に当てはまり、かつ共有知識である時に限り、慣習（convention）である」（Sugden, 2004 [1986], p.34）

この慣習の定義は純粹調整ゲームを用いて説明できる。純粹調整ゲームとは、次のような構造を持つゲームである。選択される行為は単に、AかBかである。行為者の選好は利己的と仮定する。利得行列の内、最初の数字はプレーヤー1の、次の数字はプレーヤー2の利得を表す。

<sup>11)</sup> 「慣習」論は、Lewis（2002 [1969]）における「慣習＝行為の規則性」に基づく。Lewisは慣習に続けて規範論を展開しているが（Cf. Lewis（2002[1969], pp.97-100））、規範とは、それに従わない場合にサンクションを受ける慣習であるとした。

<sup>12)</sup> 「安定した」とは進化的に安定した、との意味である。進化ゲームによる社会規範論は、例えば、Binmore（1994）、Skyms（1996）、Young（2001）を参照のこと。

表1 パレート支配のある調整ゲーム

	A	B
A	2, 2	0, 0
B	0, 0	1, 1

表2 パレート支配のない調整ゲーム

	A	B
A	2, 2	0, 0
B	0, 0	2, 2

以上の利得構造が与えられた場合、行為者は相手がどの行為を選択するのかの予想（期待）に応じて、戦略合理的に自己にとって最善の行為を選択し、かつ、この「期待」が全ての行為者において一致すれば、(A, A) または (B, B) がナッシュ均衡であり、行為の規則性を持つ。行為者が共に (A, A) または (B, B) を採る場合では、そうでない場合の (A, B) または (B, A) に比べて、個人利得が増す。パレート支配のあるゲームの場合（表1）、(A, A) は (B, B) よりも「望ましい」が、共に慣習とされる。どちらが実際に慣習として定着するか、はある意味で偶然の問題である<sup>103</sup>。パレート支配のないゲーム（表2）では、(A, A) または (B, B) はどちらも慣習である。例えば、Aを「車は右側通行」、Bを「車は左側通行」とすれば、複数の行為者が全員「右側通行」を選択すれば、(右, 右) が慣習として定着する。また、全員が「左側通行」を選択すれば、(左, 左) が慣習として定着する。このゲームでは、安定したナッシュ均衡は複数（2つ）存在し、その内の一つが「実際の慣習」として定着するが、残りの一つも、実現しなかった慣習である。

サンクションの必要性に関してはどうであろうか？ 慣習から規範への移行は、ひとたび慣習が確立されたら（「慣習」が存在するという「信念」を誰もが持つようになる）、誰もがそれに従うことに利益を見いだすであろう。従って、もし誰かが慣習に従わない行動をとったら、それは他者の利益を損なうことになるので、違反者に対して制裁、処罰が加えられるであろう。つまり、違反者に負のサンクションが与えられるであろう。この時に慣習は「規範」に移行する（Sugden, 2006 [1986], p.170）。

しかしながら、上記の調整ゲームにおける慣習と後述の囚人のジレンマゲームにおける「規範」

<sup>103</sup> 「複数の安定したナッシュ均衡」の中でどれが選択されるかの問題は、「ナッシュ均衡選択問題」（NESP: Nash equilibrium selection problem）と呼ばれる。この時には、利得支配ではなく、リスク支配等の別の効率性基準が使用される。

との最大の違いは、裏切りの誘因が存在しないことにある。純粹調整ゲームにおいては、行為の結果に対する各行為者の選好は一致している。従って、どの結果が選ばれるべきかに関して利害の対立がない。よって、サンクシヨンの必要性はないのである。上記は純粹調整ゲームを使った慣習と規範の説明であるが、一般に慣習論的社会規範論では、サンクシヨンは社会規範の必須の要件とはされない。このために、この立場に立つ議論は規範と慣習との区別は曖昧であり、規範とは結局慣習の中に包摂される<sup>94</sup>。

## 2. 制裁と社会規範

利己的な行為が必ずしも社会的に望ましい結果を生まない場合、社会規範を、利己心を抑制する行動の「基準」と考えることは自然であろう。そして、私的合理性と社会合理性（パレート効率性）の乖離という事態を「ジレンマ」と呼べば、利己心の抑制が必要となる社会的状況とは「囚人のジレンマ」的状況として定式化される。囚人のジレンマタイプのゲームは以下の利得行列を持つ（表3）<sup>95</sup>。Cは「協力」、Dは「非協力」を表す。プレイヤーの選好は利己的、従って、 $CC > DC > DD > CD$ である。利得行列の内、最初の数字はプレイヤー1の、次の数字はプレイヤー2の利得を表す。

表3 囚人のジレンマゲーム

	C	D
C	R, R	S, T
D	T, S	P, P

$$(T > R > P > S)$$

つまり、行為者は、相手が協力の時に、自分は非協力を選好する。両者が協力すれば、非協力の場合よりも利得が高く、パレート効率的である。しかし、囚人のジレンマには裏切り＝「非協力」のインセンティブがある。協力することにより断念される利得（「協力の機会費用」）を  $T - R$  とすると、 $T - R > 0$  であり、常に「非協力」のインセンティブがある。従って、ナッシュ均衡は（非協力、非協力）となり、社会的に望ましい状態と利己心の追求の結果とが乖離する。このとき、「裏切りが支配戦略である時、協力はいかにしたら可能か？」という問題がゲーム理論の解決すべき問題である。

<sup>94</sup> 「進化的に安定な戦略は慣習（convention）と見なすことができる」（Weibull, 2002 [1995], p.34）。

<sup>95</sup> 更に、 $R > \frac{T+S}{2}$  を条件に加える。これはプレイヤーが交互にC—D—C—Dを繰り返す。つまり、裏切りあうことによっては、協調した場合の利得を上回ることはできないことを示す。

この解決の一つがフォーク定理である。無限繰り返し囚人のジレンマゲームで、割引因子が十分1に近いならば、ナッシュ均衡戦略の与える利得（「留保利得」）を上回る利得を実現する戦略、トリガー戦略が存在し、この戦略はサブ・ゲーム完全ナッシュ均衡である。必ずしもパレート効率性を実現するわけではないが、ナッシュ均衡に対してパレート改善的である（Cf. Fudenberg and Tirole, 2002 [1991], pp.150-60）。近視眼的なプレーヤーは非協力＝「自白」により一時的に高い利得を得るが、長期的な利益を考えるならば、協力＝「黙秘」した方が良いのである。また、「直接的サンクション」（利得構造の変化をもたらす）を導入することにより、裏切りに対して負のサンクションを与え、非協力による利得を魅力的なものにするという方法も考えられた<sup>66</sup>。このゲームでは、社会的に望ましい状態を実現する「協力」が規範的行為であり、二人のプレーヤーが（協力、協力）という同じ行為を採る時、協力規範が成立していると考えられる。つまり、協力規範とは、「囚人のジレンマ的社会状況におけるナッシュ均衡であり、かつパレート効率的な状態」である<sup>67</sup>。「ナッシュ均衡であり、かつパレート効率的（または改善的）な状態」を規範の成立している状態と見なすことは、しかし、囚人のジレンマゲームを用いた規範論の本質を捉えるものでない。この立場では協力規範の実現・維持のためにはサンクションが不可欠であるという点にある。従って、このゲームは利己心の抑制という社会規範の要件の一つを満たしている。前述の純粋調整ゲームにおける慣習は、この立場によれば慣習ではありえても、規範ではない。

「社会的状況において規範は、個人がある行為をし、またこの行為をしない場合にはしばしばサンクションを受ける、その程度において存在する。従って、規範の存在はあるなしの問題ではなく、程度の問題である。」（Axelrod, 1997 [1984], p.47）

しかしながら、囚人のジレンマゲームを用いた社会規範論は、利己心の抑制は長期的には、自己の利益になるとの想定（「啓蒙された利己心」）に基づくものであるので、義務論的要素はその定義から欠け、「行為の規則性+サンクション」により社会規範を定義することになる。

### 3. 社会的選好論と社会規範

上記2つの立場では、行為者の選好は利己的であると想定されているが、全ての行為者は利己的選好を持つという前提に対し、ある行為者は社会的選好をもつとしたらどうであろうか？ 例えば、囚人のジレンマゲームにおいて、利己的選好の持ち主は「非協力」を選択するが、社会的

<sup>66</sup> 規範ゲームにおけるメタノームつまり、ルール違反者を罰しないプレーヤーを罰するべきとする規範に関しては、Axelrod (1997 [1984], pp.48-55)。

<sup>67</sup> どのようなタイプのゲームが「囚人のジレンマ」であるのかに関しては、議論がある。Taylor (1987) は「チキンゲーム」も「ジレンマ」状態を表現するとしている。

選好の持ち主は「協力」を選択するかもしれない。そして、社会的選好の持ち主同士がこのゲームを行えば、相手を裏切ることなく、協力が容易に実現するであろう。これにより、囚人のジレンマゲームはゲーム自体がある意味で「解消」されよう。ゲーム理論的状况において、戦略的合理性に加えて、性向という「非戦略的」要素が行為選択において存在すること、また、非戦略的サンクションの必要性を社会規範の要件とするのは、行動経済学で採用される社会的選好論である。特定の社会的状況において、ある行為者が社会的選好に合致した行為を採るとする。この時、行為者間で、「行為の規則性があり、かつナッシュ均衡」が実現すれば、それが規範の成立している状態である。つまり、社会的選好によって行為の規則性が生み出される。社会規範が成立している状態では、他者の抱く期待、つまり、誰もが規範的行為を採るだろうとの期待に反して利己的な行為を行うことは、「サンクション」(普通は負のサンクション)の対象となり、またサンクションにより規範は維持される。

「社会規範とは、いかに振る舞うべきかについて社会的に共有されている信念に基づいた行為の規則性であり、また、社会規範により指示された行為はインフォーマルな社会的なサンクションにより執行される。」(Fehr and Gächter, 2000, pp.517-18)

社会的選好には様々なタイプの選好が考えられている。代表的なものは、不平等嫌悪的選好、つまり、分配の不公平さを嫌悪し、公平性を好ましいとする選好、あるいは、互酬性、利他主義、嫉妬、等である(Camerer, 2004, pp.101-10)。もし、プレーヤーの一部のものがこの選好をもつとしたら、ゲームの「プレイの仕方」は根本的に異なるものとなろう。社会的選好論は以下の基本的前提に基づき社会規範を説明する。第一に、諸個人の選好には「異質性(heterogeneity)」(Fehr and Schmidt, 2004 [1997], p.272)があるということである。通常想定される利己的選好以外の他者顧慮的選好または過程顧慮的選好を持つ個人が存在し、社会的相互行為はこれらの異質の選好を持つ行為者の行為からなる。第二に、社会的状況の様々なタイプ(完全に競争的な、囚人のジレンマ的な、調整ゲーム的な)により、社会的選好に基づく行為選択が異なる結果をもたらすという「行為選択の状況依存性」が導入される。この社会的選好の存在と状況の論理により、利己的モデルでは説明できない一見非利己的、反利己的動機に基づく行動、例えば、公平性や互酬性といった規範に基づくものとしてしか説明できない選択のパターンが存在することが説明できるとするのである(Fehr and Fischbacher, 2004, pp.1-5)。

分配の公平性を選好する社会的選好とは、Fehr and Schmidt (2004)における「自己中心的不平等嫌悪(self-centered inequality aversion)」的選好が代表的である。これは不平等な分配への嫌悪、つまり、分配の平等性への選好の存在を示す。彼らが示した特殊な効用関数では、自分に

不利な不平等な分配に対する嫌悪度が、自分に有利な不平等な分配に対する嫌悪度よりも大きいことを意味するものとなっている。この関数式で、分配の公平性への選好は、分配の不公平の嫌悪度を示すパラメーターで示され、分配の公平性への選好とは、その人固有の性向を表すものとされる。つまり、彼らにとって社会的規範とは社会的「選好」という全く主観的なものなのである。囚人のジレンマゲームにおいて、ある行為者が社会的選好の一つである互酬的選好を持ち、他の行為者が利己的選好を持つ場合はどうであろうか。囚人のジレンマがジレンマである理由は、両者が利己的選好を持つからであるが、社会的選好論はこの前提を変える。ゲームは次のように進行する (Fehr and Fischbacher, 2004, pp.13–21)。

1. 同じ互酬的選好の持ち主と相互行為した場合には、協力を選好する。
2. 利己的選好の持ち主と相互行為した場合に、もし利己的選好の持ち主が「非協力」を選択した場合には「負のサンクション」を与える。この「負のサンクション」の執行に費用がかかり、またこの費用が将来の回収される見込みがなくとも、サンクションを与える。つまり、サンクションは戦略的行為ではなく、非一戦略的な (non-strategic) 「罰を与えたい」という欲求による。
3. この互酬性により、協力がナッシュ均衡 = 「互酬的均衡」 (Gintis, 2000, p.265) として成立する。

社会的選好論における規範論が、前述の2つの理論と異なるのは、行為の行われる以前に、規範への性向を持っているという点であり、更に、サンクションも、戦略的行為ではなく、非戦略的な (non-strategic) 「罰を与えたい」という欲求に求められている点である。利己的選好の持ち主であれば、サンクションのコストが正であれば、「非協力的な相手には罰を与える」という行為は行わないであろう。この立場の社会規範論は、非戦略的な行為選択 (義務論的な行為選択) の契機およびサンクションの必須性 (サンクションによる利己心の抑制) を強調することにより、前二者とは異なっている。しかしながら、この立場では、規範とは、効用関数の中に現れるパラメーターが示すように、「人が生まれながらにして持つ社会的性向」または個人の「タイプ」 (Fehr and Gächter, 2004, p.511) なのである。なぜ、特定の人が社会的選好を持つのか? と問う必要はないのであって、特定の人には「実際に」、この選好に基づく行為を選択していることが示されれば十分なのである<sup>98)</sup>。

<sup>98)</sup> 社会的選好論は進化ゲームと結びつくことにより、選好の内生的出現という問題が考察可能となった。CF. Bowles (2004, pp.381–90)。

#### 4. 評価

以上をまとめれば次のようになる。ゲーム理論的社会規範論では、社会規範は一般に「行為の規則性とナッシュ均衡」により定義される。これは、社会規範を行為の方向付けおよび「社会秩序」の二側面から把握していることを示している。慣習論的社会規範論では、社会規範とは「行為の規則性のあるナッシュ均衡の一つ」である。社会的合理性は、必ずしも社会規範の要件ではなく、また、サンクションの重要性も認められない。規範とは慣習のサブカテゴリーであるとの結論は、行為の規則性と安定したナッシュ均衡により規範を定義するという方法から必然的に生じよう<sup>99</sup>。囚人のジレンマゲームを用いた社会規範論では、規範とは、パレート効率的または改善的なナッシュ均衡であり、サンクションの要素は不可欠である。この立場では、利己的行為の抑制と社会的望ましさの実現という社会規範の必須の要件の一つを満たしているが、行為選択は戦略的である。最後に、社会的選好論による社会規範論では、非戦略的な行為選択の基準がゲーム理論的社会状況においても存在することを示している（「いかにふるまうべきかについての社会的に共有された信念」）。またサンクションは必須の契機である。しかし、社会的選好は個人的性向であり得るという点が最大の問題点である<sup>100</sup>。つまり、社会規範の「客観性」が保証されていない。

#### 5. ゲーム理論的社会規範論に内在する問題性

社会規範を考える上で、しかしながら、「行為の規則性とナッシュ均衡」によって社会規範を定義することは「ゲームの内部でのみ規範を定義する」あるいは「戦略的合理性により規範を定義する」<sup>101</sup>ことであるが、これは、慣習については受容しうる考えかもしれないが、社会規範という場合には、奇妙な事態が生じる。囚人のジレンマゲームによる協力規範を見てみよう。二人の犯人が実際には罪を犯しているのに共謀して自白しないことを「黙秘」、贖罪意識により犯罪を告白することを自白としよう。この時、（黙秘、黙秘）が協力規範である。つまり、黙秘は正直に犯した罪を自白することよりも「望ましい」ことになる。二人の犯罪者の間では黙秘は確かに「協力」であるが、これを規範的行為とすることはできないであろう。つまり、二人の犯罪者にとっては「規範」であるが、この二人の「外側にいる我々」からすれば、反規範的行為であろう。この場合、「外側にいる我々」は、犯した罪を正直に告白して、罰を受けることが「正しい」と確信しているのである。この場合に問題となっているのは、ゲームの構造により定義される「内部の行動

<sup>99</sup> 規範と慣習とを区別する論者の代表はElsterである。社会的規範の「合理化された利己心」的解釈への反論は、Elster (1989, 130ff).

<sup>100</sup> 同様な批判は、Bicchieri (2006, pp.109-10).

<sup>101</sup> Harsanyiの「合理的プレーヤーは同じ利得を生み出す戦略に関しては無差別である」(Harsanyi, 1977, p.113) という「合理性の公準」による。

基準」とその「基準」が「外部の我々」にとってもまた正しいものであるかどうかという「内部—外部」の関係である。あるいは、内部規範が「正しい」ものかどうかを判断する参照枠の問題である<sup>22)</sup>。

囚人のジレンマゲームではなく、次のような調整ゲームを考えよう（表4）。利得行列の内、最初の数字はプレーヤー1の、次の数字はプレーヤー2の利得を表す。

表4 調整ゲーム

	黙秘	自白
黙秘	3, 3	0, 0
自白	0, 0	3, 3

この場合、（黙秘，黙秘）も（自白，自白）ももたらす利得は等しく、どちらを選択するかの基準はない。つまり、両者は無差別である。更に定義により両者は共に慣習あるいは規範である。「ゲームの内部でのみ規範を定義する」場合、黙秘と自白のどちらが「正しい」のかを判断する基準は与えられない。我々が社会規範に求めたいのは、自白は黙秘よりも「望ましい」とする基準であろう<sup>23)</sup>。Mishra (2006) は進化ゲームを用いて、「腐敗した行為者」と「正直な行為者」の二つのタイプの行為者からなるゲームで、腐敗戦略が進化的に安定な戦略となる場合があることを示した。このとき、「腐敗は規範からの逸脱ではなくて、規範そのものである」（Mishra, 2006, p.355）。この場合にも、規範と反規範＝腐敗も区別することができないことになる。確かに、不正が蔓延し、それが行動の基準となっている集団も存在するであろうが、定義上、腐敗とは規範からの逸脱を意味しよう（法からの逸脱は違法である）。しかしながら、ゲーム理論的社会規範論による社会規範は、腐敗と規範とを区別することはできないのである<sup>24)</sup>。社会的選好による社会規範論では、行為者がある種の社会的選好を持ち、この性向により、ある行為を別の行為よりも「望ましい」「正しい」ものとして選択するということが可能であろう。この場合の問題は前述したように、この「規範」は個人の「性向」であり得るということであろう。

<sup>22)</sup> 社会規範を「ある集団の内部において」確立された行動の様式として定義するというのがゲーム理論的社会規範論の「パラダイム」である。この考えは正しい。その集団を含むより外延的な集団—社会—からみて、どれほど「反規範的」「非倫理的」であっても、その集団の内部では「規範」なのである。「社会規範は社会的利益をもたらす行為を促進することもあれば、これを阻害することもある」（Fehr and Gächter, 2004, p.518）。従って、もし複数の「行動の様式」が可能であるとしたら、その選択基準もまた「内部」にのみ求められよう。ゲーム理論ではナッシュ均衡の「安定性」と「唯一性」である。以上より、ゲームの「外」を考えることは、ある意味で約束違反であり、「正しく」ないと思惟が進行するならば、これは誤りであろう。

<sup>23)</sup> 社会規範を慣習的均衡に求めることを強く批判したのは、Elster (1989) である。Elsterによれば、行為選択は結果主義的な利己的合理性および社会規範への配慮からなされる。Cf. Elster (1989, Ch. 3)。

<sup>24)</sup> Mishraは、更に、このように腐敗が規範である場合には、「ゲームから退出する選択」、つまり、ゲームの「外」へでるという選択の可能性を挙げた。

以上、ゲーム理論的社会規範論を総括したが、この規範論は、本稿の最初で挙げた社会規範の要件の全てを満たしているわけではないことが明らかとなった。社会規範の問題の核心は、社会的状況における、行為そのものの選択の基準、つまり、ある行為の組のもたらす利得が、別の行為の組のもたらす利得と同じであっても、ある行為を別の行為に対して「望ましい」とする基準による選択、さらにこの「基準」が単なる性向によるのではない、「客観的な」基準の存在にあるのではないだろうか。

### Ⅲ. 「行為そのものの選択」とは何か。

上記は、社会規範論においては、行為そのものが「正しい」が故に選択されるという観点が必要であることを示している。

#### 1. 「行為そのものの選択」

行為そのものの選択とは、行為AとBの間で、行為の結果を顧慮することなく、AまたはBを選択するということである。明らかにこのままでは、なぜ、ある行為が選択されるのかわからない。次のような例を考えよう。「私は自分の労働により千円を得る」と「私は盗みによって千円を得る」のは、労働と盗みの「コスト」が同じであれば、利己的選好の持ち主にとっては無差別となる。しかしながら、ある種の社会的選好の持ち主にとっては、労働 $>$ 盗みとなり（労働により得た千円の効用は盗みによって得た千円の効用よりも大きいと考える）、「労働」は「盗み」よりも選好されるかもしれない。しかし、社会的選好は個人の性向であり得、「私」が盗みよりも労働を選好するにしても、「あなた」も同様であるとは限らない。私は盗みよりは労働を選好するが、あなたも盗みよりも労働を選好するという場合に、この行為選択には規則性がある。この場合には、この行為選択の規則性を生み出す「何か」、行為そのものの選択に関わる基準が想定されているのである。行為そのものの選択に関わる基準に関して、Vanberg (2006) は、「行為そのものに対する選好 (preference over actions 'as such')」を提唱した。Vanbergは「純粋に規範または規則に従った行動 (genuinely rule- or norm-guided behaviour)」を規範的行為と見なし、規範とは、「自らの行為から帰結するであろう結果に関する選好から独立し、それに先行する行為に対する選好」(Vanberg, 2006, p.9) を意味するものとし、「行為は結果に関してではなく、行為者が従うルールに一致して行われる」(ibid, p.10) とした。Vanbergは非結果主義的な行為そのものの選択に関わる「ルール」を規範としているのである。また、Bicchieri (2006) は規範に依拠した行動 (norm-based behavior) の存在を示した。Bicchieriによれば、社会規範とは、「行動のルール」であって、他人の行動に関する自己の期待を、なすべきことにマッピングする関数 (Bicchieri, 2006, p.219)

である。ここで、我々は、社会規範を行為選択の規則性を生み出す「行為選択そのものに関わるルール」とする立場に会う。

## 2. ルール顧慮的選好と「外部」

社会規範を「行為選択そのものに関わるルール」とした場合、行為者の行為の選択は、このルールを顧慮して行われると考えることができよう。選好が「行為選択そのものに関わるルールへの顧慮を示す」場合、「ルール顧慮的」選好であると定義しよう。この定義によれば、社会規範とは、問題的な社会状況における、行為の選択そのものに関わるルールである。つまり、行為の結果に必ずしも依存しない、行為者間で「共有」された行為選択のルール<sup>69</sup>があり、このルールが規範である。行為そのものの選択に関わる「ルール」の性格とはどのようなものなのであろうか。前節で我々は「行為の規則性があり、かつナッシュ均衡」を社会規範とすることは、「ゲームの内部でのみ規範を定義する」ことであり、「正しさ」という点では規範を定義できないと述べた。この「正しさ」の基準とはゲームの構造の「外部」に求めるべきではないだろうか。この「内部」のみによっては解決されない「正しさの基準」の問題を何らかの「外部」を想定することにより、解決しようとする試みは既にいくつか存在する。社会的状況は最小2人の行為より構成される（最小の「社会」）が、特定の問題的な社会状況、例えば、N人囚人のジレンマの状況では、集合行為問題が解決できないかもしれない。その場合に、この行為者の間に特定の「社会関係」（例えば、家族、友人、共同体等）を想定する。この場合、行為者を相互行為の「内部」とすれば、社会関係は「外部」にある<sup>70</sup>。

## IV. 結論

社会規範論の展開において、我々はこの「ゲームの外部」という視点を継承したい。先に、行為そのものの選択に関わるルールを社会規範としたが、ここで「外部」という概念を導入すれば、このルールはゲームの「外部」<sup>71</sup>にあるということができよう。従って、社会規範論で展開すべきものは、「ルール顧慮性とルールの外部性」という視点であろう。社会規範論の展開において「ルー

<sup>69</sup> Raub and Voss (1990) はセンによって提示された「選好に対する選好」(Sen, 2004) を「選好の合理的選択」論として展開した。

<sup>70</sup> 規範を維持に必要なサンクションは行為者の「社会関係」つまり、「義務と期待」とを生み出す「相互信頼のシステム」から生まれるとしたのがColeman (1990, Ch.11) である。

<sup>71</sup> Ostrom (2005) は規範の存在を示すものとして、デルタパラメーターを導入した。「デルタパラメーターは、共同体の規範またはルールへのコミットメントから生じるのであり、他人の利得を自分の利得の中に組み込むことではない。」(Ostrom, 2005, pp.173-34) 武藤 (2005) は、ゲームの「外部」の存在を「自明な期待」と呼び、プレイヤーは、ゲーム内部の「期待」に従う・従わない、「自明な期待」に従う・従わない、の二重のゲームを行うものとした。

ル顧慮性とルールの外部性」を恐らく初めて展開したのはアダム・スミス『道徳感情論』における「観察者」概念である。我々はスミスに従来（少なくとも現在までの）のゲーム理論的社会規範に含まれる困難を克服する可能性を見る<sup>28</sup>。

## 参考文献

- 飯田高 (2004) 『〈法と経済学〉の社会規範論』 勁草書房
- 田島慶吾 (2008) 「アダム・スミスと社会規範論」, 『経済研究』 静岡大学, 12 卷 4 号, 37-54 頁, 2008 年 2 月
- 武藤正義 (2005) 「日常世界的秩序問題のゲーム理論的分析」, 『ソシオロギス』, 19-34.
- 山岸俊男 (1991) 「社会的交換と囚人のジレンマ」 盛山和夫, 海野道郎編 『秩序問題と囚人のジレンマ』 ハーベスト社
- Ashraf, Camerer and Lowenstein (2005), “Adam Smith, Behavioral Economist,” *Journal of Economic Perspective*, vol. 19 (3)
- Axelrod, R. (1997), *The Complexity of Cooperation*, Princeton: Princeton University Press.
- Ben-Ner, A. and L. Putterman (1998), “Values, and institutions in economics analysis,” *Economics, Values, and Organization*. Eds. A. Ben-Ner and L. Putterman, Cambridge: Cambridge University Press: 3-69
- Bicchieri, C. (2006), *The Grammar of Society*, Cambridge: Cambridge University Press.
- Binmore, K. (1994, 1998), *Game Theory and Social Contract*, Vol1. Vol.2. Cambridge, MA: The MIT Press.
- Bowles, S. (2004), *Microeconomics Behavior, Institutions, and Evolution*, New York : Russell Sage Foundation.
- Camerer, C. F. (2004), *Behavioral Game Theory*, New York: Russell Sage.
- Camerer, C. F. and R. H. Thaler (1995), “Anomalities: Ultimatums, Dictators, and Manners,” *The Journal of Economics Perspective*, Vol.9, No.2 (Spring, 1995) : 209-219
- Coleman, J. S. (1990), *Foundations of Social Theory*, Cambridge, MA: The Belknap Press of Harvard University Press.

---

<sup>28</sup> アダム・スミスとゲーム理論的社会規範論の結びつきは唐突であろうか。Ashraf, Camerer and Lowenstein (2005) は、スミスは「最初の行動経済学者であった」と主張している。スミスの『道徳感情論』を社会規範論として展開する試みについては田島 (2008) を参照。

- Ellickson, R. C. (2001), “ The Evolution of social norms: A Perspective from the Legal Academy,” *Social Norms*. Eds. M. Hechter and K. Opp, New York: Russell Sage Foundation: 35–75
- Elster, J. (1989), *The Nuts and Bolts for the Social Science*, Cambridge: Cambridge University Press.
- Fehr and Fishbacher (2004), “ Third-party punishment and social norms,” *Evolution and Human Behavior*, Vol.25 (2004) : 63–87
- Fehr, E. and S. Gächter (2000), “ Fairness and retaliation: The economics of reciprocity,” *Journal of economic perspectives*, Vo.14 (3) : 159–181
- Fehr, E. and K. M. Schmidt (2004 [1997]), “ A Theory of Fairness, Competition, and Cooperation,” *Advances in Behavioral Economics*. Eds. C. F. Camerer, G. Lowenstein and M. Rabin, New York: Russell Sage Foundation: 271–296
- Fudenberg, D and J. Tirole (2002 [1991]), *Game Theory*, Cambridge, MA: The MIT Press.
- Gintis, H. (2000), *Game Theory Evolving*, Princeton: Princeton University Press.
- Griffith, W. B. and R. S. Goldfarb (1991), “ Amending the Economist’s “ Rational Egoist ” Model to include Moral Values and Norms,” *Social Norms and Economic Institutions*. Eds. K. J. Koford and J. B. Miller, Ann Arbor: The University of Michigan Press: 39–84
- Hardin, R. 1993 [1982], *Collective Action*, Baltimore: The Johns Hopkins University Press.
- Harsanyi, J. C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.
- Lewis, D. (2002 [1969]), *Convention*, Oxford: Blackwell Publishers.
- Mishra, A. (2006), “Persistence of Corruption: Some Theoretical Perspectives,” *World Development*, Vol.34 (2): 349–358
- Ostrom, E. (2005), *Understanding Institutional Diversity*, Princeton: Princeton University Press.
- Raub, W. and T. Voss (1990), “ Individual Interests and Moral Institutions,” *Social Institutions Their Emergence, Maintenance, and Effects*. Eds. M. Hechter, K. Opp and R. Wippler, New York: Aldine de Gruyter
- Sen, A. (2004), *Rationality and Freedom*, Cambridge, MA: The Belknap Press of Harvard University Press.
- Skyrms, B. (1996), *Evolution of Social Contract*, Cambridge: Cambridge University Press.
- Sugden, R. (2004 [1986]), *The Economics of Rights, Cooperation, and Welfare*, New York: Palgrave Macmillan.
- Taylor, M. (1987), *Possibility of Cooperation*, Cambridge: Cambridge University Press. (テラー『協力の可能性』松原望訳, 木鐸社, 1995年)

- Ulman – Margalit, E. (1977), *The Emergence of Norms*, Oxford: Oxford University Press.
- Vanverg, V. J. (2006), “Rationality, Rule-Following and Emotions: On the Economics of Moral Preference,” *Papers on Economics & Evolution*. Jena: Max-Planck-Institute für Ökonomie: 1 – 21
- Voss, T. (2001), “Game-theoretical Perspectives on the Emergence of Social Norms,” *Social Norms*. Eds. M. Hechter and K. Opp, New York: Russell Sage Foundation: 105 – 136
- Young, H. P. (2001), *Individual Strategy and Social Structure*, Princeton: Princeton University Press.
- Weibull, J. W. (2002 [1995]), *Evolutionary Game Theory*, Cambridge: The MIT Press.